

Open Source basierte Virtuelle Forschungsumgebungen für Projekte in der ökologischen Langzeitforschung

Das „Rapid L-TER VRE“ Baukastenkonzept
zur schnellen und nachhaltigen Realisierung

Von der Fakultät für Umwelt und Naturwissenschaften der Brandenburgischen Technischen Universität Cottbus – Senftenberg zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von
Diplom-Umweltingenieur

Mirko Filetti

aus
Offenbach, am Main (Hessen)

Gutachter: apl. Prof. Dr.-Ing. habil. Frank Molkenhain

Gutachter: Prof. Dr. rer. nat. habil. Albrecht Gnauck

Gutachter: Prof. Dr. rer. nat. habil. Gerhard Wiegler

Tag der mündlichen Prüfung: 28.06.2018

Kurzfassung

Virtuelle Forschungsumgebungen (engl. virtual research environments, im Folgenden als VRE bezeichnet), sind nach allgemeinem Verständnis Plattformen auf Basis von Informations- und Kommunikationstechnologien (engl. ICT) mit ganzheitlichem Ansatz, die das wissenschaftliche Arbeiten im Team und insbesondere dessen Kommunikation und Organisation unterstützen sollen. Sie haben zum Ziel, schneller und besser zu nachvollziehbaren Ergebnissen zu gelangen und Wissen persistent verfügbar zu machen. Derzeit existiert jedoch kein VRE-Software-Framework, das „out-of-the-box“ sofort einsatzfähig oder für alle Anforderungen geeignet wäre („one size fit them all“). Projekte in der ökologischen Langzeitforschung (engl. long-term ecological research, im Folgenden als L-TER bezeichnet) sind durch den Langzeitaspekt und Raumbezug in der Regel umfangreiche Projekte mit mittleren und großen Datenmengen (Medium Data und Big Data) sowie interdisziplinären, örtlich verteilten Forschungsgruppen. Sie sind daher prädestiniert für den Einsatz von VREs. Die vorliegende Arbeit untersucht die spezifischen Aspekte und Anforderungen an virtuelle Forschungsumgebungen von Projekten in der ökologischen Langzeitforschung. Dazu wird der Begriff „L-TER VRE“ eingeführt. Grundlage hierfür ist der Stand des Wissens zu allgemeinen VREs, Datenmanagement und ICT, die interdisziplinär in Verbindung gebracht werden mit den Zielen, Methoden und Daten der ökologischen Langzeitforschung. Es zeigt sich, dass sich die Anforderungen an L-TER VREs – aufgrund ihres Einsatzbereiches für langfristige L-TER Projekte – in der Nachhaltigkeit durch ihre Persistenz (Daten, Infrastruktur), Flexibilität (projektspezifische Rahmenbedingungen), der eingesetzten ICT und vom Faktor „Mensch“ mit seinem gesellschaftlichen Umfeld von VREs für andere Aufgaben unterscheiden. Aus den Erkenntnissen wird ein innovatives Baukastenkonzept mit dem Namen „Rapid L-TER VRE“ zur Planung, dem Aufbau und Betrieb von L-TER VREs vorgestellt. Es basiert auf Open-Source-Software und kann durch die gemeinsame Basis der Bausteine auch für allgemeine VREs Verwendung finden. Da VREs oder ähnliche ICT-Infrastrukturen im Rahmen von „e-Science“ als Rückgrat für die Organisation der Projekte prädestiniert sind, ist der Aufbau solcher Infrastrukturen oft schon zur Voraussetzung bei der Antragstellung von Projekten geworden. Ein Prototyp sollte daher bereits vor oder bei der Antragstellung existieren - spätestens zum Projektbeginn muss eine Produktivumgebung zur Verfügung stehen. Vor diesem Hintergrund ist das Konzept „Rapid L-TER VRE“ speziell für eine schnelle (engl. rapid) Entwicklung von VRE Prototypen mit einer kurzen Übergangsphase in den Produktivbetrieb ausgelegt. Das Baukastenkonzept wird auf verschiedene existierende Projekte im Umfeld des Fachgebietes „Environmental Informatics“ und der Fakultät „Umwelt und Naturwissenschaften“ der BTU Cottbus-Senftenberg konzeptionell, sowie in der Form von Prototypen angewendet und dessen Anwendbarkeit auch allgemein bewertet. Aus den Erkenntnissen werden allgemeine Schlussfolgerungen für den erfolgreichen Betrieb von L-TER VREs gezogen – hier steht der Mensch als „maßgeblicher“ Faktor im Mittelpunkt. Im Anhang finden sich zusammengefasste Informationen zu den ausgewählten Projekten sowie spezifischen Software-Empfehlungen (Open Source), die aus der praktischen Erfahrung dieser Arbeit stammen und zum Teil separat im längeren Produktivbetrieb getestet wurden.

Schlagwörter: Virtuelle Forschungsumgebung (VRE), ökologische Langzeitforschung (L-TER), Open-Source, e-Science, Langzeit-Datenmanagement, Umweltinformatik

Abstract

Virtual research environments (VREs) are information and communication technology (ICT) platforms that apply a holistic approach in supporting scientific communities. In particular communication and organization among members of a research group is necessary. With VREs scientific results can be obtained faster, more efficiently and shared with the community on the platform in an immediate, comprehensible and easily accessible form. At the moment, there is no VRE software framework existing, that would be immediately applicable - "out-of-the-box" and suitable for any field of research - "one size fit them all". Large scale and long-term research projects are often interdisciplinary, generate large data sets and may involve the collaboration of spatial distributed research groups. In this thesis project specific aspects and their associated requirements for VREs are examined on the basis of exemplary long-term ecological research (L-TER) projects. The premise for this objective is a profound comprehension of the general functioning of VREs, data management and ICT, which are merged from an interdisciplinary point of view with the objectives, methods and data management of long-term ecological research. It could be shown that in the field of sustainability the requirements of L-TER VREs differ from general VREs in terms of persistence (data, infrastructure), flexibility (project specific), ICT, and "humans" with their social interaction. Based on these findings, an innovative model kit called "Rapid L-TER VRE" was developed, which is comprised of several compatible modules. By combining them, an efficient framework for planning, setup and operation of L-TER VREs can be presented. "Rapid L-TER VRE" utilizes open-source software and can be applied in any VREs, thanks to the shared conceptual basis of the modules. In the context of "e-science" VREs or similar ICT-platforms are predestined to become the backbone of complex research projects and as a result of this, are often viewed as a prerequisite for novel research proposals. A prototype should therefore be available before or during an application process, but latest at the beginning of a research project. "Rapid L-TER VRE" was specifically designed for the rapid development of VRE prototypes having a short transition phase to an efficient infrastructural research tool in a productive environment. The model kit was applied conceptually, as well as in the form of a prototype to various existing projects from the chair of environmental informatics and the faculty of environment and natural science of the BTU Cottbus-Senftenberg. By assessing the VRE's practicality, general conclusions on the successful operations of VREs could be drawn. Humans with their communities must be the „evaluating“ and leading factor in the centre. The appendix contains further information on the presented research projects and related software recommendations (open source) from a „VRE construction kit“ that are based on the experience of project specific VRE applications. These have been partially tested in long-term applications and can be helpful for the technical implementation of individual VREs supporting an interdisciplinary team of researchers.

Keywords: Virtual Research Environments (VRE), long-term ecological research (L-TER), Open-Source, e-Science, long-term data management, Environmental Informatics

Inhalt

Inhaltsverzeichnis	I
Abbildungsverzeichnis	V
Tabellenverzeichnis	VII
Abkürzungsverzeichnis	XI
1. Einleitung	1
1.1 Motivation	4
1.2 Zielstellung und Forschungsfragen	5
2. Stand des Wissens und der Technik	7
2.1 Alte und neue Paradigmen der Wissenschaft	7
2.2 Entwicklung der ökologischen Informatik	9
2.3 Ökologische Langzeitforschung (L-TER)	11
2.4 Open Source – von der Insellösung zum „Software-Ökosystem“	13
2.5 Virtuelle Forschungsumgebungen	16
2.5.1 Aktueller Stand	17
2.5.2 Management von Forschungsdaten	22
2.5.3 Life-Cycle-Modell für VREs	24
2.5.4 Langzeitarchivierung innerhalb VREs	27
2.5.5 VRE Frameworks	29
2.5.6 Wirtschaftliche Aspekte	30
2.5.7 Erfolgskriterien für VREs	32
2.5.8 Data Policies	35
2.5.9 Rechtliche Aspekte	37
2.5.10 Technische Voraussetzungen von L-TER VREs	39
2.6 Nationale und europäische Aktivitäten	42
2.7 Relevante Technologien (ICT)	45
2.7.1 Remote Desktop, Screen- und Application Sharing	45
2.7.2 Cloud Computing	45
2.7.3 Virtualisierung und virtuelle Maschinen (VM)	47
2.7.4 Hochleistungsrechnen (HPC) und Grid-Computing	48
2.7.5 Content-Management-Systeme (CMS)	49

2.7.6 Datenbanken	50
2.7.7 Dataspace- und Semantic-Web-Modelle	52
2.7.8 Serviceorientierte Architektur (SOA)	53
2.7.9 Authentifizierung und Single Sign-On (SSO).....	53
2.7.10 Datenverbünde und Metadatenharvesting	54
2.8 Forschungsdaten in L-TER Projekten	55
2.8.1 Klassifizierung von Daten	56
2.8.2 Zeitskalen und Prozessklassen für ökologische Variable.....	58
2.8.3 Datenaufbereitung	59
2.8.4 Datenlebenszyklus.....	62
2.8.5 Metadatenmodelle und Auszeichnungssprachen.....	63
2.8.6 Modellkopplung	68
3. Voraussetzungen einer L-TER VRE	69
3.1 Einführung des Begriffes „L-TER VRE“ und weitere Definitionen.....	69
3.2 Funktionale Voraussetzungen von L-TER VREs.....	74
3.3 Projektspezifische Voraussetzungen	76
4. Das Baukastenkonzept „Rapid L-TER VRE“	79
4.1 Einsatz von Open-Source-Software	81
4.2 ICT-Basis für L-TER VREs	83
4.3 „Schlüselfertige“ Server und Virtualisierung	84
4.4 Homepage und CMS	85
4.5 VRE-Services	87
4.6 VRE-Tools.....	89
4.7 Authentifizierung und Einmalanmeldung (SSO)	91
4.8 Report-Layer.....	92
4.9 Benutzer Hilfe, Feedback und technische Dokumentation	94
4.10 Weiterentwicklung	96
5. Anwendungsbeispiele für „Rapid L-TER VRE“	99
5.1 L-TER VRE für die Nachsorge der „Altablagerung Großziethen“	101
5.1.1 Aktueller Stand des Projektes.....	101
5.1.2 Zielsetzung der L-TER VRE.....	103
5.1.3 Konzept und Prototyp.....	106

5.1.4 Bewertung der Konzeptanwendbarkeit	111
5.2 L-TER VRE für BTU Sonderforschungsprojekt „Hühnerwasser“	114
5.2.1 Aktueller Stand des Projektes.....	115
5.2.2 Zielsetzung der L-TER VRE.....	117
5.2.3 Konzept und Prototyp.....	118
5.2.4 Bewertung der Konzeptanwendbarkeit	120
5.3 L-TER VRE für Projekt „HydroEurope“ in der Lehre.....	124
5.3.1 Aktueller Stand des Projektes.....	124
5.3.2 Zielsetzung der L-TER VRE.....	125
5.3.3 Konzept und Prototyp.....	126
5.3.4 Bewertung der Konzeptanwendbarkeit	128
6. Gesamtbewertung des „Rapid L-TER VRE“ Baukastenkonzeptes	131
6.1 Bewertung projektspezifischer Rahmenbedingungen	132
6.1.1 Projektgröße	134
6.1.2 Projektstruktur.....	135
6.1.3 Projektschwerpunkte	136
6.1.4 Die Zeit in Bezug auf Projektzeitraum und Lebensphasen	136
6.1.5 Virtualisierung.....	137
6.2 Bewertung der Nachhaltigkeit.....	138
6.2.1 Persistenz.....	139
6.2.2 Flexibilität	141
6.2.3 ICT.....	141
6.2.4 Faktor „Mensch“	142
6.3 Bewertung der Entwicklungszeit.....	143
6.3.1 Virtualisierung.....	145
6.3.2 „Schlüselfertige“ Systeme	145
6.3.3 Programmieraufwand	146
6.3.4 Weiterentwicklung	147
6.3.5 Baukasten für VRE-Services und -Tools mit Softwareempfehlungen..	148
6.4 Bewertung der Konzeptbestandteile.....	149
6.4.1 Open Source	150
6.4.2 ICT-Basistechnologien.....	150

6.4.3 Schlüsselfertige Server und Virtualisierung.....	151
6.4.4 Homepage und CMS	151
6.4.5 VRE-Services	152
6.4.6 VRE-Tools.....	153
6.4.7 Authentifizierung.....	154
6.4.8 Report-Layer	154
6.4.9 Benutzerhilfe, Feedback und technische Dokumentation	155
6.4.10 Weiterentwicklung	155
7. Schlussfolgerungen.....	157
7.1 Wissenschaftlicher Beitrag der Arbeit.....	157
7.2 Offene Fragen und zukünftige Forschung.....	165
Literatur- und Quellenverzeichnis.....	167
Anhang.....	191
A Vor- und Nachteile von Open-Source-Software	193
B Tabellen für die VRE-Beispielprojekte	195
C Beispiele für VRE Basis-Softwarekomponenten	201
D „Joomla“ CMS: ausgewählten Erweiterungen für VREs	205
E Beispiele für Befehle zur Steuerung und Backup von VMs	211
F Schlüsselfertige „Linux“-Server für L-TER VREs	213

Abbildungsverzeichnis

Abb. 2.1:	Schwerpunkte von Bio-, Umwelt- und Ökologischer Informatik mit Überschneidungen	10
Abb. 2.2:	Skalenebenen in der Biologie und Ökologie mit Zugehörigkeit zur Bio-/ökologischen Informatik nach Neumann (2003).....	11
Abb. 2.3:	Unterschiedliche Zeitskalen von Prozessen und Forschungsbereichen nach Hobbie et al. (2003).....	12
Abb. 2.4:	Durch den Einsatz von Open Source und Open Data ergeben sich Open Ressources nach Fritzlar et al. (2012).....	15
Abb. 2.5:	Übergreifende Bereiche des Managements von Forschungsdaten nach Ludwig und Enke (2013).....	23
Abb. 2.6:	Lebensphasen einer VRE und deren Module (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0.....	25
Abb. 2.7:	Lebensphasen einer VRE im Detail (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0	26
Abb. 2.8:	Erfolgsfaktoren und messbare Erfolgskriterien für VREs nach Relevanzgewichtung in den Stakeholderkategorien unterschieden (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0	33
Abb. 2.9:	Erfolgsbeurteilung in Abhängigkeit vom Entwicklungsstand der VRE (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0	35
Abb. 2.10:	Beispiel einer Netzwerkstruktur für Datenverbände mit Netzwerkknoten.....	55
Abb. 2.11:	Resampling Prozedur nach Gnauck (2006)	61
Abb. 2.12:	Lebenszyklus von Forschungsdaten nach Ludwig und Enke (2013)	62
Abb. 3.1:	Dynamische Basis-Nachhaltigkeitsfaktoren einer L-TER VRE.....	71
Abb. 3.2:	Einzelne Parameter der Basis-Nachhaltigkeitsfaktoren einer L-TER VRE	72
Abb. 3.3:	L-TER VREs als Basis für eine „digitale Forschungsgemeinschaft“.....	73
Abb. 4.1:	Bestandteile des „Rapid L-TER VRE“ Baukastenkonzept	80
Abb. 4.2:	Report-Layer erstellt Berichte für verschiedene Zielgruppen und Datenverbände.....	92
Abb. 4.3:	Report-Layer stellt Berichte aus einzelnen VRE-Services und-Tools	93
Abb. 4.4:	Aufteilung in „Basis-“ und „erweiterte“ L-TER VRE für schnelle Realisierung und kontinuierliche Weiterentwicklung	97

Abb. 5.1:	Projektspezifische Eigenschaften (Größe, Schwerpunkt, Zeitraum) der Anwendungsbeispiele von L-TER Projekten	99
Abb. 5.2:	Luftaufnahme der Altablagerung Großziethen (LDS Umweltamt 2005).....	101
Abb. 5.3:	Altablagerung Großziethen - Messpunkten des Oberpegels nahe angrenzender Siedlung.....	102
Abb. 5.4:	Konzentrationsverlauf für Kohlenwasserstoffe des Unterpegels der Altablagerung Großziethen im Frühjahr 2010.....	105
Abb. 5.5:	„GeoNetwork“ Suchergebnis für „Blei 2010“ mit Vorschau	110
Abb. 5.6:	„GeoNetwork“ Schlüsselwörter für Metadatenbeschreibung mit INSPIRE Thesaurus.....	111
Abb. 5.7:	INSPIRE-konforme Suchoptionen: Anhang, Datentyp, Servicetyp, Themengebiet	111
Abb. 5.8:	Lage des Untersuchungsgebiets „Hühnerwasser“ im Tagebau Welzow-Süd, bei Cottbus (BTU Cottbus-Senftenberg 2017).....	114
Abb. 5.9:	Künstliches Wassereinzugsgebiet „Hühnerwasser“ – Primärsukzession Luftbilder 2007, 2009, 2010 (BTU Cottbus-Senftenberg 2017)	115
Abb. 5.10:	Permanente und zusätzliche Monitoreinrichtungen (BTU Cottbus-Senftenberg 2017).....	115
Abb. 5.11:	Prototyp „Hühnerwasser“ L TER VRE – Community Homepage.....	120
Abb. 5.12:	Homepage des VRE-Prototyps für „HydroEurope“ Workspace Menü.....	128

Tabellenverzeichnis

Tab. 2.1:	Unterschiedliche Auslegungen der Langzeit-Datenhaltung	27
Tab. 2.2:	Kostenarten einer VRE nach Buddenbohm et al. (2014)	31
Tab. 2.3:	Erfolgskriterien einer VRE nach Buddenbohm et al. (2014)	34
Tab. 2.4:	Vor- und Nachteile von VMs	47
Tab. 2.5:	Prozessklassen der L-TER Forschung mit Beispielen nach Müller et al. (2010a)	59
Tab. 2.6:	Beispiele für Metadaten Standards und Auszeichnungssprachen vgl. Jensen et al. (2011)	67
Tab. 2.7:	Beispiele umweltrelevanter Auszeichnungssprachen	67
Tab. 3.1:	Nachhaltigkeitskriterien von Ökosystemen (Renn et al. 2007) im Vergleich zu L-TER VREs	71
Tab. 3.2:	Abgrenzende Kriterien von L-TER VREs gegenüber allgemeinen VREs	74
Tab. 4.1:	ICT Klassen mit Beispielen für L-TER VREs	83
Tab. 4.2:	Aufgabenbereiche für VRE-Services in L-TER VREs mit Aufgaben und Beispielen	88
Tab. 4.3:	Beispiele für „innovative“ L-TER VRE Erweiterungen, gruppiert nach Klassen	98
Tab. 5.1:	Rahmenbedingungen des Projektes „Altablagerung Großziethen“ für eine L-TER VRE	103
Tab. 5.2:	Datentypen für das Projekt „AAGZ“	105
Tab. 5.3:	Basis- und erweiterte VRE-Services für die „AAGZ-VRE“	107
Tab. 5.4:	Verwendete Erweiterungen (VRE-Tools) zum „Joomla“ CMS für die „AAGZ-VRE“	107
Tab. 5.5:	Verwendete Erweiterungen (VRE-Tools) zu „MediaWiki“ für die „AAGZ-VRE“	108
Tab. 5.6:	Metadatenschema für die Hauptdatentypen auf Basis der ISO 19115:2003/19139 Norm für die „AAGZ“	109
Tab. 5.7:	Projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes für „AAGZ“	113
Tab. 5.8:	Rahmenbedingungen des Projektes „Hühnerwasser“ für eine L-TER VRE	117

Tab. 5.9:	Basis- und erweiterte VRE-Services für „Hühnerwasser“	119
Tab. 5.10:	Verwendete Erweiterungen (VRE-Tools) zum „Joomla“ CMS für „Hühnerwasser“	119
Tab. 5.11:	Verwendete Erweiterungen (VRE-Tools) zu „MediaWiki“ für „Hühnerwasser“	120
Tab. 5.12:	Projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes für „Hühnerwasser“	121
Tab. 5.13:	Rahmenbedingungen des Projektes „HydroEurope“ für eine L-TER VRE	125
Tab. 5.14:	Basis- und erweiterte VRE-Services für „HydroEurope“	127
Tab. 5.15:	Verwendete Erweiterungen (VRE-Tools) zum „Joomla“ CMS für „HydroEurope“	127
Tab. 5.16:	Verwendete Erweiterungen (VRE-Tools) zu „MediaWiki“ für „HydroEurope“	128
Tab. 5.17:	Projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes für „HydroEurope“	129
Tab. 6.1:	Gesamtbewertung des „Rapid L-TER VRE“ Baukastenkonzeptes.....	132
Tab. 6.2:	Bewertung der Flexibilität des „Rapid L-TER VRE“ Konzept.....	133
Tab. 6.3:	Bewertung der Projektgröße im „Rapid L-TER VRE“ Konzept.....	134
Tab. 6.4:	Bewertung der Projektstruktur im „Rapid L-TER VRE“ Konzept	135
Tab. 6.5:	Bewertung der Projektsschwerpunkte im „Rapid L-TER VRE“ Konzept.....	136
Tab. 6.6:	Bewertung der Zeit in Bezug auf Projektzeitraum und Lebensphasen im „Rapid L-TER VRE“ Konzept	136
Tab. 6.7:	Bewertung der Virtualisierung im „Rapid L-TER VRE“ Konzept	137
Tab. 6.8:	Bewertung der Nachhaltigkeit des „Rapid L-TER VRE“ Konzept.....	138
Tab. 6.9:	Bewertung der Persistenz in Bezug zur Nachhaltigkeit im „Rapid L-TER VRE“ Konzept	140
Tab. 6.10:	Bewertung der Flexibilität in Bezug auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept.....	141
Tab. 6.11:	Bewertung der ICT in Bezug auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept.....	142
Tab. 6.12:	Bewertung des Menschen (und Gesellschaft) in Bezug auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept.....	143

Tab. 6.13: Bewertung der „schnellen“ Entwicklung im „Rapid L-TER VRE“ Konzept.....	144
Tab. 6.14: Bewertung der Zeitersparnis durch „Virtualisierung“	145
Tab. 6.15: Bewertung der Zeitersparnis durch „schlüsselfertige“ Systeme.....	146
Tab. 6.16: Bewertung des Programmieraufwandes für das „Rapid L-TER VRE“ Konzept.....	146
Tab. 6.17: Bewertung der Weiterentwicklung im Rahmen des „Rapid L-TER VRE“ Konzeptes.....	147
Tab. 6.18: Bewertung des „VRE-Baukastens“ in Bezug zur Entwicklungszeit im „Rapid L-TER VRE“ Baukastenkonzept	149
Tab. 6.19: Bewertung der Konzeptbestandteile im „Rapid L-TER VRE“ Konzept...	149
Tab. 7.1: Gesamtbewertung des „Rapid L-TER VRE“ Konzeptes.....	160
Tab. 7.2: Open-Source-Software für eine Basis L-TER VRE nach dem „Rapid L-TER Konzept“.....	161
Tab. A.1: Vorteile von Open-Source-Software (Fritzlar et al. 2012).....	193
Tab. A.2: Nachteile von Open-Source-Software (Fritzlar et al. 2012).....	194
Tab. B.1: Komprimiertes „Rapid L-TER VRE“ Konzept für die Altablagerung Großziethen.....	195
Tab. B.2: Komprimiertes „Rapid L-TER VRE“ Konzept für „Hühnerwasser“	197
Tab. B.3: Komprimiertes „Rapid L-TER VRE“ Konzept für „HydroEurope“	199
Tab. C.1: Open-Source-Software für eine Basis-VRE nach dem „Rapid L-TER Konzept“.....	201
Tab. C.2: Verbreitete Datenbank-Management-Systeme (Open Source)	201
Tab. C.3: Software für Virtualisierung	201
Tab. C.4: Software für Remote-Desktop und Screen-Sharing	202
Tab. C.5: Verbreitete Content-Management-Systeme nach Breitenstrom et al. (2013).....	202
Tab. C.6: Verbreitete Verzeichnisdienste vgl. Klünter und Laser (2008).....	203
Tab. C.7: Verbreitete SSO-Produkte und Standards vgl. Perske (2012).....	203
Tab. C.8: „MediaWiki“ Erweiterungen für VREs.....	204

Abkürzungsverzeichnis

AAGZ	Altablagerung Großziethen
API	Application Programming Interface (Programmierschnittstelle)
AR	Augmented Reality
BDSG	Bundesdatenschutzgesetz
BMBF	Bundesministerium für Bildung und Forschung
BTU	Brandenburgische Technische Universität
CMS	Content-Management-System
CPU	central processing unit (Prozessor)
CSDGM	Content Standards for Digital Geospatial Metadata
DBMS	Datenbank-Management-System
DFG	Deutsche Forschungsgemeinschaft
DIN	Deutsches Institut für Normung
DINI	Deutsche Initiative für Netzwerkinformation e.V.
DNA	Desoxyribonukleinsäure
DOI	Digital Object Identifier
EDV	Elektronische Datenverarbeitung
EU	Europäische Union
FAQ	Frequently asked questions (deutsch: häufig gestellte Fragen)
FGDC	Federal Geographic Data Committee
FLOSS	Free Libre Open Source Software
FOSS	Free Open Source Software
FSF	Free Software Foundation
GB	Gigabyte
GDI	Geodateninfrastruktur
GESIS	Leibniz-Institut für Sozialwissenschaften
GIS	Geographisches Informatinossystem
GNU / GPL	Die GNU General Public License (GPL) ist eine von der Free Software Foundation herausgegebene Lizenz für die Lizenzierung freier Software.
HPC	High Performance Computing (Hochleistungsrechnen)

HTML	Hypertext Markup Language
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronic Engineers
INSPIRE	Infrastructure for spatial information in Europe
ISBN	International Standard Book Number
ISO	International Organization for Standardization
IT	Information Technology
KI	Künstliche Intelligenz
LDAP	Lightweight Directory Access Protocol
L-TER	Long-term ecological research (Ökologische Langzeitforschung)
MMS	Metadaten-Management-System
MS	Microsoft (Handelsmarke)
nSdT	Nachhaltiger Stand der Technik
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OECD	Organisation for Economic Cooperation and Development (deutsch: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)
OGC	Open GIS Consortium
OS	Operation System (Betriebssystem)
OSD	Open Source Definition
OSI	Open Source Initiative
OWL	Web Ontology Language
PDF	Portable Document Format (Dateiformat)
PDF-A	Portable Document Format (Dateiformat) für die elektronische Langzeitarchivierung
PHP	Hypertext Preprocessor (Programmiersprache)
RDF	Resource Description Framework
SdT	Stand der Technik
REST	REpresentational State Transfer (Architekturansatz für verteilte Systeme)
SEIS	Shared Environmental Information System
SEO	search engine optimization

SISE	Single Information Space in Europe for the Environment
SOA	Serviceorientierte Architekturen
SQL	Structured Query Language
SSO	Single Sign-on (Einmalanmeldung)
UrhG	Urheberrechtsgesetz
VM	Virtuelle Maschine
VR	Virtual Reality
VRE	Virtual Research Environment (Virtuelle Forschungsumgebung)
W3C	World Wide Web Consortium
WYSIWYG	What You See Is What You Get (Was du siehst, ist [das], was du bekommst)
XML	Extensible Markup Language
ZKI	Kommission Zukunft der Informationsinfrastruktur

1. Einleitung

Der Einsatz von Informations- und Kommunikationstechnologie (ICT) auf der Basis von ortsgebundenen Computern und mit mobilen Geräten (Laptops, Smartphones bzw. Tablets) ist aus der modernen Wissenschaft nicht mehr wegzudenken. Viele wissenschaftliche Methoden, wie z.B. Fernerkundung, Feldmessungen, Laborexperimente und Computersimulationen erzeugen große Datenmengen (Big Data). Erhobene Daten müssen gespeichert, organisiert und ausgewertet werden. Dies bringt neue wissenschaftliche Herausforderungen mit sich, die von Jim Gray als das „*vierte Paradigma*“ für die Wissenschaft eingeführt wurden (Hey et al. 2009). Dabei geht es um ICT-basierte Methoden, die bisher verborgene oder neue Zusammenhänge aus ständig wachsenden Datenmengen sichtbar machen und insgesamt die Qualität in der Wissenschaft steigern. In diesem Rahmen hat sich auch der Begriff „e-Science“ (electronic oder enhanced science) eingebürgert, der eine teamorientierte Forschung auf Basis von digitalen Infrastrukturen beschreibt und neben den technologischen Methoden (z.B. Virtualisierung und Cloud Computing) auch einen sozioökonomischen und wissenschaftspolitischen Aspekt aufweist (Taylor 2007; Yang et al. 2013). International werden Programme zum Aufbau von e-Science-Infrastrukturen in Forschung und Entwicklung politisch und finanziell gefördert.

Projekte in der ökologischen Langzeitforschung (engl. long-term ecological research, im Folgenden als L-TER bezeichnet) sind aufgrund ihrer Untersuchungsobjekte (z.B. Lebewesen, Habitate, Stoffkreisläufe oder Energieflüsse) und deren Wechselwirkungen untereinander komplexe Vorhaben, die ein großes Spektrum an wissenschaftlichen Disziplinen integrieren. In der Praxis arbeiten interdisziplinäre Forschungsgruppen verschiedener Institutionen zusammen an gemeinsamen Projekten – oft örtlich verteilt und in internationalen Kooperationen. Für ökologische Langzeitforschungen sammeln Ökologen, Ingenieure und andere Fachleute durch das Monitoring von Umweltparametern Messdaten von zahlreichen Messstationen über lange Zeiträume. Der dabei schnell anwachsende Datenbestand sollte allen Projektpartnern zur Datenaufbereitung, Auswertung, Forschung und für zukünftige Fragestellungen nachhaltig (Pillmann und Gnauck 2012) zur Verfügung stehen. Rechtliche Grundlagen, wie z.B. das Urheberrecht, müssen beachtet werden.

Dabei ist zu berücksichtigen, dass sich die Forschungsphilosophien gewandelt haben. Traditionell waren die eingesetzten ICT-Systeme meist Eigenentwicklungen (Ruchter 2002) und die Forschungsergebnisse unterlagen oft der Geheimhaltung (Thurnherr 2003). Mit zunehmender Globalisierung und der Agenda 21 von Rio sind die Öffnung des Nutzerkreises und die Pflicht zur Information der Bürger von Umweltdaten einhergegangen (Conference on Environment and Development und United Nations Conference on Environment and Development 1992). Handelt es sich um öffentliche Projekte, so sollten die aufbereiteten Daten und Forschungsergebnisse auch für

Entscheidungsträger aus der Politik und für Bürger zielgruppenorientiert zugänglich sein (Ruchter 2002).

Die Vielfältigkeit an Daten und Beteiligten führte nicht nur zu diversen einzelnen Datenbanken und unterschiedlichen Datenformaten, sondern auch zu räumlich und zeitlich nicht vergleichbaren Datensätzen, was die Weiterverarbeitung der Daten deutlich erschwert (Molkenthin 2000). Für Umweltprojekte wird daher eine moderne Informations- und Kommunikationsstruktur benötigt, die sich stark an einen individuellen Workflow der Benutzer anpassen lässt, um die gewonnenen Daten zu organisieren, vergleichbar zu machen und auszuwerten (Wohlgemuth 2009; Pillmann et al. 2011; Gnauck 2011). Hier kommt das Konzept der virtuellen Forschungsumgebungen (engl. virtual research environments, im Folgenden als VRE bezeichnet) zum Einsatz, welches Kreativität, Effektivität und Produktivität in der Forschung steigern soll (Fraser 2005).

Eine Herausforderung in der ökologischen Langzeitforschung besteht darin, die Daten langfristig persistent speichern zu können und wiederauffindbar zu archivieren. Das ist angesichts sich ständig ändernder Hard- und Software-Plattformen mit immer neuen Standards und Technologien problematisch. Zum einen müssen spezielle Dateiformate, Soft- und Hardwaretechnologien, sowie passende Softwareprodukte gewählt werden, die eine auf die Zukunft ausgerichtete Kompatibilität besitzen. Zum anderen muss der Datenbestand dauerhaft gepflegt (z.B. aktualisiert oder strukturiert), auf dem Server gegen Internetangriffe geschützt und durch ein organisiertes Backup vor Datenverlust geschützt werden. Dazu ist zu beachten, dass die Daten in L-TER Projekten oft über Generationen von Wissenschaftlern hinweg verwendet werden sollen. Dies erfordert eine gute Dokumentation des gesamten ICT-Systems und der vorhandenen Daten, so dass nachfolgende Generationen die Daten nutzen können (Nachnutzungseffekt). Zum Wiederauffinden der Daten wird ein Metadaten-Schema benötigt, welches die Daten charakteristisch beschreibt. Internationale Metadaten Standards stehen dazu zur Verfügung (Lazinger und Tibbo 2001), doch bei individuellen Datenstrukturen müssen die Metadaten flexibel an die Datentypen angepasst werden können. Auch die Eingabe und Übernahme der Metadaten ist zu betrachten. Viele ältere Datensätze haben keine Metadaten oder orientieren sich nicht an den nationalen und internationalen Metadatenstandards. Eine nachträgliche Anpassung erfordert viel Arbeitsaufwand in der Datenpflege oder der ICT-Struktur. Bei neueren Datensätzen muss ebenfalls der zusätzliche Aufwand zur Erhebung der Metadaten einkalkuliert werden, der vielen Wissenschaftlern als lästig erscheint und daher auch einen sozioökonomischen Faktor darstellt. Eine teilautomatische, ICT-gestützte Erzeugung von Metadaten ist möglich. Jedoch liegt die Schwierigkeit darin, diese genau auf das jeweilige eingesetzte System abzustimmen. Ein derartiges System im sich schnell entwickelnden ICT-Sektor muss ständig von Administratoren gepflegt und mit Updates aktualisiert werden. Es befindet sich während der gesamten Lebenszeit daher in einem Anpassungszyklus um die Hard- und Softwarekompatibilität für die Zukunft zu gewährleisten, Funktionalitäten zu verbessern und gegebenenfalls zu erweitern.

Die verschiedenen Philosophien bei der Software-Entwicklung einer derartigen virtuellen Forschungsumgebung haben Auswirkung auf das Endprodukt und damit auf das gesamte Forschungsprojekt. Prinzipiell können VREs in Eigenentwicklung realisiert, als proprietäre Software eingekauft oder aus Open-Source-Modulen aufgebaut werden. In der Vergangenheit wurden viele Projekte in Eigenentwicklung durchgeführt (vgl. Ludwig und Enke 2013) da die auf dem Markt befindlichen Produkte in der Regel teuer in der Anschaffung (bei kostenpflichtigen Lizenzen) und meist nicht flexibel genug für den jeweiligen Anwendungsfall sind (vgl. Candela et al. 2013, Carusi und Reimer 2010). Die Ergebnisse der Eigenentwicklung waren oft nicht zufriedenstellend, leicht erweiterbar oder gar kompatibel mit anderen Systemen (vgl. Hamburger 2016). Die so entstandenen und durch eine lange Entwicklungszeit kostspieligen „Insellösungen“ waren oft schon bei der Fertigstellung nicht mehr auf dem neusten Stand der Technik und nicht ohne erheblichen Aufwand auf andere Projekte übertragbar, wie Gespräche mit Fachleuten auf Konferenzen und Workshops bestätigten. Der Wirkungskreis war daher trotz hoher Entwicklungskosten relativ eingeschränkt. Einen anderen Weg geht die Open-Source-Philosophie (Free Software Foundation 1996), die versucht, Software als „Ökosysteme“ (vergleichbar zu wirtschaftlichen Ökosystemen) zu entwickeln (Janzen et al. 2013), um effiziente Softwareentwicklung durch die Nutzung von Synergien zu ermöglichen. Open-Source-Software ist nach dieser Philosophie mit offenem Quellcode frei erhältlich. Das Open GIS Konsortium (OGC) (Open Geospatial Consortium 2017) koordiniert seit 1994 zum Beispiel Open-Source-Software und Standards im GIS Bereich. Doch der Open-Source-Weg bereitet neben den Vorteilen ebenfalls Probleme, wie z.B. Sicherheitsrisiken durch eine weit verbreitete Quellcodebasis und damit im Zusammenhang ständiger „Updatezwang“ zur Aufrechterhaltung des Betriebes oder der Kompatibilität einer Vielzahl von Komponenten und deren Versionen. Meist findet sich keine Software, die „alles“ kann. So müssen verschiedene Komponenten ausgesucht und in ein System integriert werden, die den jeweiligen Anforderungen entsprechen und zueinander kompatibel sind. Wichtig bei der Verwendung unterschiedlicher Open-Source-Komponenten ist die Gestaltung einer einheitlichen Benutzeroberfläche (engl. user interface, kurz UI), so dass die gesamte VRE aus unterschiedlichen Komponenten wie aus „einem Guss“ erscheint. Gleiches gilt für die Benutzerauthentifizierung für den kontrollierten, differenzierten Zugang zu den Daten. Hier ist es von zentraler Bedeutung für die Akzeptanz der Benutzer, dass diese sich bei den verschiedenen Diensten und Funktionen im Portal nur einmalig mit einem zentralen Account anmelden müssen (Single Sign-On), der gegebenenfalls mit dem globalen Account der jeweiligen Institution gekoppelt ist. Ebenfalls ist für die Akzeptanz besonders von Bedeutung, dass trotz des hohen Komplexitätsgrades einer derartigen VRE, die Nutzung möglichst einfach, benutzerfreundlich und intuitiv ist. Dies gilt auch für die Datensuche sowie für die Darstellung von Rohdaten und aufbereiteten Daten. Eine intelligente Suche und automatisierte Berichte (engl. Reports) mit Datenvisualisierung, die zudem gezielten Fragestellungen in „menschlicher Syntax“ verarbeiten, könnten die Benutzerfreundlichkeit darüber hinaus noch steigern.

Ist ein interoperabler Austausch von Daten und Forschungsergebnissen mit anderen nationalen oder internationalen Netzwerken erforderlich, müssen entsprechende Schnittstellen bereitgestellt und genutzt werden, die je nach Disziplin und eingesetzter Software unterschiedlich sowie komplex ausfallen können. Es empfiehlt sich der Einsatz von automatisierten Data-Harvesting-Strategien, die den Austausch von Metadaten und Datensätzen mit anderen Servern ermöglichen (Rusch-Feja 2002). Eine geeignete Infrastruktur auf der Basis von standardisierten Harvesting-Protokollen ist durch die Vorgaben der INSPIRE-Direktive in Form der Geodateninfrastruktur Deutschland (GDI) auf EU-Ebene vorhanden (vgl. Kapitel 2.6).

Die Kommunikation der Projektpartner untereinander und der Austausch (engl. data exchange) bzw. das Teilen von Daten (engl. data sharing) sind ebenfalls von zentraler Bedeutung für schnellere und bessere Forschungsergebnisse. Dabei gewinnt das *Data-* bzw. *Information Sharing* gegenüber dem *Data-* bzw. *Information Exchange* zunehmend an Bedeutung, da im Gegensatz zur lokalen Speicherung (bei Austausch) die Speicherung mit Hilfe von *Cloud-Technologien* (vgl. Kapitel 2.7.2) zentralisiert wird und somit keine Synchronisation erforderlich ist. Oftmals reichen zur Kommunikation konventionelle Kommunikationsmedien, wie Telefon, Post und E-Mail nicht mehr für eine effiziente Arbeit im Team aus. Es werden geeignete ICT- und e-Science-Infrastrukturen mit Datenbanksystemen, Cloud-Technologien und schnelle, interaktive Kommunikationsmöglichkeiten wie Wikis, Messenger, Telekonferenzsysteme und Umfragetools benötigt, die auch Work-Flow-Komponenten, die Dokumentenversionierung und Nachverfolgung von Änderungen beinhalten. Mit dem „mobilen Zeitalter“ und dem Durchbruch von Smartphones und Tablets ist es wichtig geworden, allgemein Informationen auch mobil zur Verfügung zu stellen. Dies bedarf in Bezug auf VREs einer Aufbereitung der Websites und der Daten in einer für mobile Endgeräte optimierten Darstellung und Technik unter Einbeziehung der mobilen Kommunikationsmöglichkeiten der Geräte. In diesem Zusammenhang sind soziale Netzwerkstrukturen wie bei „Facebook“ und „Twitter“ zu erwähnen, die genutzt werden können, oder sich als Vorlage und Maßstab für eigene Kommunikationstools mit Applikationen für Mobilgeräte (Apps) eignen.

Entscheidend für den Erfolg derartiger Projekte ist die solide Planung mit allen Beteiligten, um eine Auswahl der Grundfunktionen des Systems zu treffen und daraufhin eine Systemarchitektur mit den passenden Technologien und der richtigen Software zu entwickeln. So können folgenschwere Fehler besonders in der Anfangsphase eines Forschungsprojektes vermieden werden.

1.1 Motivation

Projekte im Fachgebiet „Environmental Informatics“ und des Lehrstuhls für „Ökosysteme und Umweltinformatik“ der „BTU Cottbus-Senftenberg“, sowie zahlreiche Expertengespräche mit Konferenz- und Workshop Teilnehmern im Umweltsektor haben

gezeigt, dass geeignete e-Science- bzw. ICT-Infrastrukturen für Umweltprojekte und insbesondere L-TER Projekte dringend benötigt werden, um Daten digital speichern, organisieren, austauschen bzw. teilen und auswerten zu können. Die an sich oft schon interdisziplinären Umweltprojekte erfordern für digitale Infrastrukturen zusätzlich fundiertes Know-how aus dem Fachgebiet der (Umwelt-) Informatik und ICT. Oft fehlte das Verständnis unter den einzelnen Fachdisziplinen für tiefgreifender Zusammenhänge, Technologien und Trends der jeweils anderen Disziplinen.

Einerseits waren im Umfeld des Autors an der BTU und in seinem allgemeinen Forschungsumfeld diverse Projekte mit einer digitalen Infrastruktur nur eingeschränkt erfolgreich z.B. aufgrund von Informationsdefiziten sowie Kommunikationsproblemen, bedingt durch zu viele Projektpartner in der Verbindung mit der Komplexität des Vorhabens sowie der ungewünschten Preisgabe von Know-how und urheberrechtlich geschütztem bzw. vertraulichen Daten. In diesen Beispielen war der Faktor „Mensch“ ausschlaggebend. Andererseits hatten Faktoren der Nachhaltigkeit und technische Faktoren Einfluss auf den Erfolg. Dies zeigte sich z.B. bei Projekten, die mit viel Geld- und Zeitaufwand aufgebaut wurden, aber auf ICT-Eigenentwicklungen mit „Insel-Charakter“ basierten. Diese stagnierten durch den Wegfall einzelner Softwareingenieure oder entsprachen durch eine lange Entwicklungsphase bei Fertigstellung schon nicht mehr dem internationalen Stand der Technik.

Im Jahr 2011 ging eine Anfrage des „Informations- und Kommunikationszentrums“ (IKMZ) der BTU Cottbus-Senftenberg am ehemaligen Lehrstuhl „Ökosysteme und Umweltinformatik“ ein, bezüglich eines Konzeptes für eine digitale Infrastruktur im Sonderforschungsbereich „SFB/TRR 38 - Hühnerwasser“. Für das seit 2005 laufende Projekt wurde im Rahmen eines Projektantrages eine Struktur zum Datenmanagement mit interoperablem Austausch von Forschungsdaten benötigt. Dabei stand eher die technische Komponente im Vordergrund – im Gegensatz zum Ansatz der VREs, bei denen die Wissenschaftler und Forschungsprozesse im Vordergrund stehen. Dieses war der Anlass zur Entwicklung eines nachhaltigen und schnell zu realisierenden L-TER VRE Konzeptes für digitale Forschungsgemeinschaften in Projekten der ökologischen Langzeitforschung und der vorliegenden Dissertation, welches unabhängig vom Projektantrag entwickelt wurde.

1.2 Zielstellung und Forschungsfragen

Das übergeordnete Ziel dieser Arbeit liegt darin, die Produktivität, Effektivität und Kreativität in der Forschung durch den Einsatz von VREs im Umweltbereich zu steigern. Den Schwerpunkt bildet die Entwicklung eines innovativen Baukastenkonzeptes für VREs zur Unterstützung von Projekten und Forschungsgemeinschaften in der ökologischen Langzeitforschung. Grundlage dafür sind Analysen von erforderlichen Voraussetzungen und Funktionalitäten für L-TER VREs. Das Konzept soll für eine schnelle Implementierung von Prototypen und dem schnellen Aufbau von Produktivumgebungen

geeignet sein. Eine Anwendung des Konzeptes auf drei verschiedene, reale Projekte ermöglicht eine Bewertung der Ergebnisse. Die Arbeit soll insbesondere Projekt- und Entscheidungsträger bei der Planung, dem Aufbau und Betrieb von VREs unterstützen, sowie den dazu erforderlichen wissenschaftlichen und technischen Hintergrund mit konkreten Empfehlungen zu untereinander kompatiblen „Best Practice Software“ aus dem Open-Source-Bereich liefern.

Folgende Fragestellungen werden in der Arbeit untersucht:

- Unter welchen Aspekten sollte eine L-TER VRE aufgebaut oder besser die gegebenenfalls bestehende Infrastruktur erweitert werden?
- Welche Voraussetzungen und Funktionalitäten benötigen VREs allgemein und wie unterscheiden sie sich im Vergleich zu L-TER VREs?
- Welche Kriterien spielen allgemein für den Erfolg von VREs eine Rolle?
- Welche technischen Standards sollten beachtet werden?
- Welche Systemarchitekturen, Technologien und Softwareprodukte sind geeignet?
- Welche Ressourcen, Services und Tools kommen als Bestandteile einer L-TER VRE in Frage?
- Wie sieht ein nachhaltiges „best-practice“-Konzept für L-TER VREs aus?
- Welche theoretischen und praktischen Probleme sind zu überwinden?
- Wie kann die Zeit für die Entwicklung von VRE Prototypen und die Administration im Produktivbetrieb (regulärer Betrieb des Systems) deutlich reduziert werden?

Diese Fragen werden in der vorliegenden Arbeit in verschiedenen Kapiteln beantwortet und in den Schlussfolgerungen zusammengefasst.

2. Stand des Wissens und der Technik

Die vorliegende Arbeit ordnet sich thematisch vorwiegend in der fachübergreifenden Disziplin der Ökologischen Informatik bzw. Umweltinformatik ein mit Anwendung auf dem Gebiet der Langzeitforschungsökologie (Müller et al. 2016). Die Informations- und Kommunikationstechnologien stammen größtenteils direkt aus der Disziplin der Informatik.

Im Folgenden wird der Stand des Wissens und der Technik diskutiert, der für den Entwurf, Aufbau und nachhaltigen Betrieb von VREs im Allgemeinen erforderlich ist. Dazu zählen die Leitlinien der datenbasierten Wissenschaft, die Grundlagen der Ökologischen Informatik, Umweltinformatik, L-TER-Forschung und Open-Source-Philosophie sowie konkrete Literatur mit Studien zu VREs. Bedeutende nationale, europäische und weitere internationale Aktivitäten sowie Technologien im Bereich von e-Science und VREs werden dazu aufgezeigt und kurz erläutert.

2.1 Alte und neue Paradigmen der Wissenschaft

Hohe Bedeutung für die Wissenschaft allgemein haben ihre Leitlinien und Paradigmen. Eine VRE für die Wissenschaft sollte prinzipiell auf diesen aufbauen und sie in Form von Basiseigenschaften verankern. Insbesondere die *gute wissenschaftliche Praxis* (Deutsche Forschungsgemeinschaft 1998, 2013) als Leitlinie im Umgang mit Forschungsdaten und das modernere „4. Paradigma“ der *datenbasierten Wissenschaft* (Hey 2009) sind zu beachten, da sie die Anforderungen an VREs und deren Ziele direkt betreffen.

Im Sinne einer guten wissenschaftlichen Praxis müssen folgende Aspekte beim Umgang mit Forschungsdaten berücksichtigt werden (Deutsche Forschungsgemeinschaft 1998, 2013):

- Alle Schritte müssen als Grundlage wissenschaftlicher Arbeit nachvollziehbar sein,
- Primärdaten müssen persistent und wieder auffindbar aufbewahrt werden,
- zur Datenhaltung ist ein fachspezifisches Organisationskonzept erforderlich,
- Daten müssen für die Öffentlichkeit einsichtbar sein (Open Access),
- Datenschutz, Urheberrechte und Zugriffsrechte müssen beachtet werden und
- die Speicherung sollte im Rahmen definierter Standards erfolgen.

Empirische Wissenschaften basieren häufig auf Beobachtungen zu theoretischen Annahmen über die Welt. Aus ihnen wird durch Modelle (Sarbadhikari und Srivastava 2017) versucht, Rückschlüsse auf die Zusammenhänge in der Realität zu ziehen. Daher wurden die drei klassischen Paradigmen der Wissenschaft formuliert (Birken 2011):

1. Empirischer Forschung (Eberhard 1999),
2. Mathematischen Modellbeschreibungen (Imboden et al. 2008),
3. Modelllösungen und Vergleich mit der (wahrgenommenen) Realität (Archer 1998),

hinzu kommt das „4. Paradigma“, welches durch die moderneren Möglichkeiten der ICT zur Jahrtausendwende entstanden ist,

4. Datenbasierte Wissenschaft (Hey et al. 2009).

Das klassische dritte Paradigma wurde nach Jim Gray (Hey et al. 2009) neu formuliert. Dabei tritt an die Stelle einer exakten Lösung eine approximierbare Lösung nach dem Konzept des *kritischen Realismus* (Archer 1998). Diese ist erst in den letzten Jahrzehnten bedingt durch drei Punkte aufgekommen (Birken 2011):

- Verfügbarkeit leistungsfähiger und bezahlbarer Rechner,
- Entwicklung und drastische Verbesserung numerischer Verfahren, mit denen mathematische Modelle auf Computern ausgewertet werden können,
- das Problem, dass die interessanten mathematischen Modelle mittlerweile zu komplex sind, als dass sie von Menschen in einer bewältigbaren Zeit berechnet werden könnten.

Die neue Formulierung des dritten Paradigmas ist nach Hey (2009) nicht genug. Daher etablierte sich das vierte Paradigma - der „datenbasierten Wissenschaft“ („*e-Science*“). So wie Computer es ermöglichen, von theoretischen Modellen auf numerische Modelle überzugehen, ermöglichen sie ebenso von rein empirischen Beobachtungen auf das digitale Sammeln von Daten mit anschließender Auswertung zu wechseln. Diese Aussage hat deshalb eine neue Qualität, weil mittlerweile weltweit Datenmengen gesammelt werden, die immense Größen im Zetta-Byte-Bereich (10^{21} Bytes oder eine Milliarde Terrabyte) (Reinsel et al. 2017) erreichen und außerdem zunehmend Programme und Geräte zur Auswertung verfügbar sind, mit denen diese Datenmengen nutzbar und zugänglich gemacht werden können. Dies kann z.B. durch eine einfache Form der Visualisierung am Bildschirm erfolgen, oder komplexer sein mit Methoden des *Datamining* (Runkler 2015) und der *künstlichen Intelligenz* (Lek und Guégan 2000; Russell et al. 2012). Beispiele für moderne Geräte sind: Die „CAVE“ (Pape 2001; Kim 2009), bei der 3D-Daten in einem echten Raum erfahrbar werden, eine VR-Datenbrille (Virtual-Reality) oder Augmented Reality (AR) mit eingeblendeten Daten z.B. in der Kameraansicht eines Smartphones (Furht 2011; Filetti 2014).

Das „4. Paradigma“ ist als Leitlinie für VREs insofern von Bedeutung, da VREs für die datenbasierte Wissenschaft ein besonders geeignetes Werkzeug darstellen.

Das e-Science-Konzept setzt sich in den einzelnen Teilbereichen fort. Dort entstehen Begriffsbildungen wie z.B. e-Geography, e-Learning, e-Medicine oder e-Engineering. Was man vor allem in Deutschland und Großbritannien unter dem Begriff e-Science versteht, deckt sich in den USA etwa dem Konzept der "cyberinfrastructure" und in Australien dem Konzept der "e-Research" (Candela et al. 2013).

Für die meisten Bereiche sind die Begriffe VRE, cyber- oder e-Infrastruktur Synonyme. Wenn es einen Unterschied zwischen den Begriffen gibt, dann wird er am ehesten im gesamtheitlichen Ansatz der VRE im Vergleich zu den eher technisch geprägten Begriffen der verschiedenen Infrastrukturen deutlich (Fraser 2005).

2.2 Entwicklung der Ökologischen Informatik

Die Ära der Ökologischen Informatik wurde in den 1950er und 1960er Jahren durch die verstärkte Umweltforschung im Bereich Ökosystemmodellierung / -simulation (Odum 1967) und die damals neuartigen Möglichkeiten der rechnergestützten Datenverarbeitung eingeleitet. So wurden zum Beispiel zur Analyse des Eutrophierungsgrades von Seen nichtlineare mathematische Modelle erstellt, die sich wesentlich schneller auf Rechnern lösen ließen, als von Hand (Sakamoto 1966). In den 1970er und 1980er Jahren wurden große Datenmengen an multivariaten Zeitreihen gesammelt (Seen, Flüsse, Wälder und Offenland), für die eine Zeitreihenanalyse mit Hilfe von Computern essentiell war (Jørgensen 1976). Auch die Lösung von Differentialgleichungen für Analyse, Modellierung und Vorhersage von Umweltparametern und ökosystemaren Prozessen benötigten immer mehr Rechenleistung von Computern. Mit steigendem Umweltbewusstsein in den 1980er / 1990er Jahren wurde auch die Umwelt- / Ökologische Informatik populär (Recknagel 2006). In den letzten Jahrzehnten konnten durch den rasanten Fortschritt in der Computertechnologie mit immer höherer Rechenleistung, Speicherkapazität, Parallelisierung und effizienten Algorithmen erhebliche Fortschritte in der Umwelt- und Ökosystemtheorie bei Analysen, Modellierung, Informationssystemen, Datenbanken und Entscheidungsunterstützungssystemen für transparente Entscheidungsprozesse erzielt werden (Recknagel 2006; Müller und Leupelt 1998). Ebenso ist der Fortschritt bei Sensortechnologien, dem Online Monitoring der Umwelt und der Fernerkundung zu erwähnen, der große Datenmengen (Big Data) mit sich bringt, welche organisiert und ausgewertet werden sollen. Wissensbasierte Expertensysteme können in der Ökologie von Nutzen sein, um Informationen und Wissen auf Grund von Expertenerfahrungen erfolgreicher zu verarbeiten (Puppe 1993). Der Vorteil liegt hier in der Verwaltung von qualitativem Wissen.

Der rechnergestützte Ansatz der Ökosystemanalyse, Ökosystemmodellierung und Ökosystemvorhersage wird als „Ökologische Informatik“ bezeichnet (Recknagel 2006). Dabei spielen viele Fachgebiete eine Rolle. Neben der Informatik, Mathematik und anderen Naturwissenschaften beschäftigt sich die Ökologische Informatik unter anderem mit Genen, ökologischen Netzwerken, Stoff- und Energiekreisläufen, Nahrungsket-

ten, ökologischer Nachhaltigkeit, Biodiversität und die Auswirkungen des Klimawandels (Jørgensen 1976; Straškraba und Gnauck 1983). Zu den Methoden zählen unter anderem zum Teil biologisch inspirierte Algorithmen und Technologien in Form von Fuzzy Logic, künstliche Intelligenz mit Neuronalen Netzen, evolutionäre oder genetische Algorithmen, zelluläre Automaten und agentenbasierte Systeme (Recknagel 2006). Die klassischen Methoden der Datenverarbeitung liefern Werkzeuge für die Datenarchivierung, Datengewinnung und Visualisierung - in Form von Datenbanken, Datenmodellierung, Programmiersprachen, (Geografische) Informationssysteme und viele weitere.

Die Ökologische Informatik ist in der internationalen Forschung als ein neues Forschungsgebiet der Ökologie anerkannt (Neumann 2003). Auf der dritten internationalen „Conference of Ecological Informatics“ (ISEI) 2002 in Rom verabschiedete man eine Definition, wonach die Ökologische Informatik ein interdisziplinäres Rahmenwerk darstellt, das auf Grundlage der neusten Informationstechnologien mit den Prinzipien der Informationsverarbeitung die Anwendung auf den unterschiedlichen Komplexitätsebenen eines Ökosystems fördert (Neumann 2003).

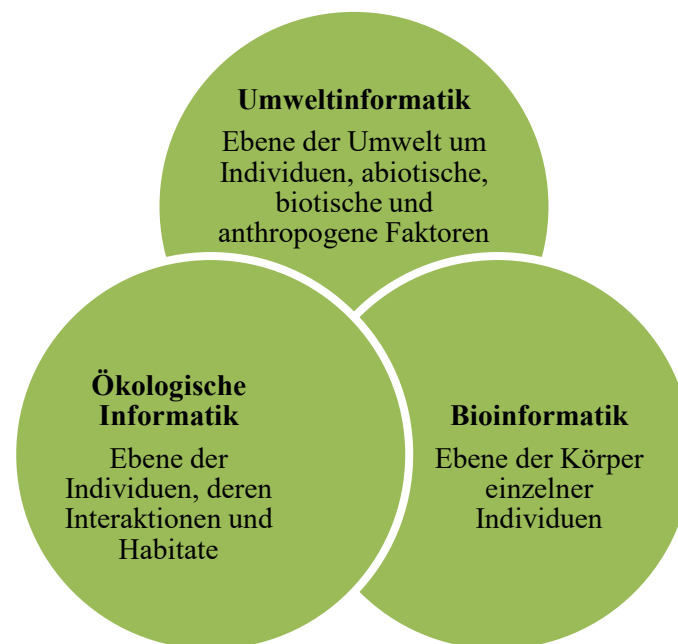


Abbildung 2.1: Schwerpunkte von Bio-, Umwelt- und Ökologischer Informatik mit Überschneidungen

Zu der Ökologischen Informatik kommen die Umweltinformatik und die Bioinformatik als Teilgebiete der Angewandten Informatik (Fuchs-Kittowski und Wohlgemuth 2011; Recknagel 2006). Diese drei Disziplinen überschneiden sich in vielen Teilbereichen (Abbildung 2.1). Das primäre Ziel der Umweltinformatik ist es durch Informationstechnologien die Umwelt zu untersuchen und Umweltbelastungen / Umweltschäden zu vermeiden, zu minimieren oder zu beseitigen (Page und Hilty 1995). Die Bioinformatik hingegen beschäftigt sich eher mit den Prozessen auf der Ebene der einzelnen

Individuen. Zum Beispiel zeigt Abbildung 2.2 zeigt die unterschiedlichen Skalenebenen im Übergang von der ökologischen Informatik zur Bioinformatik.

Die Interpretation von Beobachtungen auf unterschiedlichen Skalen wird durch die Konzepte der Hierarchietheorie in der Ökologie vereinfacht (Wiegand 1996). Sie ist hilfreich, um die hypothetische Wirklichkeit in Abhängigkeit von der Fragestellung in Beobachtungsebenen einzuteilen und damit einen geeigneten Maßstab für weitere Untersuchungen festzulegen. An diese Grundlagen knüpft das folgende Kapitel zur ökologischen Langzeitforschung an.

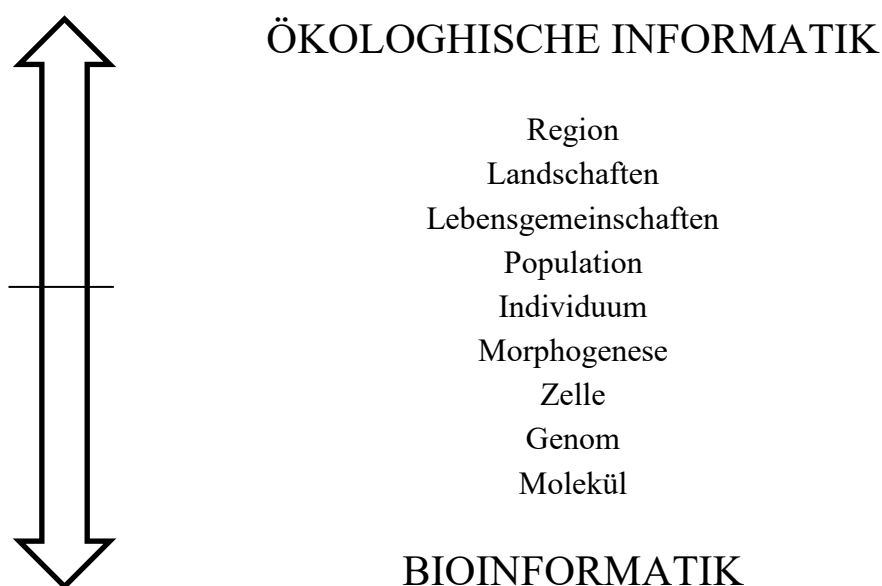


Abbildung 2.2: Skalenebenen in der Biologie und Ökologie mit Zugehörigkeit zur Bio-/ökologischen Informatik nach Neumann (2003)

2.3 Ökologische Langzeitforschung (L-TER)

Ökosysteme sind dynamische Systeme, die durch interne und externe Faktoren auf verschiedenen zeitlichen (Abbildung 2.3) und räumlichen Skalen beeinflusst werden. Bisher lag der Fokus bei der Erforschung von Ökosystemdynamiken meist auf der Betrachtung kurzer Zeiträume (oft bis zu drei Jahren), die typisch für finanziell geförderte Projekte und Studien waren (Müller et al. 2010a). Der Langzeitaspekt wurde dabei meist außer Acht gelassen, da kurzfristige Interessen von Politik und Wirtschaft im Vordergrund standen. Dieser Trend erzeugte nach Müller et al. (2010a) ein Informations- und Wissensdefizit in der ökologischen Langzeitforschung, deren Betrachtungszeitraum deutlich länger als drei Jahre ist. Einem nachhaltigen Ansatz folgt daher die Ökologische Langzeitforschung, in der weltweite Forschergruppen entstanden sind (Müller et al. 2010a; Pillmann und Gnauck 2012). In der EU Fauna-Flora-Habitat-Direktive (The council of the european communities 1992) und in der CBD (Conventi-

on on Biological Diversity 2017) wird das Ziel der „Biodiversität“ definiert, welches ein essenzielles Prinzip der Nachhaltigkeit darstellt. Dieser Aspekt findet sich beispielsweise in der EU-Wasserrahmenrichtlinie (Europäisches Parlament 2000) und bei „Natura 2000“ (Europäisches Parlament 2017). Darüber hinaus haben längerfristige Prozesse Einzug in unser Bewusstsein gefunden, wie z.B. der *Klimawandel* mit seinen Auswirkungen auf die Natur, den Menschen und seine Umwelt. Viele Umwelteinflüsse zeigen in einem Ökosystem träge Reaktionsgeschwindigkeiten, lange Wirkungsketten oder werden im System gepuffert. Dies drückt sich in Form der Resilienz gegenüber Störungen, der Stabilität und Anpassungsfähigkeit von Ökosystemen aus (Müller et al. 2010a). Langzeitbeobachtungen und spezielle Methoden für die Langzeituntersuchung sind daher unter anderem zur Einschätzung von Konsequenzen erforderlich, die durch den anthropogenen Eingriff in die Umwelt hervorgerufen werden. Der Schutz von Ökosystemen zielt langfristig darauf ab, dem Menschen in sozialen, ökologischen und ökonomischen Aspekten funktional zu „dienen“. Daraus entwickelte sich der Begriff der *Ökosystem Dienste* (engl. ecosystem services) (Müller et al. 2010a).

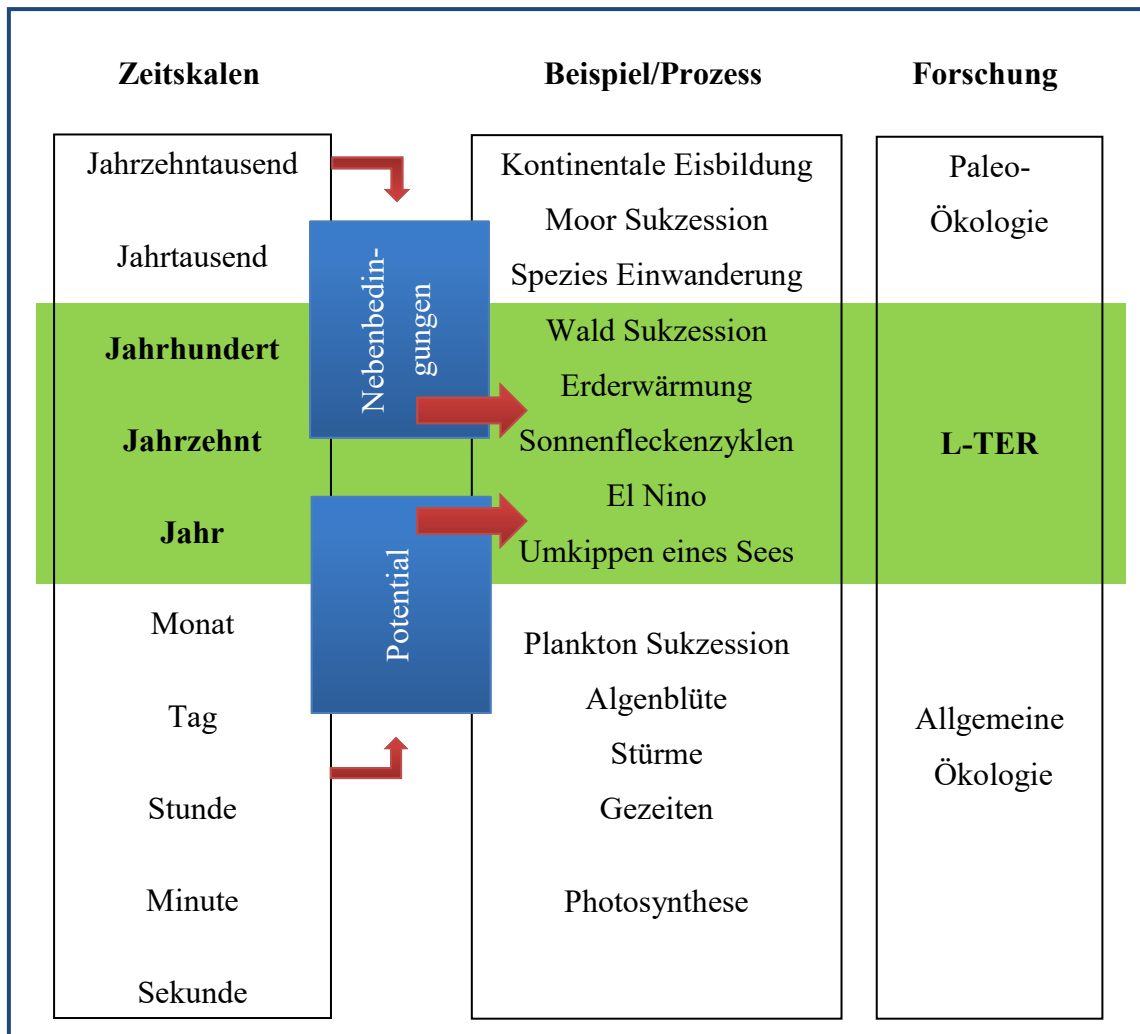


Abbildung 2.3: Unterschiedliche Zeitskalen von Prozessen und Forschungsbereichen nach Hobbie et al. (2003)

Die Analyse und Modellierung von Ökosystemen benötigen und erzeugen große Datenmengen. Daher gilt insbesondere das Management von großen Mengen an Forschungsdaten als Voraussetzung für die (ökologische) Langzeitforschung (Müller et al. 2010a). Virtuelle Forschungsumgebungen können dazu die erforderlichen Funktionen bereitstellen (Filetti und Gnauck 2011), wie in den folgenden Kapiteln dargestellt wird.

2.4 Open Source – von der Insellösung zum „Software-Ökosystem“

Open-Source-Software eignet sich besonders für den Aufbau nachhaltiger VREs. Wie die Arbeit und das folgende Konzept „Rapid L-TER VRE“ (vgl. Kapitel 4.1) zeigen werden, ist der Einsatz von Open-Source-Software ein wichtiger Baustein für den Erfolg einer VRE. Daher sollen die Philosophien dieser Softwareform näher erläutert werden.

Open-Source-Software ist ein Schlagwort, das in der ICT oft zusammen mit Kosteneinsparungen bzw. kostenloser („freier“) Software einseitig genannt wird. Die Philosophie von „Open-Source-Software“ (übersetzt: quelltextoffene Computerprogramme) umfasst mehr, als nur kostenlose Software. Das „Frei“ bezieht sich eher auf das Frei wie in „Redefreiheit“ nicht wie das „Frei“ in z.B. „Freibier“ (Stallmann 2017).

Unterschiedliche Interpretationen von „Frei“ und verschiedene Ansätze der Open-Source-Definition ziehen auch verschiedene Bezeichnungen dieser Art von Software nach sich. Daher werden häufig die Begriffe „Freie Software“ im Sinne von kostenloser Software oder „Open-Source-Software“ im Sinne von quelltextoffener Software verwendet, je nach Kontext. Generell meinen die beiden Begriffe zwar ähnliche Prinzipien, die dahinter verborgenen Philosophien unterscheiden sich jedoch in der Ausrichtung und Reichweite. Um einen möglichst übergeordneten Begriff für die Zielsetzungen der einzelnen Philosophien zu verwenden, hat sich die Verwendung von *Free-Libre-Open-Source-Software* (FLOSS) und *Free-Open-Source-Software* (FOSS) etabliert, mit dem man sich neutral zwischen „Freier Software“ und „Open-Source-Software“ ausdrücken kann (Stallmann 2017).

Im Wesentlichen gibt es nach Diedrich (2016) drei Institutionen, welche die Definition von Open Source prägen. An erster Stelle sei die von Bruce Perens (damals Debian-Projektleiter) und Eric S. Raymond 1998 gegründete, gemeinnützige Open-Source-Initiative (OSI) genannt (opensource.org 2017). Ihr Ziel als nicht kommerzielles Unternehmen ist es, weltweit die Open-Source-Entwicklung zu fördern und dazu Standards und Definitionen zu entwickeln. Die OSI versuchte erfolglos nach ihrer Gründung, „Open Source“ als Markenzeichen schützen zu lassen. So wäre sie zum „offiziellen Hüter“ der Open-Source-Definition (OSD) geworden (Diedrich 2016).

Die OSI gibt auf Ihrer Website (opensource.org 2017) unter ihrer Mission folgende Definition von Open Source:

“Open source is a development method for software that harnesses the power of distributed peer review and transparency of process. The promise of open source is better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in.”

Noch weiter geht die 1983 von Richard M. Stallmann gegründete „Free Software Foundation“ (fsf.org 2017). Die FSF propagiert weitreichende politische und gesellschaftliche Ansprüche in Verbindung mit Open Source. Es wird hier zusätzlich die Gefahr betont, die von „unfreier“ Software ausgehen könne, deren Quelltext nicht offen liegt und ggf. kostenpflichtig ist.

“The free software movement is one of the most successful social movements to emerge in the past 25 years, driven by a worldwide community of ethical programmers dedicated to the cause of freedom and sharing. But the ultimate success of the free software movement depends upon teaching our friends, neighbors and work colleagues about the danger of not having software freedom, about the danger of a society losing control over its computing.” (fsf.org 2017)

Im Wesentlichen meinen die beiden Organisationen (OSI und FSF) inhaltlich dasselbe. Im Vordergrund steht, dass Anwender mehr Rechte an der Software haben als bei proprietären Programmen (Diedrich 2016):

- das Recht, die Software nach Belieben weiterzugeben;
- das Recht, die Quelltexte zu erhalten;
- das Recht, die Software zu verändern und in veränderter Form weiterzugeben;
- das Recht, die Software für jeden Zweck einzusetzen.

Die Debian-Richtlinien (Linux Distribution) für Open Source als dritte Quelle von Definitionen für Open Source entsprechen ebenso diesen Kernaussagen. Nachfolgend wird ein Überblick zu den groben Debian Richtlinien (debian.org 2017) gegeben, die sich mit den Richtlinien der OSD decken:

1. Unbeschränkte Weitergabe
2. Quelltext Offenlegung
3. Abgeleitete Werke
4. Integrität des ursprünglichen Quelltextes
5. Keine Diskriminierung von Personen oder Gruppen
6. Keine Diskriminierung von bestimmten Einsatzbereichen
7. Weitergabe von Lizenzen
8. Keine Lizenzen für nur ein spezielles Produkt
9. Keine Auswirkungen / Einschränkungen für andere Software
10. Technologie-Neutralität

Aus den nachfolgend vorgestellten Leitlinien für die Definition von FLOSS gehen zahlreiche Lizenz-Modelle hervor, die im Einzelnen die Verwendung für private und, kommerzielle Nutzung regeln. Diese sind z.B. GPL und LGPL (gnu.org 2017), BSD-License (frebsd.org 2017), Apache License (apache.org 2017), Mozilla Public License (mozilla.org 2017). Auf die verschiedenen Lizenz-Modelle soll im Rahmen dieser Arbeit nicht weiter eingegangen werden.

Aus der von der Europäischen Kommission in Auftrag gegebenen Studie über den ökonomischen Einfluss von FLOSS im europäischen ICT Sektor (United Nations University 2006) geht hervor, dass FLOSS einen erheblichen Einfluss auf die Industrie, Wirtschaft, Haushalte und Öffentlichkeit nicht nur in Europa hat. Dabei wird FLOSS nach unterschiedlichen Kategorien unterschieden, unter anderem auch nach Webserver Software und Betriebssystemen, die wegen ihrer zentralen Bedeutung einen besonderen Stellenwert einnehmen. Außerdem sollte diese Art von Software mit Spenden für wohltätige Zwecke gleichgesetzt werden, so die Autoren der Studie.

Erfolgreiche Open-Source-Software kann zur Entstehung eines dynamischen „Software-Ökosystems“ beitragen. Es beschreibt das Zusammenspiel zwischen Organisationen und Unternehmen auf einem geteilten Markt für Softwareentwicklung und ICT-Services ähnlich eines wirtschaftlichen Ökosystems (Jansen und Brinkkemper 2013). Oft bildet eine gemeinsame technologische Plattform die Basis für Kooperation und wirtschaftliche Aktivitäten.

Open-Source-Software bietet ein großes Potential - auch für den Einsatz im wissenschaftlichen Bereich und damit meist im öffentlichen Sektor, indem Umwelt- und ökologische Daten der Öffentlichkeit prinzipiell zugänglich gemacht werden müssen (Open Data) (Conference on Environment and Development und United Nations Conference on Environment and Development 1992). Es gibt jedoch auch Ausnahmen, in denen die Forschung nicht öffentlich stattfindet, z.B. bei sicherheitsrelevanten Projekten, die der Geheimhaltung unterliegen (Thurnherr 2003), wie z.B. Umweltauswirkungen einer neuen Kriegswaffe.

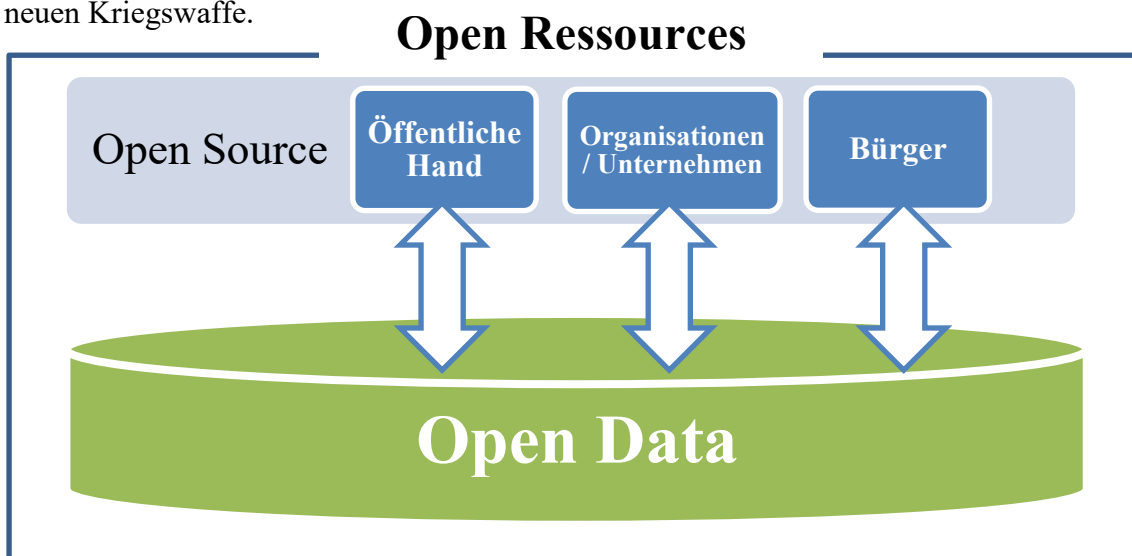


Abbildung 2.4: Durch den Einsatz von Open Source und Open Data ergeben sich Open Ressources nach Fritzlar et al. (2012)

Analog zu den Begriffen „Open Source“ und „Open Data“ kann ein System aus flexiblen und nicht-exklusiven Kapazitäten als „Open Resources“ bezeichnet werden, wie Abbildung 2.4 darstellt (Fritzlar et al. 2012). Es schließt prinzipiell keinen von der Mitarbeit aus. Leistungen in diesem System können entweder auf freiwilliger Basis (Überzeugung) stattfinden oder von wirtschaftlichem Interesse angetrieben sein. Ziel ist es, die Kompetenzen effektiv zu bündeln und auf weitaus größere Ressourcen zurückzugreifen, als sie das eigene Umfeld bietet.

Durch den Einsatz von quelloffener Software können Kosten und Entwicklungszeit gesenkt werden (Fritzlar et al. 2012; Europäische Kommission 2006). Die Abhängigkeit des öffentlichen Sektors von proprietärer Software und deren Herstellern wird gleichzeitig verringert, da FLOSS frei genutzt, verbreitet, kopiert und weiterentwickelt werden kann. Eine Anpassung der Software an die individuellen Bedürfnisse ist durch den zugänglichen Quellcode möglich. Die FLOSS-Produkte, die aus den oftmals weltweiten Entwicklergemeinden entstehen, entsprechen immer mehr den Qualitätsstandards - z.B. BSI-Sicherheitsstandard (Breitenstrom et al. 2013). Doch können sich auch Nachteile aus dem Einsatz von FLOSS ergeben, wie z.B. mangelnde Kompatibilität, rechtliche Aspekte (z.B. Gewährleistung), oder in deren Anwendung und Akzeptanz (Fritzlar et al. 2012).

Tabelle A. im Anhang A zeigt Vor- und Nachteile für den Einsatz von Open Source im öffentlichen Sektor nach verschiedenen Aspekten gegliedert.

Obwohl in den letzten Jahren gerade die Systeme auf der Basis von Open-Source-Software einen technischen Stand erreicht haben, der eine ausgereifte Entwicklung komplexer VREs und vergleichbarer Software zulässt, wurden im Rahmen dieser Arbeit weder unter den Enterprise-Produkten, noch unter den Open-Source-Projekten ein System gefunden, das allen Projektanforderungen gerecht würde oder beliebig skaliert werden kann. Dies geht aus der Auswertung der einschlägigen wissenschaftlichen Literatur zu VREs hervor (vgl. Kapitel 2.5.1 und 2.5.5). Internetrecherchen zum Thema „VRE“ & „Framework“ bestätigten dies (vgl. Kapitel 2.5.5).

Vermutlich wird es dies auch in der nächsten Zeit aufgrund der Vielfältigkeit der Anforderungen und Funktionen, sowie der rasanten Geschwindigkeit beim Fortschritt der ICT-Entwicklung nicht geben (vgl. Kapitel 2.5). Einzelne Teilbereiche entwickeln sich mitunter so schnell, dass spezielles Know-how erforderlich wird, Module nicht mehr kompatibel sind und die Struktur immer komplexer wird.

2.5 Virtuelle Forschungsumgebungen

Virtuelle Forschungsumgebungen sind nach Fraser (2005) innovative, webbasierte Arbeitsumgebungen, die gruppenorientiert, übergreifend, sicher und flexibel sind. Sie sind kein abgeschlossenes Produkt, sondern ein (vielfach mehrstufiges) anpassungsfähig-

ges und erweiterbares Framework zur Unterstützung des gesamten Forschungsprozesses oder Teilen davon. VREs bestehen aus Ressourcen, VRE-Services und VRE-Tools (vgl. Kapitel 3.1). Ressourcen sind in diesem Zusammenhang sowohl die Forschungsprimärdaten, die Ergebnisse der Forschung als auch die Personen selbst. Die Services stellen selbständige Dienste (z.B. Cloud-Service) zur Unterstützung der VRE dar, können multidisziplinär oder fachspezifisch sein und sollten sich in eine Gesamtarchitektur eingliedern. *Serviceorientierte Architekturen* (SOA) unterstützt die Einbindung der Services indem verteilte Dienste durch verschiedene Domänen bereitgestellt werden können (Elsayed 2012). VRE-Tools gliedern sich stärker in die VRE ein (nicht eigenständig) und bieten Werkzeuge (Funktionen) für Arbeitsprozesse in der Forschung, wie z.B. die Datenaufbereitung. Sie sollten auf offenen Standards basieren (Fraser 2005).

Virtuelle Forschungsumgebungen lassen sich folgenden charakterisieren. Sie sind:

1. interoperable, webbasierte Systeme auf der Basis von Webstandards (Llewellyn Smith et al. 2011).
2. individuelle und maßgeschneiderte Systeme für die praktischen Aufgaben und Anforderungen einer dynamischen Gemeinschaft (engl. community of practice) (Lave und Wenger 1991).
3. gesamtheitliche und vollständige Systeme, die in der Praxis alle Materialien und Werkzeugen bieten, um die Ziele der Gemeinschaft zu erreichen (Fraser 2005).
4. offene und flexible Systeme in Bezug auf die angebotenen Services und ihre Lebenszeit. Sie benötigen eine fein abstimmbare gruppenorientierte Zugangskontrolle und Aufbereitung der Daten zu, Zwischen- und Endergebnissen mit Parametern und Metadaten, wie z.B. Datenherkunft, Bearbeiter und Nutzungsrechte (Fraser 2005).

2.5.1 Aktueller Stand

Der aktuelle Stand in der Forschung zu VREs wurde in der Arbeit von Candela et al. (2013) untersucht, die als Basis der weiteren Untersuchung diente. In der Studie „Knowledge, networks and nations. Global scientific collaboration in the 21st century“ der Royal Society in Kooperation mit Elsevier wird bestätigt, dass wissenschaftliches Arbeiten zunehmend global, multipolar und vernetzt erfolgt (Llewellyn Smith et al. 2011). Die Studie umfasst 12 Länder (USA, UK, Deutschland, Japan, Frankreich, Kanada, Italien, Niederlande, Australien, Schweiz, China, Spanien) sowie weitere zusammengefasste Länder. Dabei wurden Tausende von Arbeiten und Autoren aus unterschiedlichsten Disziplinen berücksichtigt. Parameter waren unter anderem: Die wachsende Zahl der Beiträge mit mehr als einem Autor, Art und Weise der Zusammenarbeit, eingesetzte Technologien der Kommunikation, nationale und internationale Kooperationen und Netzwerke. Dieser Trend zeigt auf, dass innovative, dynamische und universel-

le Web-Umgebungen genutzt werden, welche die Wissenschaft bei den neuen Herausforderungen adäquat unterstützen.

Je nach Kontext haben diese „Umgebungen“ verschiedene Namensgebungen. Allgemein werden sie als *Virtuelle Forschungsumgebungen* (Carusi und Reimer 2010), *Science Gateways* (Wilkins-Diehr 2007), *Collaboratories* (Wulf 1993), *Digitale Bibliotheken* (Candela et al. 2011) oder *Inhabited Information Spaces* (Snowdon 2004) bezeichnet. Diese Begriffe gliedern sich ein in die *e-Infrastructures* (e-Infrastructure Reflection Group 2010) und die *cyberinfrastructures* (Cyberinfrastructure Council 2007), welche die „Umgebungen“ eher gesamtheitlich betrachten. Teilweise überschneiden sich diese Begriffe auch mit *grid-computing* (Foster et al. 2009) (Rechenpower des Verbundes steht im Vordergrund) und *virtual organizations* (Foster und Kesselman 1999) sowie mit *cloud-computing* (Foster et al. 2009), bei dem abstrakte IT-Infrastrukturen dynamisch auf verteilten Systemen bereit stehen. Die meisten Definitionen von VREs haben als Charakteristik eine webbasierte Arbeitsumgebung gemeinsam (Süptitz et al. 2013). Diese sei jedoch nicht ausreichend für die Funktionen einer VRE. Ebenso würden mindestens Web 2.0 Technologien (O'Reilly 2007) und Soziale Netzwerke (Wang et al. 2007) erforderlich, um der Wissenschaft eine zentrale und umfassende Umgebung für Teamarbeit und Ressourcennutzung zu geben.

Carusi und Reimer (2010) geben in ihrer Studie einen umfassenden Bericht zum Stand von VREs in Großbritannien. An erster Stelle steht ihre Erkenntnis, dass VREs nicht nur technologische Projekte sind, sondern insbesondere dazu geeignet, um Forschungsgemeinschaften (communities of science) und weltweite Netzwerke zu bilden. Hierüber besteht eine große Übereinstimmung in den verschiedenen Definitionen der Eigenschaften von VREs. Diese Communities sind virtueller Natur und folgen nicht unbedingt formalen Strukturen, wie z.B. Abteilungen oder Projekt-Teams (Lave und Wenger 1991; Carusi und Reimer 2010). Die Struktur bildet sich eher in den Köpfen der Teilnehmer und wird sichtbar durch die interaktiven Aktionen in einer VRE, z.B. bei der Lösung eines ganz bestimmten Problems. Dies wird dann deutlich, wenn die Teilnehmer das Wissen teilen, das für die Lösung eines ganz bestimmten Problems notwendig ist (Wenger 2000). Dabei sollte eine Einigung in Form eines „Interaction Agreements“ über die Art und Weise der Zusammenarbeit bestehen.

Eine weitere Eigenschaft von VREs besteht darin, ganzheitlich alle erforderlichen Werkzeuge zur Bewältigung der insgesamt erforderlichen Aufgaben für den Forschungsprozess bereitzustellen. Ein Forschungsprozess ist die chronologische Abfolge der einzelnen wissenschaftlichen Arbeitsschritte bis zum Forschungsergebnis (vgl. Ritschel et al. 2016). Dazu müssen nicht nur die einzelnen Werkzeuge bereitgestellt werden, sondern auch die Eingliederung in ein zentrales einheitliches System mit Layout, Funktionalität, Workflows und Schnittstellen ist erforderlich. Hier spielen zentrale Logins und Management von Benutzern sowie der Daten- und Informationsaustausch zwischen den einzelnen Komponenten und Nutzern eine besondere Rolle. Auch die Vielfalt an Ressourcen steigt unter diesem gesamtheitlichen Aspekt, wie z.B. Services für Datenreihen, Statistiken, Objektsammlungen, Reporting Tools, Speicher- und Re-

chenpower. Dabei entwickeln sich die Ergebnisse der traditionellen Forschung - wie z.B. klassische Forschungspublikationen und experimentelle Daten - hin zu *lebenden Berichten* (Candela et al. 2007; Candela et al. 2005), *ausführbaren* oder *interaktiven Ressourcen* (Nowakowski et al. 2011; van Gorp und Mazanek 2011), *wissenschaftlichen Workflows* (Roure et al. 2009) und *erweiterten Publikationen* (Hoogerwerf et al. 2013). Darüber hinaus sollten VREs eine vereinfachte und visualisierte Übersicht zu den Ressourcen aus verschiedenen Quellen ermöglichen. VREs müssen den Zugriff auf Daten regeln und zur Übersichtlichkeit gruppenorientiert mit ausreichend abstimmbaren Zugriffsrechten ausgestattet sein. Eine Trennung von Zwischen- und Endergebnissen ermöglicht mehr Übersichtlichkeit. Dabei sind Urheberrechte zu wahren und die Datenherkunft muss klar erkennbar sein (Ludwig und Enke 2013). Viele Wissenschaftler weigern sich, ihre Daten zur Verfügung zu stellen, wenn keine Mechanismen zum Zugriffsschutz, Urheberrecht und zur Sicherheit existieren (Roure et al. 2009). Diese Mechanismen können *explizit* sein, wie z.B. die erforderlichen Rechte zur Anzeige und Bearbeitung von Daten oder *implizit* in Form eines Frameworks, das Metadaten zur Herkunft bei der Ausgabe von Daten einfügt (Candela et al. 2013). Auch die inhaltliche und visuelle Aufbereitung der Daten für verschiedene Nutzergruppen ist erforderlich, um beispielsweise die wissenschaftlichen Erkenntnisse in allgemein verständlicher Form für Entscheidungsträger und die Öffentlichkeit bereitzustellen (Ruchter 2002).

Über den Zugriff auf wissenschaftliche Ergebnisse veröffentlichte das Research Councils UK (Research Councils UK 2005) ein Positionspapier ähnlich zu den Richtlinien des DFGs zur „guten Wissenschaftlichen Praxis“ (Deutsche Forschungsgemeinschaft 1998, 2013). Daraus geht hervor, dass sich ein nachhaltiges Datenmanagement zur Nachvollziehbarkeit in der Wissenschaft tiefgreifend in Großbritannien etabliert hat. Von besonderer Bedeutung ist der aus den USA stammende Atkins-Report (Atkins und Cyberinfrastruktur 2003). In dieser umfangreichen Studie geht es um das Cyberinfrastrukturprogramm der USA. Hier wird auf die Herausforderungen und die Strukturen für VREs eingegangen, sowie die Zielstellungen des Programmes erläutern. Die Publikation der National Science Foundation - Cyberinfrastruktur Council (USA) (National Science Foundation 2007) beschreibt die Herausforderungen und den Einsatz von Cyberinfrastrukturen im 21. Jahrhundert anhand vieler praxisorientierten Beispiele aus der Wissenschaft. Der Report eines Workshops für Cyberinfrastruktur und Sozialwissenschaften (Berman et al. 2005) betrachtet VREs zusätzlich von der soziotechnischen Seite. In einem Artikel von Clifford Lynch (Lynch 2008) werden die Herausforderungen für Lehre und Forschung in Cyberinfrastrukturen beschrieben. Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen werden von Buddenbohm et al. (2014) zusammengestellt. Dabei wird zunächst auf die verschiedenen Lebensphasen einer VRE eingegangen. Im Weiteren folgt für den Erfolg von VREs eine Zusammenstellung von Kriterien in Bezug auf die verschiedenen beteiligten Gruppen, wie Benutzer, Betreiber und Förderer. Die Arbeit zeigt zudem verschiedene Finanzierungsmöglichkeiten für den Aufbau und Betrieb von VREs.

In der Literatur finden sich nur wenige konkrete Leitfäden für die Entwicklung von VREs und ebenso für die technische Implementation. Stellvertretend für einen deutschsprachigen Leitfaden sei hier das vom BMBF geförderte Projekt einer „VRE Blaupause“ für den Aufbau der „DARIAH-DE“ Forschungsinfrastrukturen erwähnt (Munson et al. 2016). Darin wird kurz und knapp auf die Grundlagen von VREs eingegangen. Anhand der verschiedenen Phasen bei der Entstehung (Entscheidungsfindung, Vorbereitungsphase, Entwicklungs- und Produktionsphase) wird vor allem für Entscheidungsträger, Designer und Entwickler ein Wegweiser für VREs präsentiert.

EU-Projekte, wie DILIGENT (DILIGENT 2017), D4Science und D4Science-II (D4Science 2017) basieren auf Infrastrukturen, die großen Kapazitäten für Ressourcen mit viel Rechenleistung und Servern bereitstellen. Sie gehören zu den ersten, die auch Cloud-Technologien für die Erstellung von VREs verwenden (Candela et al. 2010). Dabei kommen Software-Frameworks und -Pakete für das Management der Ressourcen zum Einsatz, sowie softwaregestützte Analysen für individuell benötigte VRE-Eigenschaften. Die automatisierte, individuelle Erstellung und Bereitstellung von VREs erfolgt über „Generatoren“ (Assante et al. 2008).

Die Studie von Candela et al. (2013) geht davon aus, dass der oben beschriebene Trend in den nächsten Jahren fortgesetzt wird und VREs sich als Standard bei der Unterstützung wissenschaftlicher Aktivitäten weiter etablieren. Dabei werden sie zum Rückgrat der Organisationsstruktur für eine kooperative und sozial vernetzte *Forschungsgemeinschaft der Praxis* mit großem Potential zur Steigerung der Effizienz und Effektivität des wissenschaftlichen Arbeitens und darüber hinaus informativer für Entscheidungsträger und Öffentlichkeit. Forschungsgemeinschaften sind diesem Kontext abstrakter anzusehen als konkrete Forschungsprojekte und können sich z.B. in ihren Zielen, der zeitlich Ausdehnung und von den Mitgliedern her unterscheiden.

Die Studie von Candela et al. (2013) beschreibt weiterhin den Trend aufgegliedert in die verschiedenen Phasen einer VRE, der die Richtung in der Entwicklung von modernen VREs anzeigt:

In der Planungsphase wird das Grundsystem die Entwickler bei der Erstellung einer VRE und der Definition ihrer individuell erforderlichen Eigenschaften durch geeignete Planungssoftware unterstützen.

Die Implementation und das Management von VREs wird ein zielgerichteter Prozess sein, der auf verlässlichen Services und offenen Standards basiert. Aufgrund ihrer Komplexität werden VREs nicht sofort in vollem Umfang bereit stehen. Sie werden sich aus bereits bestehenden Infrastrukturen und Systemen entwickeln, die in einem globalen VRE-Management-Service zusammengeführt werden. Dieser Service wird von der Konzeption einer VRE (Allan et al. 2004), über die Bereitstellung und der Inhaltspflege bis hin zur Wartung und dem Monitoring des Systems gesamtheitlich alle erforderlichen Funktionen anbieten. Dabei wird sich der Prozess stetig über einen längeren Zeitraum und in einem iterativen Vorgehen vollziehen. Bei der Implementation des Systems helfen Mechanismen, die geeignete Datenstrukturen aufbauen, die Services verknüpfen,

die gesamte softwareseitige ICT-Infrastruktur generieren und bereitstellen. Dazu wird es vordefinierte Anwendungsszenarien für VREs zur Auswahl geben, die jeweils grundlegenden Eigenschaften, Module, Komponenten und Grundkonfigurationen einer VRE erzeugen.

Die Phase der Bereitstellung von VRE-Systemen könnte automatisch ablaufen. Die Management-Services werden dabei hinsichtlich ihrer Qualität und des Budgets optimiert werden. Inhaltspflege und Monitoring wird weiterhin vornehmlich durch menschliche Kontrolle ausgeführt aber durch die maschinellen Services zunehmend unterstützt werden. Die daraus resultierenden VREs werden sehr flexibel und dynamisch anpassbar sein. Jeder Benutzer wird dann seine eigenen Workflows für die Lösung eines wissenschaftlichen Problems definieren, kopieren und mit anderen Teilen können. Somit werden VREs nicht nur durch technologische sondern zunehmend durch soziale und organisatorische Prozesse weiterentwickelt.

Der Trend geht hin zu effizienteren VREs, die Arbeitserleichterung für Benutzer bieten oder im technischen Sinne (z.B. Speicherkapazität, Rechenleistung), in denen die Wiederverwendung von Daten, Modulen und Workflows ermöglicht wird. Dabei können z.B. interaktive Methoden zur Datenaufbereitung in einem Modell oder Workflow wiederverwendbar abgespeichert werden. Dies stellt eine erhebliche Zeitersparnis dar - besonders dann, wenn ähnliche Prozesse aus vorangegangenen Prozessen neu abgebildet werden und nur wenige Parameter angepasst werden müssen (Holl 2014). Der wahrscheinlich bedeutendste Trend sind allgemeine VRE Frameworks, die als Basis zur Entwicklung von verschiedenen VREs zum Einsatz kommen können (Carusi und Reimer 2010).

Die derzeitigen Herausforderungen und zukünftigen Forschungsschwerpunkte von VREs werden von Candela et al. (2013) zusammengefasst und erläutert. Diese liegen vor allem bei einer zunehmenden Automatisierung von Workflows in der VRE selbst, bei der Entwicklung von VRE Frameworks und der automatisierten Bereitstellung von einsatzbereiten VREs as a Service (VREaaS), sowie bei der Integration in den Arbeitsalltag - auch durch geeignete Data Policies.

Aus dem vorgestellten Stand der Technik zu VREs mit prognostizierten Trends aus der Literatur wird ersichtlich, welche Anforderungen an VREs bestehen und welche Funktionalitäten erforderlich sind. Derzeit existiert jedoch kein VRE-Software-Framework, das „out-of-the-box“ sofort einsatzfähig oder für alle Anforderungen geeignet wäre („one size fit them all“) (vgl. Kapitel 2.4). Es kann davon ausgegangen werden, dass es dies auch in der nächsten Zeit aufgrund der Vielfältigkeit der Anforderungen und Eigenschaften, sowie der rasanten Geschwindigkeit beim Fortschritt der ICT-Entwicklung nicht geben wird. Die Trends zeigen auf, das zukünftig noch viel Bedarf an Forschung zu VREs besteht. Insbesondere effizient arbeitende VRE-Frameworks mit Konfigurationsassistenten, automatisierter Bereitstellung und automatisierten Workflows sind zu nennen. Die vorliegende Arbeit soll hier anknüpfen und eine Basis für nachhaltige VRE-Frameworks im Bereich von L-TER entwickeln.

2.5.2 Management von Forschungsdaten

Das digitale Management von Forschungsdaten ist ein zentraler Bestandteil von VREs. Daher wird an dieser Stelle auf einige Leitfäden zum Forschungsdatenmanagement aus der deutschen Literatur eingegangen. Besonders ist das WISSGRID-Projekt zu erwähnen, dass von 2010 bis 2012 vom BMBF im Rahmen der D-Grid Initiative gefördert wurde (Ludwig und Enke 2013). Ähnliche Ansätze findet man auch in den GESIS-Leitlinien (Jensen 2012) zum Management von Forschungsdaten, die sich inhaltlich aber eher mit sozialwissenschaftlichen Inhalten beschäftigen.

Das Management von Forschungsdaten umfasst alle Maßnahmen, die sicherstellen, dass digitale und ggf. analoge Forschungsdaten nutzbar sind. Was im Einzelnen dazu zählt, unterscheidet sich teilweise sehr stark durch die unterschiedlichen Ziele und Zwecke der Nutzung. Prinzipiell lassen sich nach Ludwig und Enke (2013) vier Arten von Zwecken unterscheiden:

1. die Nutzung als Arbeitskopie für das wissenschaftliche Arbeiten,
2. die Nachnutzung von Forschungsdaten für spätere Forschung,
3. die Aufbewahrung als Dokumentation des durchgeführten wissenschaftlichen Arbeitens und
4. die Aufbewahrung, um rechtlichen oder anderen forschungsfremden Anforderungen nachzukommen.

Als Grundvoraussetzung steht die *Bitstream Preservation* (Spisla 2014), also die Erhaltung der Bitfolge der Rohdaten. Dies kann angesichts der Kurzlebigkeit und Fehleranfälligkeit von Datenträgern und ICT-Systemen besonders bei Forschungsdaten mit einem größeren Volumen schon eine Herausforderung an sich sein. Der Erhalt der Bitfolge alleine sichert noch nicht die technische Nachnutzbarkeit der Daten. Durch spezialisierte und proprietäre Dateiformate wird ein nachhaltiger Zugriff auf Daten erschwert. Dies begründet sich zum einen durch schnelle Innovationszyklen in der ICT mit denen sich die Standards und Kompatibilitäten ändern, zum anderen durch proprietäre Software und Dateiformate, die an Unternehmen gebunden sind. Auch wenn die technische Nutzbarkeit gegeben ist, bedarf es zur inhaltlichen Nutzung der Forschungsdaten immer Kontextwissen und eine nachvollziehbare Dokumentation der Daten sowie die Beschreibung mit Metadaten (Ludwig und Enke 2013).

Neben den übergreifenden Aufgaben, gibt es nach Ludwig und Enke (2013) weitere Bereiche (Abbildung 2.5), die von zentraler Bedeutung sind. Auf diese soll im Folgenden kurz eingegangen werden.

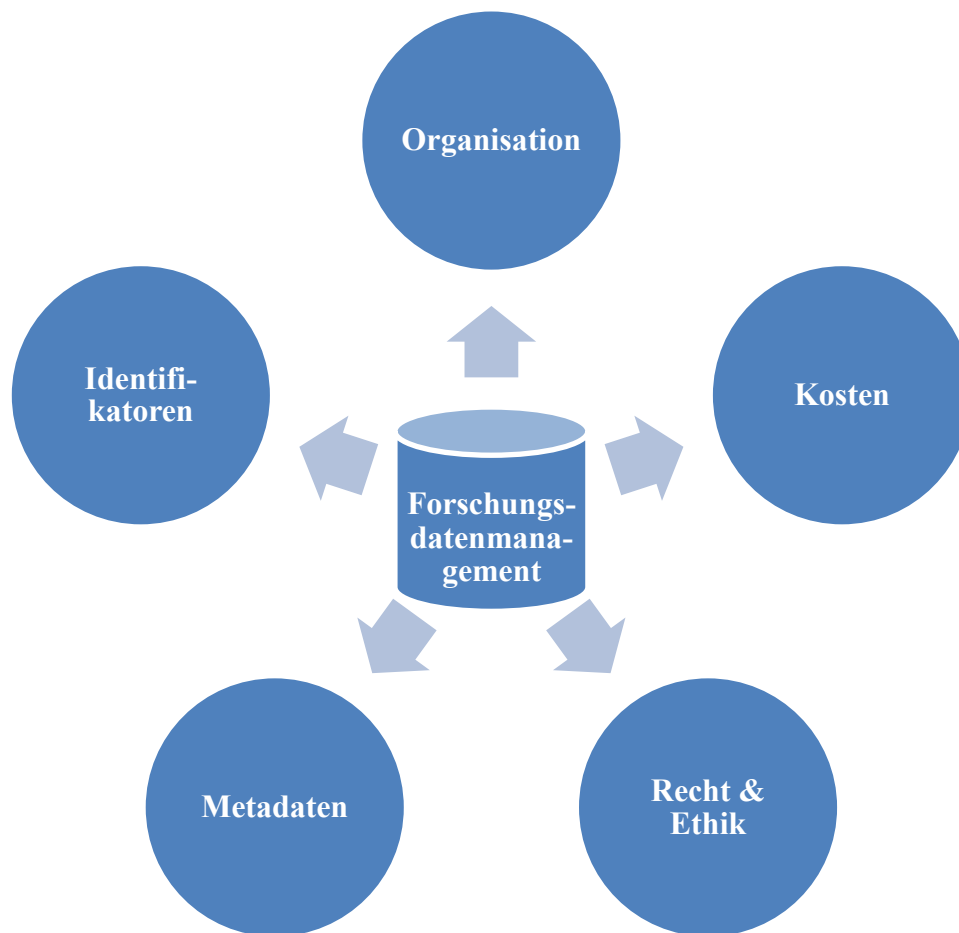


Abbildung 2.5: Übergreifende Bereiche des Managements von Forschungsdaten nach Ludwig und Enke (2013)

- **Organisation:** Für die Aufbewahrung von Forschungsdaten kann keine Technik oder Software verantwortlich gemacht werden. Die Verantwortung muss immer eine Person oder Institution tragen. Dazu müssen Prozesse definiert und dokumentiert werden, um langfristig sicherzustellen, dass alle erforderlichen Maßnahmen für die Nutzbarkeit der Forschungsdaten getroffen werden. Ob es sich dabei um ein fachspezifisches Datenarchiv, ein Konsortium oder beispielsweise um ein Universitätsrechenzentrum handelt, ist hierbei weniger von Bedeutung. Dazu müssen auch klare Regeln in Form von „Policies“ aufgestellt werden, die festlegen, wie die Nutzer mit dem System und den Daten umzugehen haben – wobei auch dafür gesorgt werden muss, dass diese eingehalten werden.
- **Kosten:** Finanzen sind normalerweise der limitierende Faktor, wenn es um ICT-Systeme für Forschungsdaten geht. Meist ist zu Anfang eine große Summe für Planung und Implementierung der Systeme erforderlich. Der finanzielle Aufwand zur nachhaltigen Nutzung der Daten kann hoch sein und ist hinsichtlich der Kostenentwicklung und der Gesamtkosten bei einer zeitlich unbefristeten Archivierung schwierig zu bewerten.
- **Recht und Ethik:** Forschung und Wissenschaft muss sich an geltendes nationales und ggf. internationales Recht halten. Besonders sollte auf den Schutz von

personenbezogenen Daten, Urheberrechte, Lizenzierungen und vertrauliche Daten geachtet werden. Der Vollständigkeit halber sein hier auch ethische und kulturelle Aspekte erwähnt, die es zu beachten gilt.

- **Metadaten:** Für die Nutzung von Forschungsdaten sind strukturierte semantische Informationen über die vorliegenden Daten in Form von Metadaten in jeder Phase des Lebenszyklus unverzichtbar. Für ICT-Systeme sollten die Metadaten digital zu verarbeiten und zu pflegen sein und nach Möglichkeit bis zu einem gewissen Grad automatisch erzeugt werden können.
- **Identifikatoren:** Der Einsatz von dauerhaften Identifikatoren (Persistent Identifier) (Schroeder 2001, 2004; Neuroth et al. 2012) für Daten und Dokumente sollte beim Datenmanagement nicht vernachlässigt werden. Identifikatoren dienen dazu, ein Objekt eindeutig anhand beispielsweise einer Nummer oder Zeichenfolge zu identifizieren, wie z.B. die ISBN bei Büchern. In der Forschung und Wissenschaft bewährt hat sich für diesen Zweck beispielsweise die Verwendung des sogenannten DOI (Digital Object Identifier) (Brase und Klump 2007), der sich auch zur bibliographischen Zitation sehr gut eignet.

2.5.3 Life-Cycle-Modell für VREs

Betreiber von VREs benötigen funktionierende und effiziente Verfahren zum Aufbau und Betrieb. Zur Unterstützung dieser Prozesse wird für VREs ein Lebensphasenmodell (Buddenbohm et al. 2014) mit Parallelen zur klassischen Softwareentwicklung (Tarr und Wolf 2011) vorgeschlagen, das auf die erfolgskritischen Punkte für einen Übergang in den nachhaltigen Betrieb besonders eingeht. Dieses Modell basiert auf Arbeitsgemeinschaften, Literaturrecherchen, Expertengesprächen und Workshops. Es konzentriert sich insbesondere auf die Sicht der Betreiber und Förderer von VREs. Das Lebensphasenmodell ist ein bedeutendes Element bei der Planung, der Entwicklung und dem nachhaltigen Betrieb von VREs, da es die Rahmenbedingungen zeitlich strukturieren kann. Dabei ist die frühe Einbindung der beteiligten Nutzergruppen entscheidend für den Erfolg einer VRE.

Hinsichtlich der Lebensphasen sind zwei Situationen zu unterscheiden (Abbildung 2.6). Für VREs, die als eigenständige Infrastrukturen betrieben werden, ist das hier beschriebene Schema sequentiell aufgebaut. Besteht die VRE jedoch aus verschiedenen unabhängigen Modulen, gelten die Lebensphasen für das jeweilige Modul. Die Überführung einzelner Module in der Transferphase erfolgt dann nicht zwingend in eine externe Struktur, sondern die Module können zur Wiederverwendung in der VRE genutzt werden. Das detaillierte Lebensphasenmodell einer VRE ist in Abbildung 2.7 ersichtlich.

Lebensphasen einer VRE

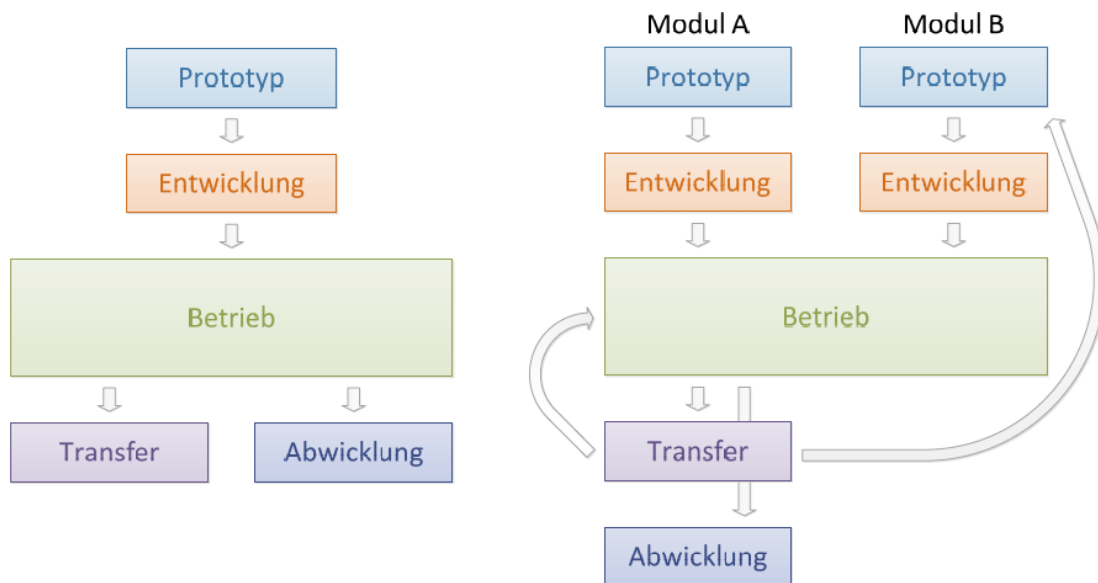


Abbildung 2.6: Lebensphasen einer VRE und deren Module (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0

Der Lebenszyklus einer VRE lässt sich in die fünf Phasen Aufbau, Entwicklung, Betrieb, Transfer und Abwicklung aufteilen. Sie können sowohl in den Anforderungen an die VRE, als auch in externen Parametern, wie beispielsweise der Finanzierung, deutlich variieren. Dabei muss jedoch beachtet werden, dass eine schematische Darstellung immer eine idealisierte Situation beschreibt, und daher das vorgestellte Schema nicht exakt auf bestehende VREs übertragbar ist.

Die schematische Darstellung der fünf Phasen einer VRE und ihrer einzelnen Module zeigt Abbildung 2.7. In der Aufbauphase werden die Rahmenbedingungen für die Struktur der VRE festgelegt. Dazu gehören Vorarbeiten, organisatorische und technische Spezifikationen und deren Umsetzung, welche zur Bereitstellung eines Prototyps für die Benutzer führt. Die Realisierbarkeit des Systems soll damit unter Beweis gestellt werden und eine Evaluierung unter Betriebsbedingungen kann erfolgen. Auch die Akzeptanz der VRE innerhalb ihrer Nutzergruppe lässt sich hier einschätzen.

Der Prototyp ist die Basis für die folgende Entwicklungsphase. Sie verfeinert das Konzept mit Aspekten, die in der Aufbauphase nicht ausreichend bearbeitet werden können. Dazu zählen ein Betriebskonzept, indem VRE-Dienste und die dafür benötigten Ressourcen gegenüber gestellt werden, die Organisationsform (inklusive Rechtsform), ein Kosten- und Finanzierungsplan, sowie ein Nachnutzungskonzept. Danach handelt es sich bei der VRE um ein ausgereiftes Produkt, das in die Betriebsphase übergeht.

In der Betriebsphase steht die VRE den Nutzern in einer ausgereiften Form für ihre wissenschaftliche Arbeit zur Verfügung. Von Seiten der Betreiber liegt von nun an der Schwerpunkt auf der Betreuung der Nutzer, der Administration des Systems und auf der fortdauernden technischen und organisatorischen Weiterentwicklung der VRE.

Kommt es zum Zeitpunkt, den Betrieb der VRE einzustellen, dient die Transferphase dazu, erhaltenswerte Komponenten wie beispielsweise Forschungsdaten oder Programmcode in andere (Infra-)Strukturen zu überführen (z.B. Archive). Verbleibenden technischen und organisatorischen Strukturen werden parallel dazu in der Abwicklungsphase letztendlich aufgelöst.

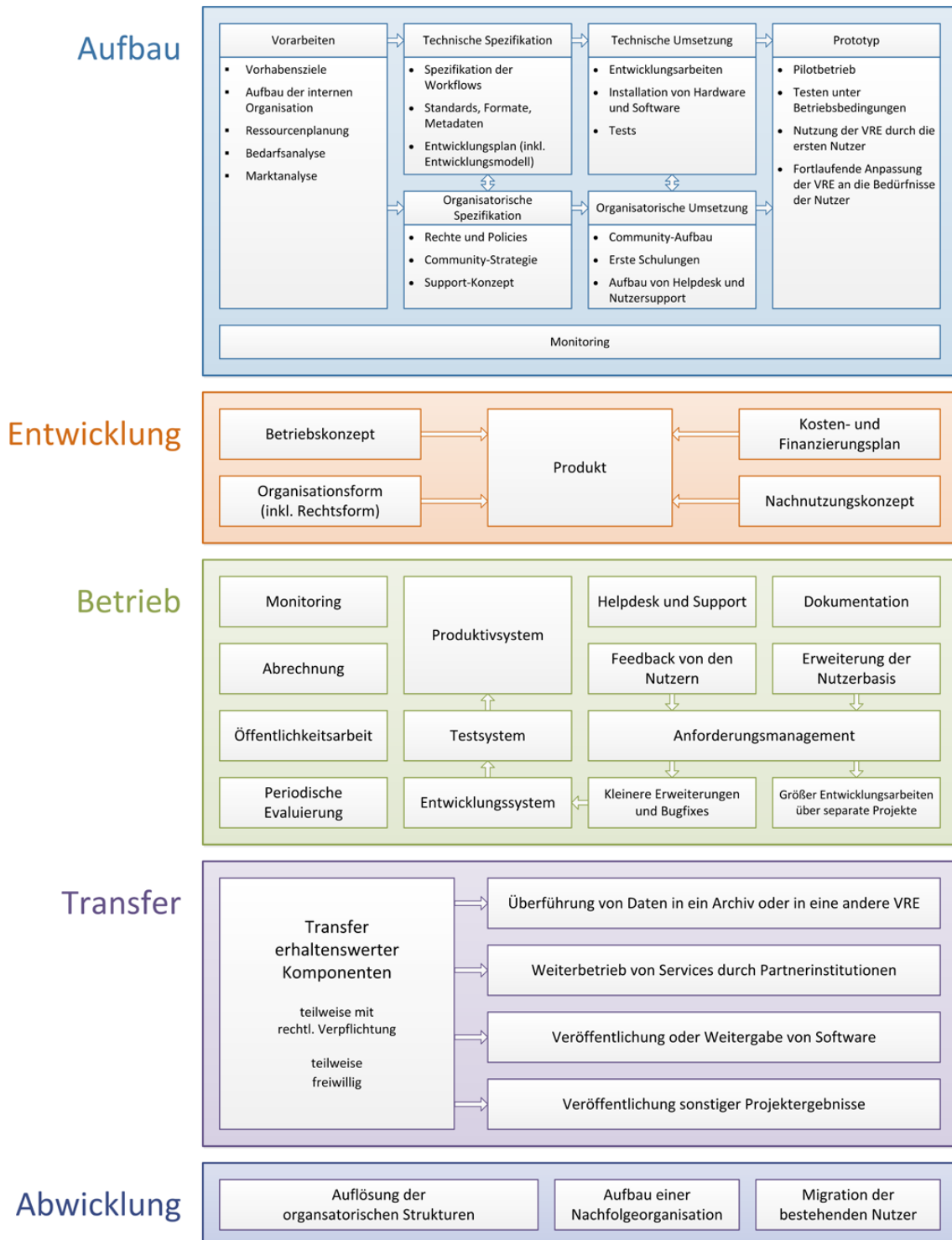


Abbildung 2.7: Lebensphasen einer VRE im Detail (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0

2.5.4 Langzeitarchivierung innerhalb VREs

Bislang gibt es keine international anerkannte Definition von Langzeitarchivierung (Neuroth et al. 2012). Wie lange Forschungsdaten genutzt werden, ist vom jeweiligen Einsatzzweck und von den rechtlichen Rahmenbedingungen abhängig.

Wie verschieden z.B. die Auslegung von „Langzeit-Datenhaltung“ unterschiedlicher Berufsgruppen und Bereichen sein können, zeigt die folgende Tabelle 2.1. Sie dient als Anhaltspunkt für die Abschätzung der zeitlichen Anforderungen für allgemeine Projekte. Meistens sind die zeitlichen Vorgaben gesetzlich, vertraglich oder anderweitig schriftlich für das jeweilige Projekt geregelt.

Tabelle 2.1: Unterschiedliche Auslegungen der Langzeit-Datenhaltung

Berufsgruppe / Bereich	Zeitraum
Informatiker (Neuroth et al. 2012)	> 5 Jahre
Deutsche Forschungsgemeinschaft (DFG) (Deutsche Forschungsgemeinschaft 2009)	> 10 Jahre
Sonderforschungsbereich (SFB) (Watzke 2017)	> 12 Jahre
Ingenieure (Neuroth et al. 2012)	> 30 Jahre
Meteorologen, Linguisten (Neuroth et al. 2012)	> 100 Jahre

Um eine möglichst lange Nutzbarkeit von Forschungsdaten zu garantieren, ist ein Management über die gesamte Lebensdauer der Daten unerlässlich (Ludwig und Enke 2013; Neuroth et al. 2009). Im Kontext der Forschung geht es darum, Forschungsinhalte in Form von Forschungsdaten und Publikationen langfristig digital zur Verfügung zu stellen, wobei die Inhalte verifizierbar, nachnutzbar und interpretierbar sein müssen, damit sie für zukünftige Forschungen von Nutzen sind (Neuroth et al. 2012). Dies spiegelt sich auch in den Richtlinien der DFG zur „guten wissenschaftlichen Praxis“ wider (vgl. Kapitel 2.1).

Bei der Langzeitarchivierung können die Daten als physisches Objekt, als logisches Objekt und als konzeptuelles Objekt gespeichert werden (Thibodeau 2002). Bei der Speicherung des physischen Objekts werden einfach die Bits und Bytes auf einem Datenträger ohne eine nähere Beschreibung des Objekts durch Morphologie, Syntax, oder Semantik (Metadaten) gespeichert. Die Bitstream-Preservation (oder Bitstrom-Speicherung) gilt als Basis der Langzeitarchivierung. Diese Bitstream-Preservation kann nach den folgenden Vorgaben des Kooperationsverbundes „nestor“ – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland – erfolgen (Baumann 2014):

1. **Redundante Datenhaltung:** Von den Daten sollten mehrfache Kopien existieren. Um die Kopien vor äußeren Einflüssen zu schützen, empfiehlt sich auch eine räumlich getrennte Aufbewahrung der Kopien.
2. **Diversität eingesetzter Speichertechnologien:** Die Sicherung der Daten sollte auf mindestens zwei unterschiedlichen Datenträgertypen erfolgen.
3. **Standards:** Die Speichermedien sollten internationalen Standards entsprechen und eine weite Verbreitung aufweisen.
4. **Datenträger:** Verwendete Speichertechnologien und Datenträger müssen regelmäßig durch neue ersetzt werden (Spisla 2014).

Für die Datensicherheit allgemein ist das Risiko abzuwiegen, welches sich aus Schadensumfang und Eintrittswahrscheinlichkeit für verloren gegangene Daten berechnet. Daraus lassen sich mit den entsprechenden Ressourcen und finanziellen Mitteln geeignete Strategien für die Speicherung und für das Backup der Daten auf verschiedenen Sicherheitsniveaus entwickeln. Eine Strategie ist es, möglichst viele „Snapshots“ von unterschiedlichen Zeitpunkten (Archiv) und aktuelle Backups von gesamten (virtuellen) Maschinen (als Container-Datei), Dateisystemen und Datenbanken an verteilten Orten redundant aufzubewahren.

Aktuelle Entwicklungen:

Mit Methoden zur „Ultra-Langzeit-Datenspeicherung“ können digitale Daten in DNA-Molekülen abgelegt werden (Grass et al. 2015). Bei Kühlung der DNA-kodierten Information auf minus 18 Grad sollen die Daten über einer Million Jahre erhalten bleiben können. Pro Gramm DNA lassen sich theoretisch Datenmengen bis in den Petabyte-Bereich persistent speichern. Beschädigungen der Information werden durch neue Algorithmen für die Fehlerkorrektur besser repariert, wie z.B. durch den Einsatz von Reed-Solomon-Codes (Gries 1986; Karve 2017). Die eigentlichen Daten werden dabei vor der Synthese des DNA-Strangs zusätzlich mit untereinander abhängigen Zusatzinformationen verknüpft. Durch äußere und inneren Code-Wörtern sollen nicht nur einzelne Fehler innerhalb von Teilsegmenten, sondern auch Ausfälle kompletter DNA-Segmente kompensiert werden (Grass et al. 2015). Das Verfahren eignet sich jedoch nicht für relativ schnelle Schreib- und Lesezugriffe, wie sie bei herkömmlichen Festplatten und im Arbeitsspeicher eines Computers im Bereich von Millisekunden benötigt werden. Im Vergleich haben die herkömmlichen Speichermedien zur Langzeitarchivierung digitaler Daten, wie DVDs oder Blu-rays, eine Lebensdauer von maximal 100 Jahren. Mikrofilme verlieren ihre Datenspeicherfunktion nach etwa 500 Jahren (Spisla 2014).

Für die Langzeitarchivierung von Daten in VREs sollte daher ein klarer zeitlicher Rahmen definiert und das Risiko für den Datenverlust sowie das langfristig verfügbare Budget einkalkuliert werden. Daraus lässt sich das Maß an geeigneten Strategien einschätzen. Dazu zählen z.B. die mehrfach redundante Datenhaltung auf verschiedenen

standardisierten Speicherformaten und die Archivierung von Formatbeschreibungen sowie Metadaten. Auch die Sicherung der gesamten Infrastruktur als „lauffähige“ Umgebung in einer *Virtuellen Maschine* (vgl. Kapitel 4.3) kann dazu beitragen. Wie weit der betriebene Aufwand dabei gehen sollte, hängt neben dem dafür erforderlichen Budget unter anderem vom definierten Zeitraum, dem zukünftigen Entwicklungsfort- bzw. Rückschritt und der gewünschten Datensicherheit ab. Für die „Ultra-Langzeit-Datenspeicherung“ von Massendaten könnte sich in Zukunft die Archivierung (Lesezugriff) mit DNA-Molekülen eignen, wenn den Wert der Daten die derzeit noch nicht kalkulierbaren Kosten für das Verfahren rechtfertigen.

Auch der Einsatz von *Blockchain-Technologien* (Lee und Deng 2018) im Bereich von „e-Science“ ist denkbar (Ozturan 2016; Lin et al. 2017), wie sie z.B. bei der Cyberwährung „Bitcoin“ zum Einsatz kommen. Dabei handelt es sich um die kryptografisch abgesicherte Verkettung einzelner Datenblöcke (Konst 2000), die in *verteilten Datenbank-Management-Systemen* (Nakamoto 2008) („relativ“) manipulations- und ausfallsicher gespeichert werden. Diese könnten genutzt werden, um z.B. Forschungsdaten und deren Bearbeitungsschritte als Forschungsprozess über die gesamte Lebensdauer hinweg persistent, d.h. manipulations- (gewollt oder ungewollt) und ausfallsicher mit kryptografisch abgesicherter Verkettung in verteilten Datenbank-Management-Systemen mit verschiedenen Zugriffsrechten zu speichern. Daraus könnte eine ganz neue Qualität von transparenten und nachvollziehbaren Forschungsergebnissen hervorgehen. Aus der Literatur gingen jedoch keine Quellen hervor, die konkret auf einen bisherigen Einsatz der Blockchain-Technologie für die persistente Langzeitdatenspeicherung in VREs hinweisen.

2.5.5 VRE Frameworks

Der wahrscheinlich bedeutendste Trend ist die Entwicklung von allgemeinen VRE Frameworks, die als Basis zum Aufbau von verschiedenen VREs verwendet werden können (Carusi und Reimer 2010). Diese Frameworks bieten allgemeine Basis-Services (wie z.B. Authentifizierung, Rechtemanagement, Datenmanagement, Projektorganisation und Community-Tools) und ermöglichen zusätzlich eine individuelle Anpassung. Derzeit finden sich jedoch nur wenige konkrete VRE-Frameworks mit den beschriebenen Eigenschaften (vgl. Kapitel 2.5.1), zu denen „eSciDoc“ (escidoc.org 2016) und „gCube“ (gcube-system.org 2017b) gehören.

„eSciDoc“ wurde im Rahmen der e-Science-Initiative des Bundesministeriums für Bildung und Forschung (BMBF) als gemeinsames Projekt vom FIZ Karlsruhe (Leibniz-Institut für Informationsinfrastruktur) und der Max-Planck-Gesellschaft von Mitte 2004 bis Mitte 2009 gefördert. 80 Deutsche Institute mit ca. 13.000 Mitarbeitern auf dem Gebiet der Naturwissenschaften waren an dem Test beteiligt (Carusi und Reimer 2010). Als Ergebnis entstand ein Open-Source-Framework für VREs. Da „eSciDoc“ ein reines Framework ist, bietet es keinerlei Weboberfläche, die als Basis einer virtuellen For-

schungsumgebung eingesetzt werden und als Grundlage einer Weiterentwicklung dienen kann.

„gCube“ ist ein VRE-Framework, das in Zusammenarbeit mit der EU entwickelt wurde, um serviceorientierte Infrastrukturen zu implementieren. Es lässt sich für verschiedene Szenarien modular erweitern. Zudem besteht die Möglichkeit, weitere Services z.B. der Europäischen Grid Initiative (EGI) (egi.eu 2017) und des „European Grid for E-science project“ (EGEE 2017) einzubinden. Seit 2007 wird „gCube“ als Basis für verschiedene durch die EU geförderte e-Science-Projekte verwendet, wie z.B.: „DILIGENT - A Digital Library Infrastructure on Grid ENabled Technology“ (2004-2007), „D4Science - DIstributed colLaboratories Infrastructure on Grid ENabled Technology for Science“ (2008-2009), „D4Science-II - Data Infrastructure Ecosystem for Science“ (2009-2011), „iMarine - Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources“ (2011-2017), „EUBrazilOpenBio - EU-Brazil Open Data and Cloud Computing e-Infrastructure for Biodiversity“ (2011-2013), „ENVRI - Common Operations of Environmental Research Infrastructures (2011-2017).

Beide Projekte sind im Förderungszeitraum zu leistungsstarken Frameworks mit innovativen Ansätzen geworden. Nach dieser Zeit scheinen die Projekte jedoch nicht mehr dem aktuellen Stand der Technik zu folgen oder zentral weiterentwickelt zu werden (Förderungszeitraum bei „eSciDoc“ endete 2009). Die Zahlen der Publikationen lag bei maximal fünf pro Jahr und sinkt für „gCube“- die letzte Publikation wurde laut Homepage (gcube-system.org 2017c) im Jahr 2015 veröffentlicht. Sechs Applikationen sind derzeit für „gCube“ verfügbar (gcube-system.org 2017a).

VRE Frameworks sind ein bedeutender Trend, da sie als Basis zur Entwicklung von verschiedenen VREs zum Einsatz kommen können (Carusi und Reimer 2010), (vgl. Kapitel 2.5.1).

2.5.6 Wirtschaftliche Aspekte

Durch den Verfall der Preise für Speichermedien könnte der Eindruck entstehen, dass die Aufbewahrung von Daten kaum Kosten oder Personalaufwand nach sich zieht. Diese Fehleinschätzung stellt oftmals die größte Gefahr für den langfristigen Erhalt von Forschungsdaten dar (Buddenbohm et al. 2014).

Relevante Kostenfaktoren können auf unterschiedliche Weise gruppiert werden. Es bietet sich z.B. eine kaufmännische Einteilung nach Kostenarten an (Tabelle 2.2). Geeignet ist aber auch eine zeitliche Einteilung, die durch das Life-Cycle-Modell (vgl. Kapitel 2.5.2) gegeben ist.

Tabelle 2.2: Kostenarten einer VRE nach Buddenbohm et al. (2014)

Kostenart	Beispiel
Kapital vor Gründung	Vorgründungskosten, Gebühren und ggf. Einlagen
Personalkosten	Gehalt für Entwickler, Administratoren, Support
Betriebskosten außer Personal	Strom, Kühlung, Gebäudenutzung (Miete)
Sachkosten	Hardware, Software
Investitionen	Neue Anschaffungen / Infrastruktur

Die folgende Auflistung erläutert die einzelnen Kostenarten nach Buddenbohm et al. (2014):

- **Kapital vor Gründung / Rechtsformen:** Im Fall, dass der betriebliche Rahmen einer VRE eine von übergeordneten Institutionen unabhängige Rechtsform erhalten soll, muss Kapital in Form von Einlagen zur Gründung bereit stehen.
- **Personalkosten:** Personalkosten stellen den mit Abstand größten Kostenfaktor dar. Sie werden nach Buddenbohm et al. (2014) bis zu ca. 86% geschätzt. Direkte Personalkosten umfassen die Gehälter für Mitarbeiter zur Entwicklung, dem technischen Betrieb und den Support der Nutzer. Die indirekten Personalkosten entstehen durch die Verwaltungskosten für das Personal.
- **Betriebskosten:** Die Betriebskosten umfassen über die Personalkosten hinaus die Bewirtschaftungskosten für die technische Infrastruktur. Dazu zählen beispielsweise Kosten für Abschreibung oder Miete für Hard- und Software, Strom zum Betrieb der Geräte und ggf. zur Kühlung bei größeren Rechenanlagen sowie die Aufwendungen für Gebäudenutzung (z.B. Miete, Umlagen und Instandhaltung) und Versicherungen.
- **Sachkosten:** Eine weitere Kostenart sind die Sachkosten. Sie entstehen z.B. durch Mittel für die Durchführung von Weiterbildungen, Workshops, Dienstreisen, Verbrauchsmaterial, Öffentlichkeitsarbeit und Fremdleitungen.
- **Investitionskosten:** Die Investitionskosten entstehen durch neue Anschaffungen und Erweiterungen einer VRE, beispielsweise in Form von Hardware oder Softwarelizenzen. Diese Kosten grenzen sich von den laufenden Betriebskosten ab.

Die Kosten mit zeitlichem Bezug zeigt folgende Übersicht nach Ludwig und Enke (2013):

- **Nutzungskosten:** Die Kosten des Datenmanagements, die Aufbewahrung und Pflege der Daten sind Teil der Nutzungskosten. Sie sind nicht zusätzlich oder optional anzusetzen, da die Nutzung der Daten das Ziel ist.
- **Höchste Kosten zu Beginn:** Bei einer Einteilung der Datenhaltung in Übernahme-, Speicher- und Zugriffsphase lassen sich die Kosten ca. zur Hälfte für

die erste Phase der Datenübernahme kalkulieren. Die Zugriffsphase ist die zweitaufwendigste gefolgt von der Speicherphase, die am günstigsten ist.

- **Abwarten ist teuer:** Die Kosten für die Anfangsphase können kaum aufgeschoben werden, weil spätere Nacharbeiten meist teurer sind, wie z.B. das nachträgliche Erheben von Metadaten oder die Qualitätskontrolle. Hierbei müsste ggf. ein Teil der Arbeit mehrfach durchgeführt werden.
- **Jährliche Kosten:** Sind die einmaligen hohen Anfangskosten durch Hard- und Softwareinvestitionen sowie Personalaufwendungen überwunden, sinken die Aufbewahrungskosten durch die zunehmende Effizienz der Technologien. Die Wartungskosten durch Personal steigen jedoch mit dem Daten- und Nutzervolumen und Frequenz der erforderlichen Updates.
- **Anreize schaffen:** Fehlende Anreize können die Durchführung von erforderlichen Maßnahmen verhindern. So können z.B. Erstverwertungsrechte oder das Zitieren von Forschungsdaten ein wichtiger Faktor im Forschungsdatenmanagement sein, um eine erforderliche Anerkennung der Teilnehmer zu gewährleisten.

Die Gründung einer eigenständigen Organisation als Träger der VRE kann nach Buddenbohm et al. (2014) einige Vorteile haben. Zum einen ermöglicht diese Rechtsform das Schließen von Verträgen zum Beispiel mit Dienstleistern zur Bereitstellung der Infrastruktur. Oftmals ist es problematisch, die bestehenden Infrastrukturen der beteiligten Institutionen selber zu nutzen, da hier Kapazitäten, Sicherheitsaspekte und rechtliche Rahmenbedingungen eine Rolle spielen. Zum anderen kann die Verwaltung und Organisation der VRE an zentraler Stelle erfolgen. Dies vereinfacht den Betrieb und die Kosten.

Folgende Finanzierungsmodelle für VREs sind nach Buddenbohm et al. (2014) naheliegend:

- Indirekte öffentliche Förderung (Fördermittel, Trägerorganisation, etc.),
- Finanzierung durch Beiträge (Institutionen oder Communities zahlen Pauschalbeitrag),
- Direktes Nutzungsmodell (auf Basis der tatsächlichen Nutzung),
- Ausgleich von Ressourcenverbrauch und Ressourcenbereitstellung (jede Institution ist selbst für ihre Finanzierung verantwortlich).

2.5.7 Erfolgskriterien für VREs

Die Bestimmung von Erfolgskriterien und deren Indikatoren setzt eine Diskussion voraus, was unter einer erfolgreichen VRE verstanden wird. Der Erfolg kann allgemein definiert werden oder mit Fokus auf einzelne Institutionen und Stakeholder. Idealerweise werden die zu erreichenden Ziele zu Beginn definiert, beispielsweise in einem initialen Antragsverfahren. Dabei geben die vielfältigen Eigenschaften der späteren

Nutzercommunity den benötigten Leistungsumfang der VRE vor. Von Seiten der Betreiber und Entwickler wird durch ein Angebot versucht, diesen Anforderungen gerecht zu werden. Der optimale Abgleich zwischen den Anforderungen und dem Angebot stellt die Grundlage für den Erfolg einer VRE dar. Dazu müssen bei der Wahl der Erfolgskriterien ggf. Zielkonflikte berücksichtigt werden. Umfangreiche Funktionswünsche seitens der Nutzer können zum Beispiel in Konflikt mit den begrenzten Ressourcen der Betreiber und Entwickler stehen. Zum Schluss ergibt sich daraus eine individuelle Erfolgsdefinition für die jeweilige VRE.

Die Erfolgsfaktoren und messbare Erfolgskriterien aus der Sicht der einzelnen beteiligten Akteur-Gruppen (Benutzer, Betreiber, Förderer) sind in Abbildung 2.8 dargestellt. Die Auswahl der Kriterien richtet sich insbesondere nach den disziplinär spezifischen Anforderungen und individuellen Gesamtzielen. Dabei zielen sie auf den konkreten Nutzen der VRE für die wissenschaftliche Arbeit der Benutzer ab. In erster Linie geht es dabei um die Kriterien, die sich aus den gegebenen VRE-Definitionen herleiten lassen.

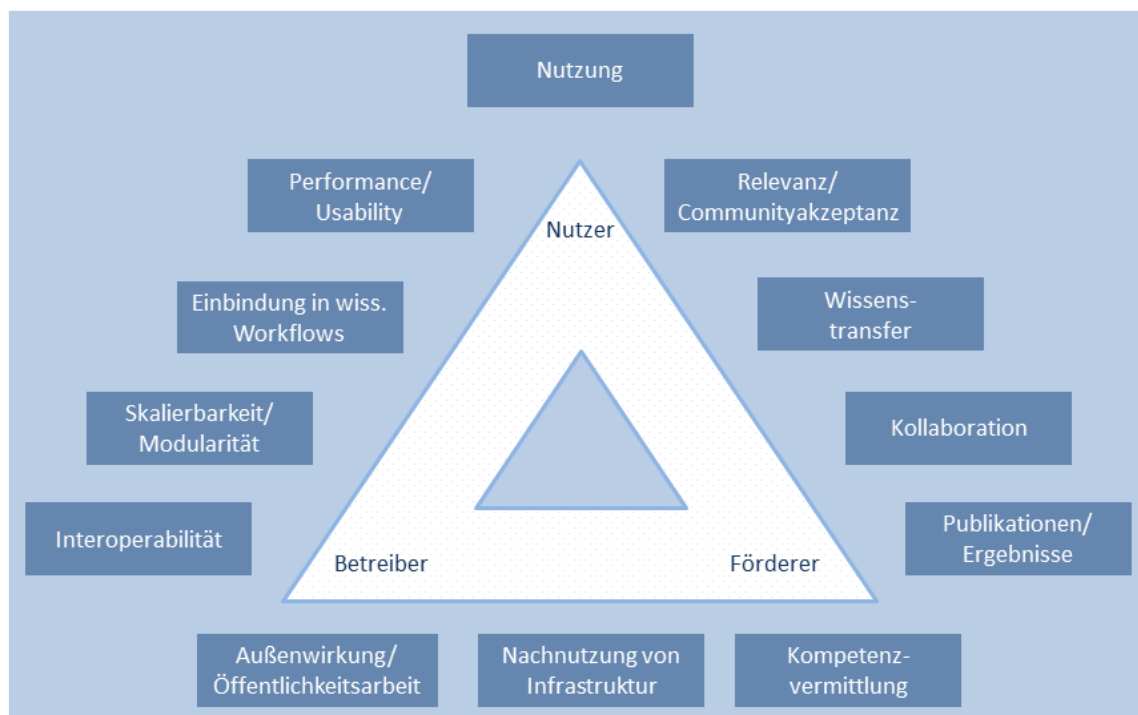


Abbildung 2.8: Erfolgsfaktoren und messbare Erfolgskriterien für VREs nach Relevanzgewichtung in den Stakeholderkategorien unterschieden (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0

In der Abbildung 2.8 sind die drei wichtigsten Akteur-Gruppen im Dreieck als Vorlage für eigene Kriterien dargestellt. An den Seiten sind die jeweiligen Erfolgskriterien angeordnet. Dies verdeutlicht die unterschiedlichen Schwerpunkte für eine Gewichtung der einzelnen Kriterien in der Erfolgsbeurteilung. Tabelle 2.3 enthält die Zusammenfassung aus der Studie von Buddenbohm et al. (2014).

Die Gruppe der Entwickler trägt ebenfalls zum Erfolg von VREs bei, da sie das Konzept einer VRE technisch implementieren. Über das notwendige technische Know-

how hinaus müssen sie interdisziplinär allgemeinverständlich kommunizieren und sich bei der Entwicklungsplanung beteiligen, um z.B. das technisch Machbare und den erforderlichen Aufwand abzuschätzen. Entwickler können ebenfalls gemeinsame oder konträre Interessen mit den drei anderen Stakeholderkategorien haben, wie z.B. Außenwirkung oder im Gegensatz weniger Programmieraufwand.

Tabelle 2.3: Erfolgskriterien einer VRE nach Buddenbohm et al. (2014)

Kriterium	Beschreibung
Nutzung	Anzahl der Benutzer, potentielle Nutzergruppe, Intensität der Nutzung, Nutzung bestimmter Komponenten, Forschungsabschnitt
Relevanz/Community-Akzeptanz	Wahrnehmbarkeit der VRE in der potentiellen Nutzergruppe, Auswirkung der VRE auf die Disziplin, Bewertung der Nützlichkeit, Wirkung außerhalb der Fachcommunity
Wissenstransfer	Erfahrungsaustausch, neue Methoden und Arbeitsziele, Projektdokumentation
Kollaboration	Kommunikationsformen, Praktische Umsetzung, Alleinstellungsmerkmale anderen Infrastrukturen
Publikationen/ Ergebnisse	Anzahl Publikationen, Arten von Publikationen, Anzahl an zusätzlichen Publikationen durch die VRE
Kompetenzvermittlung	Anteil der VRE an der Kompetenzvermittlung (Lehre und Forschung), Erwerb wissenschaftlicher Qualifikationen (Abschlussarbeiten, Promotionen, etc.), nicht wissenschaftsbezogene Kompetenzen (Teamfähigkeit, sozio-ökonomische Aspekte, etc.)
Nachnutzung von Infrastruktur	Nutzung von VRE-Komponenten für andere Projekte, Eignung für Nachnutzung in neuem Zusammenhang, Qualität und Umfang von Dokumentationen, Bereitstellung von Informationen, Erkenntnissen und Ergebnissen
Außenwirkung/ Öffentlichkeitsarbeit	Erfolg der Öffentlichkeitsarbeit, Sichtbarkeit in der potentiellen Nutzergruppe, Vermittlung der VRE an die Förderer, Wahrnehmung in anderen Fachdisziplinen, Wahrnehmung in der allgemeinen Öffentlichkeit
Interoperabilität	Einsatz etablierter Standards in der ICT, Einsatz etablierter Standards in der Forschung, Daten- und Metadatenaustausch mit anderen Systemen
Skalierbarkeit/ Modularität	Anpassbarkeit der VRE an: erweiterte Benutzergruppen, andere Forschungsfragen, Nachnutzung einzelner VRE-Bestandteile
Einbindung in wissenschaftliche Workflows	Umsetzung der wichtigsten Workflows, Abdeckung von Abschnitten und Arbeitsprozessen
Performance/ Usability	Erlernbarkeit des Umgangs mit der VRE, Benutzerfreundlichkeit, Anpassung an den Arbeitsalltag

Kriterium	Beschreibung
	der Nutzer, Vergleich zu anderen generischen Werkzeugen, Geschwindigkeit von rechenintensiven Prozeduren

Für die Ermittlung des Erfolgs können die Kriterien unterschiedlich gewichtet werden. Es geht dabei nicht in erster Linie um eine quantitative Messbarkeit, welchen gegebenenfalls einen falschen Eindruck in Form einer Pseudogenauigkeit suggeriert. Vielmehr ist hier die Definition der Zielvorgaben vorrangig, woran später der Erfolg einer VRE festzumachen ist. Ein wichtiges Instrument können Meilensteine auf diesem Weg sein, die das Ziel in kleinere Teilschritte untergliedern, wie in Abbildung 2.9 dargestellt (Buddenbohm et al. 2014).

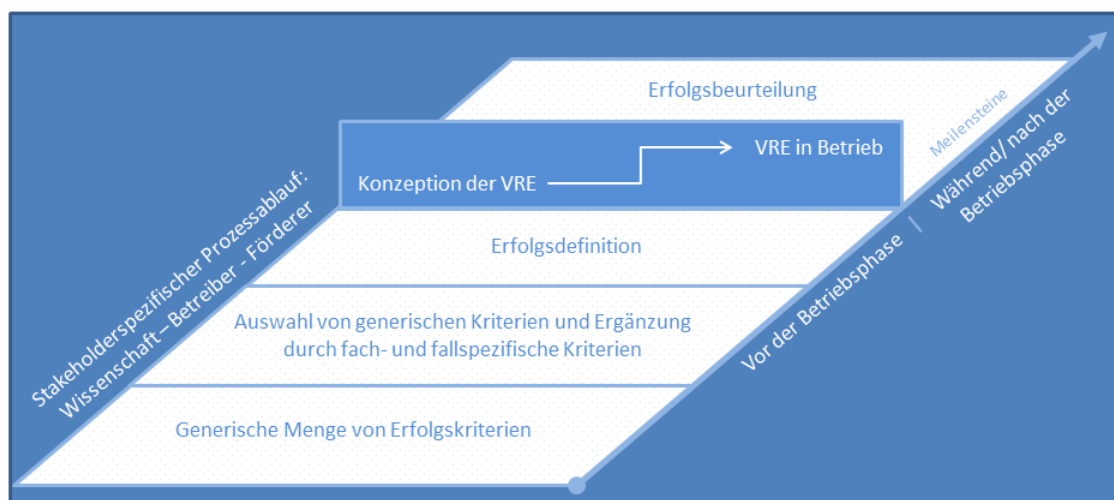


Abbildung 2.9: Erfolgsbeurteilung in Abhängigkeit vom Entwicklungsstand der VRE (Buddenbohm et al. 2014), Lizenz: Creative-Commons Attribution 4.0

2.5.8 Data Policies

Für den wissenschaftlichen Umgang mit Daten müssen allgemeingültige und verbindliche Regeln definiert werden, um eine „gute wissenschaftliche Praxis“ (Deutsche Forschungsgemeinschaft 1998, 2013) auch in der Realität einzuhalten und den offenen Zugang zu Daten zu gewährleisten (open access). So müssen Policies (interne Rahmenvorschriften, Leit- und Richtlinien) aufgestellt werden, welche die Aufnahme, die Bereitstellung und den Zugriff auf Daten im Archiv regeln (Büttner et al. 2011). Solche verpflichtenden Regeln gelten nur innerhalb eines organisatorischen Rahmens. Ein Beauftragter in der Datenhaltung sollte bei größeren Projekten festgelegt werden, der die Prozesse koordiniert, als Ansprechpartner dient und die Verantwortung für die Verarbeitungsprozesse der Daten nach den Vorgaben trägt.

Die Vorgaben und Regeln der Data Policies können in Projektbeschreibungen, Richtlinien, Arbeits- und Dienstanweisungen schriftlich fixiert und ebenso in der Software durch Integration in Arbeitsabläufe verankert werden. Zum Beispiel mit Hilfe

geeigneter Infrastruktur (Hard- und Software) in der Datenerfassung (Messgeräte, Sensoren, etc.), die automatisch Metadaten erzeugen oder den Benutzer bei der Eingabe von Metadaten aktiv unterstützen.

Probleme im Bereich der Data-Policies entstehen oft weniger durch technische Grenzen, als durch anfängliche Akzeptanzschwierigkeiten mit ggf. vermeintlichem Mehraufwand der Beteiligten (Büttner et al. 2011). Hier kann durch rechtzeitige Einbindung der Beteiligten in die Entscheidungsprozesse und durch Schulungen ggf. mit externen Experten die Sensibilität und die Akzeptanz verbessert werden. Problematisch bleibt meist der Umgang mit kleinen, unstrukturierten Daten, da hier der individuelle Aufwand bei der Metadatenerfassung und Datenhaltung nicht im Verhältnis zur Datenmenge steht und die Datenstrukturen heterogener Natur sind (vgl. Klump und Bertelmann 2013).

Die Regeln können in einer Data Policy fest integriert sein, die z.B. Bestandteil sind:

- in ICT-Systemen,
- in Websites,
- in Handbüchern und Dokumentationen,
- in Dienstanweisungen,
- in Studienordnungen,
- in Verträgen,
- in Memorandum of Agreement.

Probleme bei der Umsetzung ergeben sich meist in den folgenden Bereichen:

- Akzeptanz bei Personal / Anreize,
- zusätzlicher Aufwand (Dilemma der Prioritäten),
- Small Science Data (Aufwand / Datenmenge),
- Integration in Arbeitsabläufe (Workflow und ICT),
- schneller Fortschritt in ICT (Kompatibilität, Standards),
- Interoperabilität (Austausch von Metadaten),
- Verantwortung für Datenhaltung,
- allgemeine Rechtsgrundlagen, Richtlinien, Arbeitsanweisungen.

2.5.9 Rechtliche Aspekte

Der Umgang mit Forschungsdaten betrifft im Wesentlichen zwei rechtliche Bereiche (Ludwig und Enke 2013). Zum einen sind das die Aspekte des Datenschutzes mit dem Schutz personenbezogener Daten und zum anderen die Fragen des Urheberrechtes. Das Urheberrecht gliedert sich zudem in zwei Unterbereiche, welche die Benutzung oder Nachnutzung von urheberrechtlich geschützten Werken und als zweites die Erstellung bzw. Kontrolle eigener neuer urheberrechtlich geschützter Werke regeln.

Forschungsdaten unterliegen unter anderem mit besonderen Bestimmungen nationalen und internationalen Gesetzen. An zentraler Stelle steht das Bundesdatenschutzgesetz¹ (BDSG), welches gesetzliche Bestimmungen über die Erhebung, Verarbeitung und Nutzung personenbezogener Daten regelt. Dieses Rahmengesetz ist verbindlich für staatliche und nichtstaatliche Institutionen und wird durch die Bundesländer in eigenes Landesrecht umgesetzt. Danach dürfen personenbezogene Daten nur erhoben werden, wenn dies gesetzlich zulässig ist (z.B. durch richterliche Anordnung) oder wenn die betroffene Person ihre Einwilligung gegeben hat (im Normalfall schriftlich). Ziel des BDSG ist es, den Einzelnen vor Missbrauch durch personenbezogene Daten zu schützen.

Im BDSG ist für Forschungsdaten eine besondere Bestimmung zur Verwendung personenbezogener Daten verankert. Danach dürfen personenbezogene Daten, die in der Forschung und Wissenschaft erhoben wurden, auch nur im Rahmen dieser Zwecke verarbeitet oder genutzt werden. Dabei müssen die Daten anonymisiert werden, sobald dies der Forschungsprozess zulässt. Zuvor müssen alle personenbezogenen Daten gesondert aufbewahrt werden und dürfen nur mit anderen Daten kombiniert werden, soweit es den Forschungszweck erfüllt.

Die Veröffentlichung personenbezogener Daten durch wissenschaftliche Institutionen setzt das Einverständnis der betroffenen Personen (Forscher und Probanden) voraus oder darf erfolgen, wenn diese Daten für die Präsentation der Forschungsergebnisse von aktuellen Ereignissen unverzichtbar sind.

Forschungsdaten fallen in Deutschland zudem unter das Urheberrecht² (UrhG), sofern ein Werk das Erfordernis des persönlichen Schaffens, der wahrnehmbaren Formgestaltung, des geistigen Gehaltes und der eigenpersönlichen Prägung erfüllt (Lutz 2009). Dies ist jedoch bei Forschungsdaten nicht immer gegeben, wenn z.B. nur unstrukturierte Messdaten erzeugt werden. Werden die Daten hingegen in einer Datenbank gesammelt

¹ Bundesdatenschutzgesetz in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 7 des Gesetzes vom 30. Juni 2017 (BGBl. I S. 2097) geändert worden ist

² Urheberrechtsgesetz vom 9. September 1965 (BGBl. I S. 1273), das zuletzt durch Artikel 1 des Gesetzes vom 1. September 2017 (BGBl. I S. 3346) geändert worden ist"

oder wird für die Erstellung, Darstellung oder Auswertung der Forschungsdaten ein eigenes Werkzeug eingesetzt, so kann dies einen Urheberschutz begründen. Dies sollte daher bei der Projektplanung berücksichtigt werden.

Fällt das Werk nach den genannten Kriterien unter den Schutz des deutschen Urheberrechts, sind nur wenige Nutzungsarten ohne Zustimmung des Urhebers rechtmäßig. Hierzu zählen:

- Vervielfältigung der reinen Fakten im Rahmen einer eigenen Interpretation und Wortwahl,
- Vervielfältigung und Nutzung zum privaten Gebrauch,
- Vervielfältigung, Verbreitung und öffentliche Wiedergebe in der Form von Zitaten.

Prinzipiell sollte bei allen fremden Daten davon ausgegangen werden, dass sie dem Urheberrecht unterliegen. Das gilt grundsätzlich auch bei fremdem geistigem Eigentum in Forschung und Wissenschaft. Besteht für ein Werk nach den genannten Kriterien ein urheberrechtlicher Schutz, so ist die Einwilligung des Urhebers zum Zwecke der Vervielfältigung oder Weiterverarbeitung unerlässlich. Auch in Datenbanken gesammelte einzelne Daten ohne eigene Schutzrechte unterliegen nach deutschem Recht einem spezifischen Schutzrecht in der Datenbank als Gesamtheit. Demnach hat der Ersteller der Datenbank das alleinige Recht ihrer Verbreitung und Vervielfältigung. Nur die Verwendung eines *unwesentlichen* Teils der Datenbank kann ohne die Zustimmung des Erstellers erfolgen. Ausgenommen davon ist die Verwendung eines *wesentlichen* Teils der Datenbank zu privater Nutzung oder zu Lehrzwecken. Wie diese unbestimmten Rechtsbegriffe im Einzelnen auszulegen sind, wird von Zurth (2015) beschrieben: Ob der genutzte Teil der Datenbank wesentlich oder unwesentlich ist, hängt von den finanziellen, menschlichen und technischen Investitionen des Datenlieferanten in diesem Teil der Datenbank ab (qualitative Wesentlichkeit). Unerheblich in diesem Zusammenhang sind die diesem Teil innewohnenden objektiven Werte sowie dessen Werte für den Nutzer. Alternativ kann die Wesentlichkeit eines Teils quantitativ festgestellt werden. Dabei ist das Verhältnis des verwendeten Teils in Bezug auf das gesamte Volumen der Datenbank zu bewerten.

Ebenso wie beim Umgang mit fremden Daten sollte der Schutz der eigenen Daten durch das Urheberrecht bedacht werden und ebenso die Methoden, um diese Rechte zu kontrollieren zu können. Die Urheberrechte verbleiben normalerweise beim Urheber, wie es der Namen „Urheberrecht“ vorsieht. Der Urheber kann jedoch Nutzungsrechte zur weiteren Verwendung der Daten erteilen. Wenn keine Nutzungsbeschränkung vorgesehen ist oder nur spezifische Restriktionen erwünscht sind, eignet sich z.B. die offene GPL (gnu.org 2017) oder Creative Commons Lizenz (creativecommons.org 2017). Hierbei müssen jedoch die Fälle beachtet werden, in denen bei der Erstellung von eigenen Programmen Teile fremder Software (Quelltexte) verwendet werden. Die für die fremde Software gültigen Lizenzmodelle können dadurch die Wahl der eigenen Lizen-

sierungsmöglichkeiten einschränken, z.B. durch das „Copyleft-Prinzip“. Zudem kann bei bestimmten Datenarten das Patentrecht greifen, wenn die verwendeten oder erzeugten Daten durch wissenschaftliche, technische oder methodische Patente geschützt sind (Ludwig und Enke 2013).

Schutzgüter nach dem Urheberrecht unterliegen einer Frist, nach deren Ablauf die Urheber und andere Rechteinhaber ihre Exklusivrechte verlieren. Danach stehen diese Daten gemeinfrei der Öffentlichkeit zur Verfügung (bei Texten z.B. 70 Jahre nach dem Tod des Urhebers). Es ist also möglich, dass sich für den Zeitraum der Datenaufbewahrung die Rechte ändern und somit nach Ablauf einer Frist die Forschungsdaten nicht mehr urheberrechtlichem Schutz unterliegen. Damit wäre eine weitergehende oder uneingeschränkte Nutzung der Daten möglich, wobei zu beachten ist, dass diese Fristen länderspezifisch unterschiedlich sein können (Ludwig und Enke 2013).

2.5.10 Technische Voraussetzungen von L-TER VREs

Die folgenden technischen Voraussetzungen für VREs sind gegeben (Munson et al. 2016):

Benutzerverwaltung:

- **Benutzeraccounts** sind erforderlich zur Authentifizierung des Zuganges und Individualisierung der VRE, ebenso wie der Nachvollziehbarkeit von Änderungen.
- **Single Sign-on (SSO)** sollte als zentrale Anmeldung und Authentifizierung für alle Dienste und Module ermöglicht werden. Dies schafft eine starke Vereinfachung und dadurch eine erhöhte Akzeptanz auf Benutzerseite.
- **Benutzergruppen** ermöglichen einen gruppenorientierten Zugriff auf Inhalte in verschiedenen Ebenen und mit speziellen Zugriffsrechten. Sie eignen sich auch, um institutionelle Rollen abzubilden.

Ressourcenmanagement:

- **Persistente Datenspeicherung** muss gewährleistet sein, damit die Daten unabhängig von Personen und Prozessen reproduzierbar gespeichert werden können.
- **Metadaten** müssen für die Suche, Analyse und Bewertung von Daten erzeugt und gespeichert werden.
- **Versionsmanagement** ist erforderlich, um die Bearbeitungsstände nachvollziehbar zu kennzeichnen und die Wiederherstellung früherer Versionen zu ermöglichen.

- **Ressourcenorganisation** gibt den Daten eine Gliederung, die am besten nach einer Methode erfolgt, die den Benutzern bereits vertraut ist, wie z.B. eine Hierarchie oder Graphen.

Ressourcenaustausch:

- **Import** von Daten in die VRE muss einfach aber flexibel erfolgen, um sie dort bearbeiten zu können.
- **Export** von Daten aus einer VRE ist erforderlich, damit eine Bearbeitung mit externen Werkzeugen erfolgen kann.
- **Offene Datenschnittstellen** ermöglichen den Import, Export und das Teilen von Ressourcen.
- **Data-Harvesting** ermöglicht den automatischen Datenaustausch zwischen externen Systemen und VREs.

Suche:

- **Metadaten** erleichtern das Auffinden von Ressourcen in der VRE.
- **Volltextsuche** ermöglicht die Suche nach Texten und Zahlen in den Ressourcen.
- **Kategorien** erlauben, eine Suche zu filtern und ähnliche Ressourcen zu finden.

Forschungswerkzeuge:

- **Datenanalyse** ist eine der wesentlichen Funktionen von VREs.
- **Datenaggregation** hilft bei der Zusammenfassung von Daten und Gewinnung neuer Erkenntnisse für die Forschung.
- **Datenvisualisierung** zeigt graphisch die Eigenschaften der Daten sowie deren Zusammenhänge und kann für die Publikation von Ergebnissen hilfreich sein. Automatisierungen sind hier nützlich.
- **Dokumentation** von z.B. Versuchen und verwendeten Parametern ist notwendig für die Nachvollziehbarkeit.
- **Modellierung** mit Parametrisierung, Kalibrierung, Validierung und Verknüpfung von Forschungsmodellen stellt Zusammenhänge dar.
- **Simulation** eignet sich für verschiedene Szenarien in Forschungsmodellen.

Projektmanagement:

- **Aktionspläne** halten fest, welche Forschungsaufgaben wann und von wem bearbeitet werden sollen.

- **Meilensteine** spiegeln die Ziele der Forschung wieder und können in den Forschungsplan eingebunden werden.
- **Problemmanagement** schafft eine systematische Methode, wie auftretende Probleme kommuniziert und gelöst werden.
- **Optimierungsvorschläge** an der VRE sollten zentral kommuniziert werden, wobei Umfragen zur Priorisierung dienen können.

Interne Kommunikation:

- **Kontaktinformationen** sind erforderlich, damit die Benutzer sich untereinander kontaktieren können, um Informationen auszutauschen. Darüber hinaus ist es sinnvoll, weitere Details auf einer Profil-Homepage der Benutzer zu hinterlegen, wie z.B. Expertisen, Interessen und sonstige sachdienliche Angaben.
- **Kalender** helfen bei der Organisation von Terminen, Treffen und Konferenzen.
- **Mailingliste/Foren** ermöglichen den Austausch von Informationen mit der gesamten Community.
- **Wikis** bieten in Form von strukturierten Informationen eine Möglichkeit zur gemeinsamen Dokumentation der VRE und zur Erstellung von Forschungsbeiträgen im Team.
- **Benachrichtigungen (Notifier)** informieren über aktuelle Vorgänge in der VRE, wie z.B. neue Ressourcen oder Änderungen von beobachteten Objekten.
- **Dynamische Benutzergruppen** ermöglichen die individuelle Erstellung von zeitlich veränderlichen, themenbezogenen Gruppen mit Bündelung der Kommunikation.
- „**Social Networks**“ im Rahmen von VREs vereinen die interne Kommunikation und bieten dynamische Gruppen, wie bei den klassischen Vorbildern („Facebook“, „LinkedIn“, „Twitter“). Sie schaffen darüber hinaus ein Gefühl der Gemeinschaft.

Externe Kommunikation:

- **Websites/Homepages** sind der zentrale Einstiegspunkt für alle Benutzer der VRE. Hier sollte beschrieben werden, zu welchem Zweck sie dient und welche Funktionen angeboten werden.
- **Blogs** veröffentlichen Artikel oder Berichte zu einem bestimmten Thema, das für die Öffentlichkeit von Interesse ist. Sie können auch dazu dienen, Informationen mit anderen Social-Networks zu teilen (z.B. „Like Button“).
- **Profilseiten** der Benutzer oder Institutionen geben darüber Auskunft, wer mit der VRE arbeitet.

Publikation:

- **Zugang und Archiv** zu den Resultaten der Forschung in Form von Publikationen und Daten im Sinne von „Open Access“.

Bibliographie:

- **Publikationsquellen**, die als Basis für die Forschung dienen, sollten in die VRE importiert werden können.
- **Schlagworte** organisieren die Literatur.
- **Anhänge/Daten** auf denen eine Publikation basiert, können in der VRE weiterverarbeitet werden.

2.6 Nationale und europäische Aktivitäten

VREs wurden in Deutschland zuerst nur vereinzelt systematisch erfasst oder eingeordnet (Deutsche Initiative für Netzwerkinformation e.V. 2011). Es gingen anfänglich wenige VREs von der Aufbauphase über in den normalen Betrieb (Carusi und Reimer 2010). Mittlerweile wird das Management von Forschungsdaten in VRE-ähnlicher Weise oft zur Voraussetzung (und Hürde) bei geförderten Projekten. Viele Institutionen in Deutschland beschäftigen sich deshalb mit dem Thema. Ausgewählte Projekte sind:

- **Deutsche Initiative für Netzwerkinformation e.V. (DINI):** Die DINI zeigt auf, dass die wissenschaftliche Forschung in allen Disziplinen zunehmend in vernetzten Strukturen erfolgt. Dies ermöglicht zugleich neue Forschungsfelder und -methoden. Aufgabe aller Informationsserviceeinrichtungen ist es, Forschungsumgebungen für Wissenschaftler bereitzustellen, die ihnen zeit- und ortsunabhängige Zusammenarbeit auf webbasierten Plattformen ermöglicht (Deutsche Initiative für Netzwerkinformation e.V. 2008).
- **Wissenschaftsrat:** Für den Wissenschaftsrat sind Informationsstrukturen (zu denen VREs grundlegend gehören) eine „unverzichtbare Voraussetzung für Forschung, Lehre und Nachwuchsförderung in allen Disziplinen“ (Wissenschaftsrat 2011) und daher ein wesentlicher Bestandteil des Wissenschaftssystems. Sie sammeln, bewahren und pflegen wissenschaftliches und nichtwissenschaftliches Wissen und machen es nachhaltig für weitere wissenschaftliche Auswertungen verfügbar.
- **Kommission Zukunft der Informationsinfrastruktur (ZKI/KII):** Die Kommission (ZKI 2017) definiert „Informationsinfrastruktur“ als nationales, disziplinübergreifendes „Netz“ von Einrichtungen. Diese nehmen im öffentlichen

bzw. institutionellen Auftrag die Versorgung von Wissenschaft und Forschung mit Information und damit zusammenhängenden Dienstleistungen wahr.

- **INSPIRE Direktive für Geodaten:** Um die grenzübergreifende Nutzung von Geodaten in Europa zu erleichtern, verabschiedete das Europäische Parlament die INSPIRE Richtlinie 2007/2/EG, die am 15. Mai 2007 in Kraft getreten ist (INSPIRE 2007). Mit dem Inkrafttreten hatten die Mitgliedsstaaten der EU zwei Jahre Zeit, die Richtlinie in nationales Recht umzusetzen. Das Kürzel INSPIRE steht für „Infrastructure for Spatial Information in Europe“ und soll gemeinschaftliche umweltpolitische Entscheidungen unterstützen. Für die Umsetzung fordert INSPIRE webbasierte Online-Dienste für den Zugriff auf Geodaten, welche die Suche, die Visualisierung und den Download der Daten ermöglichen. Die Richtlinie bildet nur den rechtlichen Rahmen für den Aufbau von Geodateninfrastrukturen - fachliche und technische Details sind in ihr selbst nicht geregelt. Für die Mitgliedsstaaten werden lediglich rechtlich verbindliche Festlegungen mittels Durchführungsbestimmungen festgelegt. INSPIRE ist ausschließlich für Geodaten gedacht, die Raum und Umwelt in digitaler Form beschreiben. In den drei Anhängen zur Richtlinie werden 34 Themenfelder beschrieben, auf den sich der Anwendungsbereich bezieht. Die INSPIRE Richtlinie verlangt zudem die Beschreibung der Geodaten durch geeignete Metadaten. Weitere Details zu den konkreten fachlichen Inhalten werden über die Rahmenbedingungen zur Durchführung geregelt, wobei die Mitgliedsstaaten verantwortlich für die Identifizierung der relevanten Datensätze und die INSPIRE-konforme Darstellung im eigenen Land sind. Durch die Mitgliedschaft in der EU nimmt INSPIRE auch Deutschland als Ganzes in die Pflicht. Die Geodaten des Bundes, der Länder und der Kommunen sollen nach den Vorgaben interoperabel verfügbar sein, wobei das grundlegende Prinzip dafür zentrale Strukturen und dezentrale Daten vorsieht. Das bedeutet konkret, dass die Geodateninfrastruktur Deutschland (GDI-DE) (geoportal.de) als technisch-fachliches Netzwerk für die zentralen Strukturen genutzt werden und die Daten dezentral von den fachlichen Institutionen „on demand“ bereitgestellt werden sollen. Dazu ist es erforderlich, dass alle von INSPIRE betroffenen staatlichen Organe fest in die GDI-DE eingebunden werden.
- **ICT – ENSURE:** Mit dem Hauptziel, die Bedeutung von ICT für die Sicherung einer nachhaltigen Umwelt einzuschätzen, startete 2008 das von der Europäischen Kommission im Rahmen des „7th European Research Framework Programm“ (Europäische Union 2017) geförderte Projekt „ICT for ENvironmental SUstainability Building the European REsearch Area“ (ICT-ENSURE 2017; Gnauck und Pillmann 2011). Es wurde auf zwei Jahre angesetzt und hatte die Aufgabe einen Überblick zu dem Bedarf an ICT und deren Werkzeuge im Umweltsektor in Europa zu geben. Dazu sollte eine Studie auch die Schwerpunktbereiche für ICT in der Umwelt identifizieren, ein Rahmenprogramm erstellen, Informationen und Berichte bereitstellen, Netzwerke zwischen Experten und Insti-

tutionen stärken und weitere Forschungsschwerpunkte und Anforderungen für die Zukunft analysieren (Wohlgemuth 2009).

- **SISE und SEIS Informationssystem:** Mit dem Entwurf des Grundlagenpapiers zu *e-Environment* (Elektronischer Zugriff auf Umweltinformationen), der durch das Komitee für *e-Democracy* der EU (CAHDE 2009) erarbeitet wurde, entstand ein Informationssystem für den digitalen Zugriff auf öffentliche Umweltinformationen mit dem Namen „Shared Environmental Information System (SEIS)“ (European Environment Agency 2016). Dieses Informationssystem soll zukünftig als Basis für e-Environment dienen, mit dessen Hilfe Zugang zu öffentlichen Umweltinformationen geschaffen und somit eine aktive Teilnahme an Entscheidungsprozessen und Umweltschutz für Bürger, Experten und Entscheidungsträger ermöglicht wird. Das „7th European Research Framework Programm“ (Europäische Union 2017) schlug weiterhin die Entwicklung eines „Single Information Space in Europe for the Environment“ (SISE) vor, der einheitlich und übergeordnet für die EU Umweltinformationen bereitstellt. Die SISE Datenbank war erforderlich für den Aufbau des Informationssystems SEIS und der Entwicklung und Implementation von dessen Funktionalitäten und Werkzeugen. Die Verbindung dieser beiden Systeme schaffte die Möglichkeit, reine Umweltdaten mit Expertenwissen zu kombinieren und somit als Entscheidungsunterstützung für den nachhaltigen Umweltschutz zu dienen.

Die nationalen und europäischen Aktivitäten legen dar, dass unter dem Schlagwort „e-Science“ gegenwärtig nachhaltige digitale Infrastrukturen, beispielsweise in Form von VREs, von vielseitigem, wissenschaftlichem Interesse auch im Umweltbereich sind.

2.7 Relevante ICT

Aus der Literatur (vgl. Kapitel 2.5) gehen verschiedene ICT hervor, die sich speziell für den Einsatz in VREs eignen. Auf diese wird im Folgenden näher eingegangen.

2.7.1 Remote Desktop, Screen- und Application Sharing

Mit *Remote-Desktop-Protocol* (vom Englischen für „entfernter Schreibtisch“ oder Fernsteuerung, kurz RDP), *Screen-Sharing* (vom Englischen für „Bildschirm teilen“) und *Application Sharing* (engl. „Teilen von Anwendungen“) lassen sich Computer und Programme fernsteuern. Diese Funktionen erlaubt es durch verschiedene Protokolle, Anwendungen auf dem Desktop eines Computers (Server) auszuführen, die auf einem anderen Computer (Client) dargestellt und bedient werden – so, als würde man direkt vor dem entfernten Computer sitzen. Bei Remote-Desktop und Application-Sharing läuft die Sitzung unabhängig von einer anderen eventuell laufenden Sitzung (unbeaufsichtigter Zugriff ohne Anzeige auf dem Server). Bei Screen-sharing hingegen erfolgt ein beaufsichtigter Zugriff mit Anzeige auf dem Server (Anderson und Griffin 2010).

Im Bereich von VREs lassen sich Remote-Desktop, Screen- und Application Sharing für verschiedene Aufgabenbereiche einsetzen. Nicht nur für Administratoren, die dadurch die ICT fernwarten können, sind diese Funktionalitäten interessant. Insbesondere lassen sich durch RDP und Screen-Sharing z.B. die gemeinsame Nutzung von Softwarelizenzen und Hardware auf einem einzigen Computer realisieren, was mit Kostenersparnis einhergehen kann. Dazu sind allerdings die Lizenzbedingungen der jeweiligen Software für diesen Einsatzzweck genauer zu prüfen. Außerdem entstehen im Fall von MS Windows Servern mit RDP Kosten für die Anzahl der Remote-Desktop-Benutzer durch die *Client Access Licenses* (CAL). Durch die Bereitstellung eines ferngesteuerten Computers sind auch komplexe, arbeitsplatzorientierte Forschungsumgebungen realisierbar. Strenggenommen ist jedoch ein einzelner ferngesteuerter Computer (auch wenn er über das Internet kommuniziert) keine VRE, da VREs laut Definition webbasierte Umgebungen darstellen. Sie können jedoch in webbasierte VREs zweckmäßig eingegliedert werden, wodurch mancher Aufwand zur Umsetzung vergleichbarer webbasierter Services eingespart werden kann.

Im Anhang C in Tabelle C.4 befindet sich eine Übersicht zu geeigneter Software und Protokollen.

2.7.2 Cloud Computing

Cloud Computing (deutsch: Rechnen in der Wolke) ist ein Schlagwort, das momentan kontrovers in der IT-Branche diskutiert wird. Mit Cloud Computing lassen sich ICT-Infrastruktur und ICT-Leistungen als Service über das Internet mittels technischer Schnittstellen und Protokollen bereitstellen (Terzo und Mossucca 2015). Diese können

z.B. Speicherplatz, Rechenleistung oder Anwendersoftware sein, die nicht auf dem lokalen Rechner installiert sein müssen. In der nachfolgenden Übersicht finden sich die verschiedenen Typen von Cloud-Services.

Typen von verschiedenen Cloud-Services nach Terzo und Mossucca 2015 sind:

- PaaS – Platform as a Service,
- SaaS – Software as a Service,
- Public Cloud – öffentlich,
- Private Cloud – privat,
- Hybrid Cloud – hybrid (privat / öffentlich),
- Community Cloud – gemeinschaftlich,
- Virtual Private Cloud – privat innerhalb einer öffentlichen Cloud,
- Multi Cloud - Bündelung verschiedener Cloud Computing Services.

Viele Anbieter bringen neuerdings eigene Cloud-Services in ihren Softwarepaketen mit, um z.B. im Team gemeinsam an einer Datei zu arbeiten. Auch die neueste Generation der Windows-Betriebssysteme (Windows 10) integriert Cloud-Services im System, um z.B. auf jedem Windows-Gerät eines Benutzers die persönlichen Daten zu synchronisieren. Google bietet mit „DropBox“ einen verbreiteten Cloud-Service für Datenspeicherung. Zu erwähnen sind auch die Produkte „Microsoft Azure“ und „Amazon Web Services“ mit denen ganze Infrastrukturen und Plattformen als Service bereitgestellt werden können.

Mit Cloud-Services wächst jedoch auch die Angst vor Kontrollverlust über die eigenen Daten, da die Anzahl der Personen und Unternehmen, die potentiell Zugriff auf diese Daten haben, steigt und die „Cloud“ dadurch oftmals sehr undurchsichtig ist (Blokland et al. 2016). Der Schutz von Daten wird meist technisch durch zahlreiche Datenschutzmechanismen und institutionell über Data Policies versprochen - doch gerade bei Fremdanbietern insbesondere im Ausland kann man nicht sicher sein, wo die Daten gespeichert werden und wer sie ggf. einsehen kann. Daher empfiehlt es sich in Hinblick auf den Aufbau von VREs verschlüsselte Clouds und vertrauenswürdige Anbieter (z.B. das Rechenzentrum einer Institution) zu nutzen oder eigene Cloud-Services anzubieten, über die man die volle Kontrolle behalten kann. Drittanbieter stellen oft einfache und / oder kostenlose Cloud-Services zur Verfügung, wobei sie Daten beliebig sammeln können.

Cloud-Services eignen sich besonders durch ihre serviceorientierte Architektur (SOA) für den Aufbau von VREs in der Wissenschaft (Terzo und Mossucca 2015). Daher lassen sie sich Cloud-Services sehr vielseitig in VREs einsetzen. Zur eigenen Implementierung kann man sich durchaus der einschlägigen Open-Source-Software für die entsprechenden Cloud-Services bedienen. So lassen sich z.B. die Speicherung und Synchronisation von Daten auf verschiedenen Geräten und im Team, die gemeinsame Nutzung von Cloud-Software oder andere Cloud-basierte Infrastrukturen realisieren.

Die Cloud-Technologie wird im weiteren Verlauf der Arbeit für die zentrale Speicherung von Daten, Software, Services und Infrastrukturen im Rahmen des „Rapid-L-TER VRE“ Konzeptes verwendet.

2.7.3 Virtualisierung und virtuelle Maschinen (VM)

In der Informatik bezeichnet *Virtualisierung* die virtuelle Nachbildung von physikalischer Hard- und Software mit Hilfe eines Abstraktions-Layers für z.B. für den Einsatz von *virtuellen Maschinen* (VM) (Portnoy 2016; Gertsch und Walker 2015). Dies ermöglicht, *emulierte* (nicht physische) Geräte oder Dienste, wie z.B. Hardware, Betriebssysteme, Datenspeicher oder Netzwerk-Ressourcen virtuell zu generieren. Besonders im Server-Bereich lassen sich dadurch beispielsweise Computer-Ressourcen transparent zusammenfassen oder Betriebssysteme innerhalb eines anderen ausführen. Die Virtualisierung kann erhebliche Vorteile durch Zeiteinsparung und im Ressourcen-Management bieten. Einsparungen von physischer Hardware und Arbeitszeit haben in Hinblick auf die IT-Kosten und Green-IT-Maßnahmen (Umwelteffekt) ebenso eine deutliche Wirkung. Dabei gilt es die Vor- und Nachteile (Gertsch und Walker 2015) gegenseitig abzuwiegen, siehe Tabelle 2.4: Vor- und Nachteile von VMs.

Tabelle 2.4: Vor- und Nachteile von VMs

Vorteile	Nachteile
<ul style="list-style-type: none"> ▪ Mehrere Betriebssysteme / Server gleichzeitig auf einer Hardware ▪ Plattformunabhängigkeit ▪ Dynamische Optimierung und Skalierbarkeit ▪ Schneller Einsatz von schlüsselfertigen Serversystemen ▪ Einfaches Backup der Container-Datei ▪ Snapshots zur Wiederherstellung des Systems ▪ Einfaches Übertragen der VM auf neue Hardware ▪ Einsparung bei Hardwarekosten ▪ Green-IT (Umwelteffekt) 	<ul style="list-style-type: none"> ▪ Effizienzverlust der VM durch Hostsystem ▪ Treiberprobleme in Verbindung mit der virtualisierten Hardware, z.B. Grafikkarte ▪ Gegenseitige Beeinflussung von VMs ▪ Neue Herausforderungen bei Sicherheit und Datenschutz ▪ Neue Herausforderungen hinsichtlich der Lizenzierung von Betriebssystemen ▪ Wird die Container-Datei beschädigt, sind alle Daten einer VM auf einmal verloren

Virtuelle Maschinen laufen als Gast-System (engl. guest system) auf einem Wirt-System (engl. host system). Das Wirt-System stellt als Gastgeber die echte Hardware über den Abstraktions-Layer als virtuelle Hardware dem Gast-System zur Verfügung. Bei der Virtualisierung von Betriebssystemen mittels einer Container-Datei wird so die komplette VM in einer Datei auf dem Wirt abgespeichert. Es gibt auch optimierte Verfahren der Virtualisierung, für den Fall, dass mehrere ähnliche VMs auf einem Wirt-System laufen sollen. Dabei kommen z.B. LX-Container unter Linux zum Einsatz, die einmalig den Overhead einer gemeinsamen VM bereitstellen und die weiteren dedizierten VMs nur die jeweils benötigte Konfiguration und die jeweiligen Daten beinhalten

(Kumaran und Senthil 2017). Dieses Verfahren bietet eine deutlich höhere Performance und weniger erforderliche Speicherkapazität auf dem Wirt-System – ist dadurch aber auch wesentlich komplexer im grundlegenden Aufbau.

Die Technologien zur Virtualisierung stellen durch die beschriebenen Eigenschaften ein Schlüsselement für das in dieser Arbeit vorgestellte Konzept „Rapid L-TER VRE“ dar. Dies gilt besonders hinsichtlich der schnellen Entwicklung von Prototypen, der Einsparung von Hardware und der vereinfachten Administration (vgl. Kapitel 4.3).

2.7.4 Hochleistungsrechnen (HPC) und Grid-Computing

Hochleistungsrechnen (englisch high-performance computing, HPC) ist eine Form des *computergestützten Rechnens* (Raj et al. 2016, Nagel et al. 2016). Dazu gehören alle Rechenarbeiten, die eine hohe Rechenleistung und ggf. einen hohen Speicherbedarf im Vergleich zu einem üblichen Desktop Computer benötigen. Ermöglicht wird dies durch eine auf parallele Verarbeitung ausgerichtete Architektur der Hochleistungsrechner und Rechner-Cluster. Anwendungen sollten für die parallele Verarbeitung vorbereitet sein, damit sie diese Architektur in geeigneter Weise nutzen können. Für die parallele Programmierung kommen beispielsweise die Programmbibliotheken „OpenMP“ und „MPI“ oder die Autoparallelisierung im Compiler zum Einsatz.

Grid-Computing wird für Aufgaben des *verteilten Rechnens* eingesetzt (Preve 2013). Ein virtueller „Supercomputer“ entsteht dabei aus einem *Cluster* von lose gekoppelten Computern. Die *Grid-Architektur* wurde hauptsächlich zur Lösung rechenintensiver Aufgabenstellungen entwickelt, und findet heute breiten Einsatz in kommerziellen und wissenschaftlichen Bereichen, wie z.B. der Klimaforschung. Im Unterschied zu normalen Computerclustern ist beim Grid-Computing die Kopplung wesentlich freier, die Cluster heterogener und meist auch geographisch weiter verteilt. Grid-Systeme sind in der Regel auf eine spezielle Anwendung spezialisiert und nutzen dazu standardisierte Programmbibliotheken und Middleware, wie z.B. „gEclipse“ (Resch et al. 2008) sowie Systemarchitekturen auf der Basis der Open Grid Services Architecture (OGSA) (Joseph und Fellenstein 2004; toolkit.globus.org 2017).

Hochleistungsrechnen und Grid-Computing sind vor allem auch für wissenschaftliche Berechnungen von großer Bedeutung. Sie dienen als Hilfsmittel zur Berechnung, Modellierung und Simulation von komplexen Systemen und der dabei anfallenden Verarbeitung riesiger Datenmengen. Zu den typischen Anwendungsbereichen zählen beispielsweise die Meteorologie und Klimatologie, Astro- und Teilchenphysik, Systembiologie, Genetik, Quantenchemie und Strömungsmechanik. Viele dieser Anwendungsgebiete haben einen direkten Bezug zum Umweltbereich. Somit können HPC und Grid-Computing auch zum integrativen Bestandteil von L-TER VREs werden.

2.7.5 Content-Management-Systeme (CMS)

Content-Management-Systeme (kurz CMS) sind Software Systeme, die bei der Erstellung von Websites im Internet oder Intranet eingesetzt werden und deren Inhalte organisieren (Pausch 2017). Dabei wird der Inhalt meist nicht statisch, sondern dynamisch aus einer Webdatenbank zusammengefügt. Die meisten CMS erlauben das gemeinsame Management von Inhalten durch mehrere Personen und in verschiedenen Benutzerrollen mit angepassten Bearbeitungsrechten. Über eine grafische Benutzeroberfläche können die Funktionen der Website verwaltet und die Inhalte eingepflegt werden. Die meisten CMS lassen sich in ihren Funktionen durch eigene oder Komponenten von Drittanbietern erweitern. Die CMS bauen häufig auf der sogenannten Model-View-Controller-Architektur (MVC) auf. Dadurch ist die Programmierung von CMS systematischer, weil das Datenmodell, die Präsentation und die Steuerung der Website voneinander getrennt betrachtet werden. Das MVC-Konzept wurde 1979 für Benutzeroberflächen entwickelt (Seeheim-Modell) (Vu und Proctor 2011). Es gilt mittlerweile aber als De-facto-Standard für den Grobentwurf vieler komplexer Softwaresysteme (Dimoullis 2015).

Folgende grundlegende Funktionen sind in den meisten CMS integriert (Vincent 2017):

- Rollen- / Benutzer- Management (Anwender, Autor, Editor, Administrator),
- Benutzerschnittstellen (z.B. Editoren, RSS-Feeds),
- Redakteur System (Schreiben, Editieren und Veröffentlichen von Inhalten),
- Content Integration (Zusammenführung verschiedener Inhalte),
- Metadaten (Beschreibung der Daten),
- Arbeitsprozesse (Workflow),
- Layout Muster (Templating, Design der Website),
- Versionskontrolle (Inhalte und Software),
- Globalisierung (multilingual),
- Website-Erzeugung (dynamische Website-Generierung),
- Suchen (einfache und erweiterte Suchmechanismen),
- Personalisierung (auf den einzelnen Anwender zugeschnitten),
- Privilegierung (Rechtmanagement),
- Cache Management (Beschleunigung der Ladezeit von Websites),
- Syndication (Verbreitung der Website und deren Informationen),
- Cross-Media Publishing (Multimedia und Einbettung von Inhalten anderer Websites).

Content Management Systeme können nach unterschiedlichen Schwerpunkten typisiert werden, wobei die Trennung oft unscharf verläuft (Vincent 2017):

- **Web CMS** – Die Funktionalitäten dieser Systeme sind vorrangig auf die Erstellung und Verwaltung von Web-Inhalten ausgerichtet. Zu den Inhalten für die nach außen hin sichtbare Website (Frontend) zählen in der Regel Texte und multimediale Elemente zur Aufwertung und Attraktivitätssteigerung. Die Inhalte

können meist von mehreren Benutzern über die Administrationsoberfläche (Backend) verwaltet werden. Dabei kommen komplexe Freigabemethoden und deren Workflows für unterschiedliche Benutzergruppen zum Einsatz. Zu den Standardanforderungen gehören Funktionen, wie z.B. Mehrsprachigkeit, Volltextsuche und eine Newsverwaltung.

- **Communities / Social Publishing** – In erster Linie liegt hier der Fokus auf Online-Communities mit Web-2.0-Funktionen, bei denen der Funktionsumfang deutlich über das übliche Maß der Inhaltsbereitstellung für passive Benutzer hinausgeht. Die Nutzer können selbst aktiv Inhalte erstellen und mit anderen Benutzern interagieren und kommunizieren.
- **Blogs und News-Systeme** – sind in der Regel einfach gehaltene Systeme, die sich auf die schnelle Erstellung von häufig zu bearbeitenden Inhalten spezialisiert haben. Die Strukturierung der Inhalte über Kategorien mit zeitgesteuerter Veröffentlichung gehört zu den Basisanforderungen. Web-2.0-Funktionen für Interaktionen der Benutzer sind ebenso obligatorisch.

CMS sind daher für den Aufbau von VREs und als Basisplattform geeignet (vgl. Kapitel 4.4).

Tabelle C.5 im Anhang C gibt eine kurze Übersicht zu den am weitesten verbreiteten Open Source CMS.

2.7.6 Datenbanken und Datenbanksysteme

Datenbanksysteme (DBS) sind Systeme zur elektronischen Datenverwaltung von Daten in Datenbanken (Dehler et al. 2015). Ihre Aufgabe ist es, große Mengen an Daten effizient, widerspruchsfrei und dauerhaft zu speichern. Benutzer und Anwendungsprogramme können die Daten aus einem DBS in Teilmengen und in unterschiedlichen Darstellungsformen abfragen. Der Aufbau eines DBS besteht aus zwei Teilen. Für die Organisation der Datenbank ist die Verwaltungssoftware, das sogenannte Datenbank-Management-System (DBMS) zuständig (Gerken 2016). Den zweiten Teil stellen die zu verwaltenden Daten (Datenbasis) mit der Datenbank (DB) selbst dar. Das DBMS organisiert intern die strukturierte Speicherung der Daten und überwacht alle Lese- und Schreibzugriffe auf die Datenbanken. Für die Abfrage der Daten kommen in Abhängigkeit vom DBMS spezifische Datenbanksprachen zum Einsatz.

Es gibt verschiedene Datenbankmodelle für die Konstruktion von DBS. Sie unterscheiden sich z.B. in der Struktur, wie sie die Daten speichern und verwalten, im Referenz-Code zur Abfrage der Daten oder der Hierarchie im Netzwerk. Die gebräuchlichste Form sind die *relationalen Datenbanken*, bei denen einzelne Tabellen (Entitäten) in der Datenbank durch einen Index (Relation) miteinander durch ein *Entity-Relationship-Modell* verknüpft werden (z.B. Kunde x <bestellt> Artikel y).

Mit stetig steigendem Datenvolumen gewinnen seit einigen Jahren zunehmend auch andere Datenbankmodelle wieder an Bedeutung, die flexibler in ihrer Struktur sind und sich noch besser für Massen an unstrukturierten Daten eignen

Der allgemeine Begriff „Datenbank“ wird in der Praxis oft übergreifend für die hier beschriebenen Systeme verwendet. Folgend einige Beispiele zu Datenbankmodellen mit verschiedenen Ansätzen:

SQL-Datenbanken:

SQL ist eine Datenbanksprache für relationale Datenbanken (Schicker 2017). Die Abkürzung steht für *Structured Query Language* (auf Deutsch: "Strukturierte Abfrage-Sprache"). Mit ihrer Hilfe lassen sich Datenstrukturen erstellen, Daten bearbeiten, Zugriffsrechte setzen und Abfragen auf die Datenbestände realisieren. SQL basiert auf der relationalen Algebra, wobei eine relativ einfache und semantisch an die englische Umgangssprache angepasste Syntax den Umgang erleichtert.

OQL-Datenbanken:

Die *Object Query Language* (OQL) ist eine Abfragesprache für Objektdatenbanken mit Anlehnung an SQL und wurde entwickelt, um die Interaktion zwischen objektorientierten Programmen und einer Datenbank zu vereinfachen (Erbs et al. 2003).

NoSQL-Datenbanken:

NoSQL-Datenbanken (engl. für Not only SQL) brechen die lange Tradition der relationalen Datenbanken durch ein skalierbares Datenmodell ohne festgelegte Tabellenschemata und Verknüpfungen (Ahlf 2016). Sie eignen sich besonders gut für große Mengen an unterschiedlichen Daten und sind bei häufigen Schreib- und Lesezugriffen schneller. Im akademischen Umfeld werden diese Datenbanken oft als *Strukturierter Speicher* (engl. structured storage) bezeichnet, weil die Daten dynamisch ihre eigenen Strukturen „mitbringen“. Zu Beginn der Datenbankplanung muss daher keine spezielle Struktur festgelegt werden, was spätere Anpassungen deutlich erschweren würde. NoSQL-Datenbanken sind dadurch besonders flexibel, skalierbar und performant für Massendaten (Deka 2016). Jedoch muss bei der Programmierung der Datenabfrage und Gültigkeitsprüfung aufgrund der anfänglichen Einsparung einer Struktur etwas mehr Aufwand betrieben werden. Dies geht jedoch nicht zu Lasten der Bedienerfreundlichkeit bei Endbenutzern. Auch bei verteilten Datenbanken für Redundanz und Lastenausgleich haben NoSQL-Architekturen große Stärken. Jedoch bieten sie meist nur „schwache“ Garantien hinsichtlich ihrer Konsistenz, wie beispielsweise *eventual consistency* oder auf einzelne Datensätze eingeschränkte Transaktionen (Deka 2016).

Eine VRE kommt normalerweise nicht ohne Datenbanken und Datenbanksysteme aus. Zum einen müssen Daten gespeichert werden, die zum Betrieb des Systems (der

VRE) erforderlich sind und zum anderen sollten die wissenschaftlichen Daten in Datenbanken gespeichert werden und nicht in einzelnen Dateien (z.B. MS-Excel Tabellen). Der Zugriff auf wissenschaftliche Daten über Datenbanksysteme ist ratsam, damit der Zugriff und die Datensicherung zentral, schnell und einheitlich erfolgen können und im Team so kein „Versions-Chaos“ entsteht.

Meist erfordern einzelne Komponenten einer VRE spezifische Datenbanksysteme, so dass verschiedene DBS im Rahmen einer VRE verwendet werden. Für die klassischen Komponenten, wie z.B. Content-Management-Systeme für dynamische Websites, Benutzerdatenbanken und Moduldatenbanken werden vorzugsweise relationale DBS eingesetzt. Für georeferenzierte Daten eignen sich besonders SQL-basierte DBS, die für GIS-Funktionalitäten erweitert sind, wie z.B. „PostgreSQL“ mit „PostGIS“ Erweiterung. NoSQL-Datenbanken können für die Flexibilität von VREs bei der Speicherung von Forschungsdaten von Vorteil sein, da sie besser für in längerfristigen Projekten wechselnde Datenstrukturen und unstrukturierte Massendaten geeignet sind, wie zuvor beschrieben. Eine Analyse der vorhandenen Daten und eine Vorausplanung sind erforderlich, um sich je nach Anforderungen für ein geeignetes Datenbankmodell mit einem passenden DBS zu entscheiden. Tabelle C.2 im Anhang C zeigt beispielhaft einige DBMS-Softwareprodukte zu den Datenbankmodellen.

2.7.7 Dataspace- und Semantic-Web-Modelle

Die Verbindungen und Beziehungen von Daten in einer VRE untereinander werden durch *Datenraummodelle (Dataspace-)* (Borgida 2009; Elsayed 2012) und *Semantik-Web-Modelle* (Devedžić und Gašević 2009; Hřebíček et al. 2000; Rosina 2016; Wu und Chen 2008) strukturiert. Dataspace-Modelle folgen dem Ansatz der physischen oder virtuellen Informationsintegration in einem *Datenraum* (Franklin et al. 2005). Der Erstellungsaufwand einer Integrationsstruktur soll möglichst gering ausfallen – es wird daher vorab keine feste Zielstruktur benötigt. Diese Basis bietet Möglichkeiten, die Datensouveränität einzelner Quellen zu gewährleisten, da die Daten bis zu ihrem Aufruf im Wirkungsbereich des Datenurhebers verbleiben können (Elsayed 2012). Semantik Web gehört in der „Web-Versionshistorie“ zu den Schwerpunkten von „Web 3.0“ (Hitzler 2008).

Dieser Ansatz eines verteilten Datenraummodells in Kombination mit semantischen Beziehungen zwischen Datensätzen, Forschern und Institutionen ist besonders für die Wissenschaft und *open linked data* (Volk 2011) geeignet, um Daten zu strukturieren, zu organisieren und für die Langzeitforschung persistent (Neuroth et al. 2012) zu speichern (Elsayed 2012). In diesem Zusammenhang kommt die *Semantic Web* Technologie mit dem *Resource Description Framework (RDF)* Datenmodell (Zächerl 2007) und der *Web Ontology Language (OWL)* (Liebig 2013) zur formalen Definition von *Ontologie* zum Einsatz. Ressourcen werden in diesem Kontext weltweit eindeutig durch den *Uniform Resource Identifier (URI)* gekennzeichnet. Informationen zwischen Rechnern sind

somit einfacher austauschbar, da z.B. zu dem Begriff „Kohl“ die explizite Information gespeichert ist, ob es sich um die Nahrungspflanze oder den ehemaligen Politiker handelt.

2.7.8 Serviceorientierte Architektur (SOA)

Der Begriff *serviceorientierte Architektur* (SOA, englisch: service-oriented architecture) basiert auf einem dienstorientierten Architekturmuster aus der Informationstechnik für verteilte Systeme (Bengel 2014). Das Ziel ist hierbei, ICT-Komponenten wie z.B. Datenbanken, Server und Websites zu strukturieren und als Dienste zu kapseln, um dadurch koordiniert ihre Leistungen zu höheren Diensten zusammenzufassen (Oey 2012). Flexibilität und Kostenersparnis soll hier durch die Wiederverwendung der einzelnen Services erreicht werden (Dietzel et al.). Eine strikt umgesetzte SOA erfordert tiefgreifende Integration der einzelnen ICT-Komponenten und sollte schon bei der Auswahl der Software berücksichtigt werden. Das technische Resultat einer SOA ist die Bereitstellung von Diensten über das Internet oder in der Cloud. Üblicherweise erfolgt die Kommunikation zwischen den Diensten über Protokolle, wie SOAP (Weerawarana 2008), REST (Massé 2012), XML-RPC (Wenz und Hauser 2004) oder ähnliche Schnittstellen. Darum muss sich der Benutzer jedoch nicht kümmern. Er weiß nur, ob ein entsprechender Dienst angeboten wird, welche Eingabeparameter erforderlich sind und welches Ergebnis dabei herauskommt. Services mit ihrer SOA gehören in der „Web-Versionshistorie“ zu den Schwerpunkten von „Web 2.5“ (Hitzler 2008).

VREs bestehen per Definition (vgl. Kapitel 3.1) meist aus einer Vielzahl von unterschiedlichen Services. Daher kann für VREs eine SOA nützlich sein, um verschiedene Aufgaben als Services zu strukturieren und wiederverwendbar einzusetzen.

2.7.9 Authentifizierung und Single Sign-On (SSO)

VREs bestehen aus mehreren einzelnen Services, von denen meist jeder Informationen über Benutzer-Accounts und deren Zugriffsrechte auf Daten bei der Authentifizierung von Benutzern benötigt. Sogenannte *Verzeichnisdienste* sind für VREs daher von zentraler Bedeutung, weil sie den Administrationsaufwand und die Benutzerfreundlichkeit in Zusammenhang mit der Authentifizierung von Accounts bündeln und somit umso effizienter sind, je mehr Benutzer das System beinhaltet.

Ebenso ist die Anmeldung durch Single Sign-on (SSO, auch als „Einmalanmeldung“ übersetzt) unerlässlich für die Akzeptanz der Benutzer und den Erfolg des gesamten Systems (vgl. Kapitel 2.5.1). Bei einer Vielzahl von VRE-Services (vgl. Kapitel 3.1) kann es Benutzern schwerlich zugemutet werden, sich ständig neu zu authentifizieren, um einzelne Komponenten des Systems nutzen zu können.

Im Folgenden werden diese Mechanismen näher beschrieben:

Verzeichnisdienste (engl. „directory services“) stellen im Netzwerk hierarchisch gegliederte Strukturen zur Speicherung von Daten einer bestimmten Art zur Verfügung (Larisch 2000; Klünter und Laser 2008). Dabei handelt es sich nicht um die „Dateiordner“ in einem Computer – mit „Verzeichnis“ ist beispielsweise ein Mitarbeiterverzeichnis oder Telefonbuch gemeint. Der Verzeichnisdienst basiert auf dem Client-Server-Prinzip, wobei Daten gesucht, verglichen, erstellt, geändert und gelöscht werden können. In größeren Institutionen werden Verzeichnisdienste oft dazu eingesetzt, Daten von Mitarbeitern und deren Accounts zentral zu sammeln und zu pflegen, um sie weiteren Applikationen zur Verfügung zu stellen. So müssen Änderungen nicht mehrfach eingegeben werden, wie es bei einer dezentralen Systemarchitektur der Fall wäre. Da bei Verzeichnisdiensten überwiegend Lesezugriffe stattfinden, sind sie für diesen Modus im Vergleich zu anderen Datenbanken hinsichtlich der Lesegeschwindigkeit optimiert. Im Anhang C in Tabelle C.6 findet sich ein Überblick zu oft genutzten Verzeichnisdiensten.

Single Sign-on ist eine Technologie, die es Benutzern nach einmaliger Anmeldung mit Authentifizierung an einem Gerät ermöglicht, verschiedene Services und Ressourcen zu nutzen, ohne sich dabei jedes Mal neu anmelden zu müssen (Tietz et al. 2017; Strecker 2010). Wechselt der Benutzer jedoch das Gerät oder überschreitet eine vorgegebene Zeit, ist eine erneute Anmeldung erforderlich. Der SSO Mechanismus übernimmt dabei die Aufgabe, den Anwender zu authentifizieren (die erkannte physische Identität zu verifizieren), z.B. in Verbindung mit einem Verzeichnisserver der Benutzername und Passwort verwaltet. Weiterhin kann er an die untergeordneten Services und Netzwerke die Identität des angemeldeten Benutzers durch ein Token (Methode zur Autorisierung von Services) vererben, ohne dass weitere Informationen zum Anwender selbst bekannt gegeben werden müssen (Datenschutz). Dabei darf das SSO-Verfahren von der Sicherheit her nicht schlechter sein, als das Authentifizierungsverfahren selbst (Feldmann 2015). Das kann z.B. auf der Basis der Security Assertion Markup Language (SAML) erfolgen (shibboleth.net 2017), die ein XML-Framework zum Austausch von Authentifizierungs- und Autorisierungsinformationen bietet. Im Anhang C in Tabelle C.7 findet sich ein Überblick zu oft genutzten SSO-Produkten.

2.7.10 Datenverbünde und Metadatenharvesting

Die Open Archives Initiative (OAI) (Open Archives Initiative 2017) ist eine Initiative, die den Austausch von Metadaten über eine offene Schnittstelle zum Ziel hat. Dazu werden die Daten über einen Datenverbund von einem Datenlieferant (*Data Provider*) bereitgestellt und durch einen Service Anbieter (*Service Provider*) gesammelt und weiterverarbeitet. Das hierbei zum Einsatz kommende „OAI Protocol for Metadata Harvesting“ (OAI-PMH) wurde im Jahre 2000 entwickelt und dient seither z.B. im Bibliothekswesen zum Austausch von Metadaten und dem Aufbau von großen übergreifenden Katalogen (Open Archives Initiative 2017). Es wurde auf der Basis von XML (Goldfarb und Prescod 2000) und REST (Massé 2012) implementiert. Aufgrund der

vielen unterschiedlichen Metadatenformate ist als kleinster gemeinsamer Nenner das *Dublin-Core-Datenmodell* (Lynch 1998) erforderlich. Mit dem Protokoll können auch andere Daten übertragen werden. Es wird inzwischen von einer wachsenden Anzahl an Institutionen unterstützt. Das folgende Beispiel (Abbildung 2.10) zeigt eine Netzwerkstruktur zum interoperablen Austausch von Daten unterschiedlicher Netzwerkknoten.

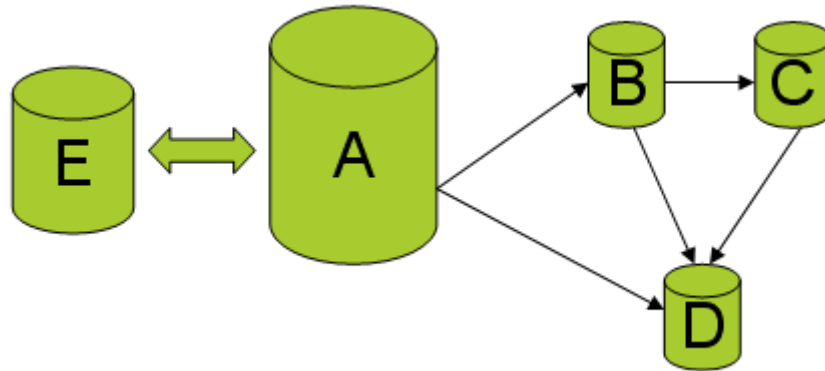


Abbildung 2.10: Beispiel einer Netzwerkstruktur für Datenverbünde mit Netzwerkknoten

1. Knoten (A) erzeugt Metadaten (a), (B,b; C,c)
2. Knoten (B) bezieht (a) von (A)
3. Knoten (C) bezieht (a) von (B)
4. Knoten (D) bezieht (a,b,c) von (A), (B) und (C)
5. Knoten (E) und (A) kommunizieren bidirektional

In VREs ist der Austausch von Daten ein wichtiger Prozess. Durch Datenverbünde z.B. auf der Basis von OAI-PM lassen sich Daten mit anderen Netzwerken oder VREs interoperabel austauschen. Dies kann in einer Netzwerk-Hierarchie erfolgen, in der die Verantwortlichkeiten für die Daten und die Datenhoheit (Copyright, Nutzung) durch die Netzwerk-Struktur gegeben ist und z.B. nur die Metadaten ausgetauscht werden (vgl. Abbildung 2.10). Auch die Wahrnehmbarkeit der VRE mit ihren Daten und Projekten in der Öffentlichkeit kann dadurch erhöht werden. Weiterhin sollen die Geodaten des Bundes, der Länder und der Kommunen (also ggf. auch die Daten in einer VRE) nach den Vorgaben interoperabel verfügbar sein bezüglich INSPIRE und GDI (vgl. Kapitel 2.6). Als Software kann z.B. „GeoNetwork – Open Source“ (geonetwork-opensource.org 2017) auf der Basis von OAI-PM zum Einsatz als Komponente in einer interoperablen VRE kommen (vgl. Kapitel 5.1.3).

2.8 Forschungsdaten in L-TER Projekten

Forschungsdaten sind ein wesentlicher Bestandteil von VREs. Von der Art der Daten ist es abhängig, welche Eigenschaften und Funktionen in VREs bereitgestellt werden müssen. Im Folgenden wird daher näher auf die Eigenschaften der Forschungsdaten in L-TER Projekten eingegangen. Hierfür werden eine Klassifizierung von Daten vorge-

nommen sowie unterschiedliche Zeitskalen und Prozessebenen betrachtet. Datenlebenszyklen helfen, bei der Analyse von Daten für den jeweiligen Betrachtungszeitraum geeignete Methoden zu entwickeln und einzusetzen. Um die Langzeitdynamik von Ökosystemen zu untersuchen, sind Werkzeuge zur Datenaufbereitung und zur Datenanalyse erforderlich. An dieser Stelle kommt besonders die interdisziplinäre Rolle der Umweltinformatik zum Tragen. Metadaten, Ontologie und DataSpace-Modelle unterstützen die Strukturierung und Organisation der Forschungsdaten. Verschiedene Datenmodelle müssen miteinander verbunden werden, wobei Standards zur Datenkopplung eingesetzt werden.

2.8.1 Klassifizierung von Daten

Die Klassifizierung von Daten vereinfacht die Erstellung und Wiederverwendung von Datenstrukturen gemäß ihrer Eigenschaften. Dazu sind verschiedene Klassifizierungskriterien möglich, die nachfolgend kurz erläutert werden. Aus der Klassifizierung lassen sich besonders gut Schlussfolgerungen für die Datenverarbeitung ziehen, wie z.B. Speicherplatzbedarf, erforderliche Systemperformance oder Automatisierungsvorgänge auf Basis von homogenen Datenstrukturen. Diese Einteilung ist jedoch zu oberflächlich, wenn es um den Inhalt der Dateien geht. Hier ist die Betrachtung nach Datenherkunft sinnvoller.

Einteilung nach Struktur der Daten:

- **Unstrukturierte Daten** sind digitale Daten, die in einer nicht formalisierten Struktur vorhanden sind. Die automatische Nutzbarkeit unstrukturierter Daten ist eingeschränkt, da für sie kein *Datenmodell* (Franklin et al. 2005) und meist keine Metadaten vorliegen.
- **Strukturierte Daten** erhalten ihre *Datenstruktur* (abstrakter Datentyp) durch eine exakte Beschreibung zur Datenhaltung und der dazu erforderlichen Operationen. Grundlegende Datenstrukturen sind z.B. Datensätze (z.B. Tupel), Arrays, verkettete Listen, Stapelspeicher, Graphen und Baumstrukturen (Dehler et al. 2015).

Für die jeweiligen strukturierten oder unstrukturierten Datenformen eignen sich unterschiedliche Datenbank-Management-Systeme (DBMS), die nach diesem Kriterium (Datenstruktur) ausgewählt werden können (vgl. Kapitel 2.7.6). Die Wahl der Datenstrukturen (vgl. Kapitel 2.7.7) und des DBMS bereits *in der Planungsphase* einer VRE hat Einfluss auf die Flexibilität, Nachhaltigkeit, und den Erfolg einer VRE.

Einteilung nach Datenmenge:

- **Big Data:** Besonders große Datenmengen, wie sie z.B. bei der Verarbeitung von Satellitenbildern oder kontinuierlichem Umwelt Monitoring entstehen, werden

als *Big Data* bezeichnet (García Márquez und Lev 2017). Sie lassen sich mit Hilfe von Standard-Datenbanken oder -Management-Tools aufgrund ihrer Größe nur teilweise oder gar nicht im Arbeitsspeicher verarbeiten. Das Datenvolumen geht üblicher Weise in die Größenordnung von Terra-, Penta- oder Exabytes (Hey et al. 2009; Qiu 2017). In der Ökologie werden beispielsweise Satellitenbilder mit unterschiedlichen Spektren in hoher Auflösung zur Bestimmung und Analyse von Vegetation und Habitaten herangezogen (Woodward 2014). Dabei können die hochauflösenden Aufnahmen pro Bild mehrere hundert Megabyte benötigen. So sammeln sich große Datenmengen über Jahre an. Um Big Data auszuwerten, bedarf es spezieller Technologien des *Data Mining* (Pedrycz und Chen 2017).

- **Small Data:** Small Data sind in der Regel kleine, unstrukturierte Daten und einzelne Dateien, die individuell bei der täglichen Arbeit anfallen, wie Excel-Tabellen unterschiedlicher Form, Textdokumente, Datenauszüge, Statistiken, Berichte, Studien, einzelne Fotos, etc. Das Datenvolumen bereitet hierbei bei der Verarbeitung mit Standard-Tools weniger Probleme. Jedoch ist es aufgrund ihrer Heterogenität schwer, die Daten zusammen zu fügen und durch ein zentrales System zu verarbeiten. Das Datenvolumen der einzelnen Dateien liegt normalerweise im Bereich von Kilo- und Megabytes.
- **Medium Data:** Zwischen Big Data und Small Data besteht ein fließender Übergang, der sich als Medium Data bezeichnen lässt. Hier können sowohl größere Datenmengen aus stark schwankenden Dateigrößen und inhomogene sowie homogene Datenstrukturen gemischt auftreten.

L-TER Projekte und L-TER VREs bewegen sich aufgrund des Langzeitaspektes und des Umfangs von Forschungsdaten eher im Bereich von Medium bis Big Data.

Weiterhin lassen sich Daten nach Ludwig und Enke (2013) auch hinsichtlich ihrer inhaltlichen Art, sowie nach Art ihrer Erstellung und Erfassung klassifizieren. Dies hat den Vorteil, dass ihre Bewertung hinsichtlich einer Entscheidungsfindung leichter möglich ist. Dabei ist es zunächst unerheblich, welche technischen Datentypen eingesetzt werden:

- **Beobachtungen und Messungen:** Diese Form der Daten repräsentiert spezifische Phänomene zu einer bestimmten Zeit oder an einem bestimmten Ort, wie sie z.B. bei Umwelt Monitoring entstehen. Sie sind in der Regel ein einzigartiges Ereignis und lassen sich nicht wiederholen.
- **Experimente:** Hierbei handelt es sich um Daten, die direkt aus Labor- oder Feldexperimenten mit definierten Randbedingungen und Parametern gewonnen werden. Prinzipiell lassen sich die Daten erneut herstellen, was sich in der Praxis bei Feldexperimenten jedoch als schwierig erweisen kann, da nicht alle Rahmenbedingungen exakt gleich sind.

- **Modelle / Simulationen:** Die Daten stammen hier aus der Berechnung von Modellen und Simulationen und sollten reproduzierbar sein. Dazu gehören auch die Parameter der Modelle oder der Simulationen sowie deren Randbedingungen und Methoden.
- **Abgeleitete Daten:** Durch die Verarbeitung von Rohdaten in Verbindung mit anderweitig zur Verfügung stehenden Daten werden durch spezifische Verfahren neue Daten erzeugt. Hier ist die Herkunft und Dokumentation der Methoden und Daten besonders relevant.
- **Metadaten:** Metadaten sind allgemein Daten über Daten (vgl. Kapitel 2.8.5).
- **Kanonische oder Referenzdaten:** Diese Datenform beschreibt eine eindeutige Darstellung von Daten (z.B. das Datum oder eine Telefonnummer) anhand von gemeinsamen Regeln (lat. canon – „Norm, Regel“) oder „übersetzt“ Daten in eine andere Form bzw. Standardformat, wobei der Übergang zu Metadaten hierbei fließend ist.

2.8.2 Zeitskalen und Prozessklassen für ökologische Variable

Die Dynamik von Ökosystemen wird stark beeinflusst durch die Abhängigkeit von Randbedingungen. Umweltfaktoren, wie z.B. die mittlere Temperatur oder der mittlere Niederschlag, verändern sich sowohl in geologischen Zeiträumen als auch in kleineren Rhythmen von Dekaden, Jahren und Tagen (Gnauck 2001; Müller et al. 2010a). Effekte, die das Klima beeinflussen, spielen sich z.B. auf einer kurz- bis mittelfristigen Zeitskala ab (vgl. Kapitel 2.3). Ökosysteme reagieren nicht linear und verzögert auf diese Einflüsse. Um Langzeitfolgen abschätzen zu können, genügt es nicht, Daten aus kurzen Zeiträumen zu extrapolieren (Müller et al. 2010a). Eine Langzeitbeobachtung ist daher erforderlich zur Untersuchung von Einwirkungen und deren potentiellen Folgen. Typische Fragestellungen in Bezug auf den Klimawandel sind beispielsweise die Folgenden (Müller et al. 2010a):

- Wie reagieren verschiedene Spezies langfristig auf einen Temperaturanstieg?
- Wie werden die Verteilungsmuster der Individuen und Populationen beeinflusst?
- Welche Spezies sind evtl. vom Aussterben bedroht?
- Was sind die Folgen des Aussterbens einer Spezies in Bezug auf die Stoff- und Energiekreisläufe und Speicherkapazitäten des Ökosystems?
- Wie verändert sich die Resilienz eines gesamten Ökosystems?
- Was wird die Folge für den Menschen sein, der die Ökosystem-Dienste in Anspruch nimmt?

Für die Untersuchung von zeitlichen Zusammenhängen in Ökosystemen ist nicht nur eine Einteilung in Form von Zeitskalen entscheidend. Auch die Prozesse lassen sich einteilen, z.B. nach der Geschwindigkeit von Prozessen, seltenen und wiederkehrenden

Prozessen, Prozesse mit hoher Variabilität oder Prozesse mit besonders komplexen Interaktionen (Dröschmeister 1998).

In Tabelle 2.5 sind die Methoden der L-TER-Forschung anhand verschiedener Prozessklassen mit Beispielen aufgeführt (Dröschmeister 1998).

Tabelle 2.5: Prozessklassen der L-TER Forschung mit Beispielen, nach Müller et al. (2010a)

Prozess Klasse	Beispiel der Beeinflussung
Langsame Prozesse	<ul style="list-style-type: none"> ▪ Populationsdynamik Methusalem-Organismen (Arnaud-Haond et al. 2012) ▪ Sukzession ▪ Änderung auf der Ebene einer Spezies ▪ Einwanderung invasiver Spezies ▪ Bodenentwicklung ▪ Ökochemische Abbauprozesse ▪ Ökotoxikologische Effekte ▪ Chronische Effekte ▪ Eutrophierung ▪ Effekte des Klimawandels ▪ Effekte durch Änderung der Landnutzung ▪ Renaturierungsmaßnahmen ▪ Studien zur Naturgeschichte
Prozesse mit seltenen und wiederkehrenden Ereignissen	<ul style="list-style-type: none"> ▪ Gradation ▪ Seuchen und Krankheiten ▪ Predator Abundanzen ▪ Konsequenzen von extremen Wetterereignissen (Trockenheit, Flut, Erosion, Feuer, Wind, etc.) ▪ Zyklische Populationsschwankungen
Prozesse hoher Variabilität	<ul style="list-style-type: none"> ▪ Rauschen in Zeitreihen ▪ Trendvorhersage von variablen Prozessen ▪ Resilienz, Elastizität, Pufferkapazität (Ableitung von Stabilitätsfaktoren) ▪ Entwicklung räumlicher Strukturen
Prozesse mit komplexen Interaktionen	<ul style="list-style-type: none"> ▪ Meta-Populationsdynamik ▪ Verhaltenswechsel von Organismen ▪ Genetische Unterschiede ▪ Extinktionen ▪ Konsequenzen von Fragmentierungen ▪ Mosaik-Dynamiken ▪ Ökoton-Dynamiken

2.8.3 Datenaufbereitung

Für ökologische Forschungsprojekte werden Daten aus verschiedensten Quellen zusammengetragen. Diese Daten müssen meist für den neuen Kontext aufbereitet werden. Eine L-TER VRE sollte daher Aspekte zur Datenaufbereitung integrieren, indem vordefinierte und individuelle, technische Funktionen für die Datenaufbereitung bereitstellt

werden. Da oft Zeitreihen (Müller et al. 2010a) in L-TER Projekten eine Rolle spielen und dies ein Forschungsschwerpunkt des ehemaligen Lehrstuhls „Ökosysteme und Umweltinformatik“ an der BTU Cottbus-Senftenberg gewesen ist, wird im Folgenden darauf beispielhaft eingegangen. Eine andere Form der Datenaufbereitung erfordert z.B. die Aufbereitung von Bildmaterial aus Fernerkundungen oder von räumlichen Daten.

Oft handelt es sich bei Zeitreihen um quantitative Messdaten, die mit Messfehlern behaftet sind. Die charakteristischen Eigenschaften dieser Daten weisen meist heterogene Datenstrukturen verbunden mit verschiedenen Unsicherheitsfaktoren auf (Recknagel 2003). Das Problem dabei sind die verschiedenen Datenstrukturen /-formate, Genauigkeitsfaktoren, Sampling-Rates, fehlende Meta-Beschreibungen, subjektive Bewertungsmethoden und urheberrechtliche Aspekte, die bei der weiteren Datenaufbereitung und Verwendung der Forschungsdaten zu beachten sind (Recknagel 2006; Müller et al. 2010a; Müller et al. 2010b; Luther und Gnauck 2012).

Faktoren der Genauigkeit von Daten sind zum Beispiel: Das Auftreten von zufällig verteilten Variablen, unvollständigen bzw. fehlerbehafteten Datensätze, ungenaue Schätzungen anstelle von genauen Messungen, unvergleichbare Datensätze (z.B. von verschiedenen Beobachtungsbedingungen) (Müller et al. 2010a; Recknagel 2006).

Darüber hinaus sind nicht alle ökologischen Daten direkt oder mit vertretbarem Aufwand messbar, wie z.B. der Zustand oder die Resilienz (Müller et al. 2016) eines gesamten Ökosystems oder die genaue Anzahl von Fischen in einem Fluss. Daher müssen die gewonnenen Daten im *Pre-* und *Postprocessing* aufbereitet und bewertet werden (Müller et al. 2010b; Luther und Gnauck 2012). Ökologische Forschungsprojekte beinhalten meist viele Teildisziplinen mit verschiedenen Experten, die ihre Ergebnisse miteinander abstimmen und zusammentragen müssen.

Fehlende Daten im Langzeit-Monitoring führen zu generellen Problemen für die Forschung und Simulation. Sie führen zu Schwierigkeiten bei der Identifizierung von Prozessen, der Parameterschätzung sowie der räumlichen und zeitlichen Interpretation von Variablen (Müller et al. 2010a; Gnauck und Luther 2006). Die numerische Verarbeitung von Datensätzen mit Datenlücken ist oft nur unzureichend in Computerprogrammen umgesetzt. Nach Little und Rubin (2014) können fehlende Daten eingeteilt werden in:

- Kompletter Datenausfall, zufällig verteilt
- Teilweise Datenlücken, zufällig verteilt
- Nicht zu ignorierende Datenlücken

Mathematische Verfahren sind dazu geeignet, die Datenlücken zu kompensieren. Dazu zählen z.B. Glättungsverfahren, Interpolationen, Trendfunktionen und Referenzkurven (Müller et al. 2010a; Pollock et al. 1999). Der Erfolg dieser Methoden kann im Einzelfall recht unterschiedlich ausfallen. Bei der Interpolation von unregelmäßigen Messdaten werden Zeitreihen mit gleichem Sampling Intervall begünstigt. Approximationsmethoden liefern hier bessere Ergebnisse (Müller et al. 2010a). Eine weitere Me-

thode ist das *Resampling* von Daten um Störungen und „Ausreißer“ zu identifizieren und auf einer Zeitskala zu harmonisieren. Sie setzt eine vorige Interpolation der Daten voraus. Folgende Möglichkeiten ergeben sich für das Resampling der Daten (Müller et al. 2010a; Luther und Gnauck 2012):

- Füllen von Datenlücken durch Interpolation,
- Datenrekonstruktion mit analytischen Funktionen und Signalschätzungen,
- Füllen von Datenlücken mit Hilfe von Referenzkurven und -Messwerten.

Mathematische Gleichungen für ökologische Prozesse können eine Zeit- oder Frequenzabhängigkeit der Variablen (Funktion der Zeit oder der Winkelgeschwindigkeit) beschreiben (Pollock et al. 1999). Um diese Daten aufzubereiten, werden beispielsweise Regressionsanalysen (Gnauck und Straškraba 1985) oder im Fall von zyklischen Ereignissen die Fourier Approximation (Alegue und Gnauck 2006) genutzt. Das Resampling kann durch digitale Signalfilter ergänzt werden, wie z.B. durch Hoch- und Tiefpassfilter (Gnauck 2006). Abbildung 2.11 verdeutlicht den Resampling Prozess (Müller et al. 2010a).

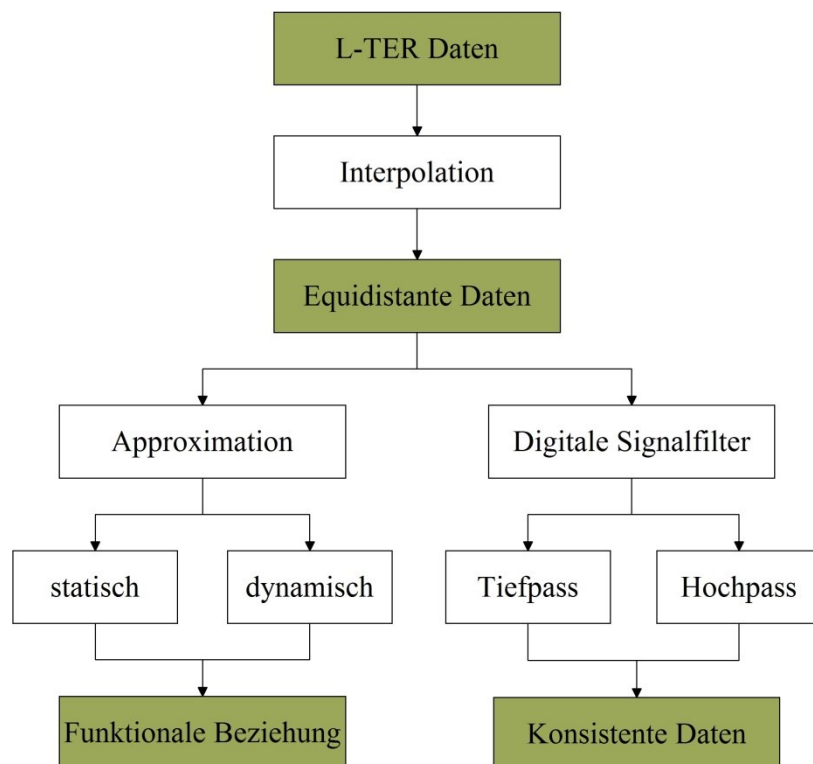


Abbildung 2.11: Resampling Prozedur nach Gnauck (2006)

Um aus Zeitreihen mit verschiedenen Sampling-Intervallen gleichverteilte Daten zu gewinnen, werden die Methoden zur Interpolation ebenso angewendet. Nach Müller (Müller et al. 2010a) werden aufgrund der *zufälligen Verteilung* von Zeitreihen keine R^2 Statistik angewendet, da die Daten meist nicht *normal verteilt* sind. Die Eignung von Interpolationsmethoden wird bewertet durch den Standard Fehler $SE = s/\sqrt{n}$ (s :=Standardabweichung einer einzelnen Messung, n := Stichprobenumfang) zwischen

der interpolierten und der originalen Zeitreihe. Meistens seien die linearen Methoden dafür besser geeignet, als Methoden wie z.B. „Nächste Nachbarn“ (engl. nearest neighbour), polynomische oder kubische Spline Funktionen. Für die Parameterschätzung eigne sich beispielsweise die Methode der „kleinsten Quadrate“ oder „Maximum Likelihood“. Nützliche Prozeduren sind außerdem multivariate Regressionsmodelle, Polynomische Modelle oder time-diskrete Transfer Funktionen (Müller et al. 2010a; Gnauck 2004; Gnauck 2006). Die Bewertung erfolgt mittels linearer oder nicht-linearer Koeffizienten für die Vorhersage. Ebenso eignen sich *Wavelets* zur Zeitreihenanalyse, da sie nicht nur das Frequenzspektrum, sondern auch den Zeitbereich berücksichtigen (Gnauck et al. 2010; Gnauck 2004).

2.8.4 Datenlebenszyklus

In der Literatur findet man eine Vielzahl von Lebenszyklus-Modellen für digitale Informationen, wobei sich das Digital Curation Center (DCC) Life-Cycle-Model besonders für Forschungsdaten etabliert hat (Higgins 2008). Im Folgenden soll auf ein vereinfachtes Modell nach Ludwig und Enke (2013) eingegangen werden (Abbildung 2.12). Es ist ebenso für L-TER VREs geeignet, wobei der Nutzungszeitraum durch den Langzeitaspekt größer ist, als bei vielen anderen Projekten.

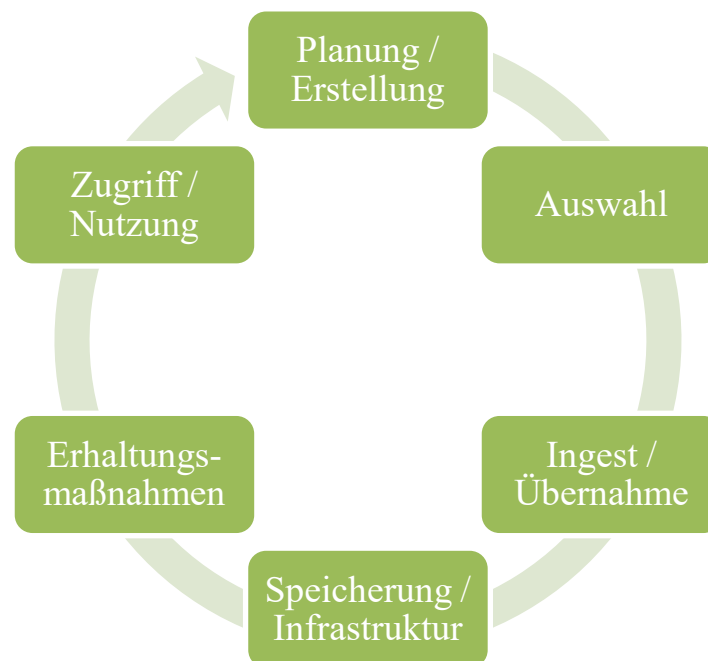


Abbildung 2.12: Lebenszyklus von Forschungsdaten nach Ludwig und Enke (2013)

Folgende Stadien im Lebenszyklus der Forschungsdaten müssen Berücksichtigung finden (Ludwig und Enke 2013):

- **Planung und Erstellung:** Welche Daten werden erhoben und wie werden sie gesammelt? Welche Methoden und welche zeitliche Auflösung werden genutzt?

Um das Management der Daten später zu vereinfachen, ist es bereits in dieser Phase nützlich, die Daten gemäß den Anforderungen in einem geeigneten Format zu speichern und dazu die richtigen Standards zu wählen (vgl. Kapitel 2.5.3).

- **Auswahl:** Müssen oder können alle Forschungsdaten auf Dauer gespeichert werden? Hierbei muss aus den jeweiligen Kriterien die Dauer der Aufbewahrung von Forschungsdaten bestimmt werden (vgl. Kapitel 2.5.7).
- **Ingest / Übernahme:** Welche Probleme bei der Datensammlung machen eine Aufbereitung der Daten (vgl. Kapitel 2.8.3) erforderlich, bevor sie in einem Datenarchiv längerfristig gespeichert werden? Hier steht üblicherweise ein aufwendiger Prozess an, bei dem die Daten z.B. verschiedene Plausibilitätsprüfungen unterlaufen, homogenisiert werden und ggf. weitere Aufbereitungen erfolgen, bevor die Archivierung inklusive der ursprünglichen Rohdaten ansteht.
- **Speicherung / Infrastruktur:** Die persistente Speicherung von Forschungsdaten sollte auf Verfahren basieren, die ein Risiko von Datenverlust minimieren und nachhaltig verfügbar sind (vgl. Kapitel 2.5.4).
- **Erhaltungsmaßnahmen:** Welche Anforderungen und Dokumentationen sind erforderlich, damit digitale Forschungsdaten auch außerhalb der ursprünglichen Erstellungs- und Nutzungsumgebung nutzbar bleiben? Hier müssen technische und organisatorische Maßnahmen getroffen werden, damit Veränderungen in der Technik und in den Standards eine Nutzung der Daten nachhaltig erlauben (vgl. Kapitel 2.5.4).
- **Zugriff und Nutzung:** Wie werden die Daten gefunden, und welche Benutzergruppen mit welchen Rechten haben Zugriff auf die Daten? Dazu sind eine geeignete Nutzeroberfläche mit verschiedenen Suchfunktionen und Filtern, sowie verschiedenen Gruppenrichtlinien erforderlich (vgl. Kapitel 2.5.1).

2.8.5 Metadatenmodelle und Auszeichnungssprachen

Nach Wohlgemuth (2009) bieten die Eigenschaften von Metadatenmodellen zur Beschreibung von Daten mit Metadaten zwei entscheidende Vorteile gegenüber Daten ohne diese Modelle: Erstens werden die Daten für den Benutzer so strukturiert, dass dieser die Daten leichter erfassen und verarbeiten kann. Zweitens wird es digital ermöglicht, gewünschte Daten über das Metadatenmodell aufzurufen. Mit Metadatenmodellen können Systeme einschließlich ihrer Rahmenbedingungen beschrieben werden. Metadatenmodelle sind somit ein geeignetes Konzept für Informationsmanagement und Wissenschaft, mit Hilfe dessen große Datenmengen strukturiert verwaltet werden können.

Für L-TER VREs müssen die Daten mit Metadaten nach Standards (z.B. INSPIRE) oder eigenen Definitionen beschrieben werden. So kann sich z.B. die Nachvollziehbarkeit des Forschungsprozesses mit einem Attribut „Änderungsdatum“ erhöhen. Dazu werden Komponenten für das Metadaten-Management mit geeigneten Metadatenmodel-

len benötigt. Besonders Hilfreich sind Systeme, die den Benutzer bei der Erstellung von Metadaten aktiv unterstützen können.

Die häufigste Definition von Metadaten ist: „Metadaten sind Daten über Daten“ (Sicilia 2014). Als Metadaten werden strukturierte Daten bezeichnet, die Daten über andere Datenressourcen enthalten (Steigerwald 2013). Metadaten beschreiben also die eigentlichen Daten. Sie werden erforderlich, wenn es größere Datenmengen zu verwalten gibt. Ein ausgesprochenes Merkmal von Metadaten ist oft, dass sie maschinell lesbar und auswertbar sind (Goldfarb und Prescod 2000).

Für die Erstellung von Metadatenystemen stellt sich zunächst die Frage nach Ziel und Zweck der Bereitstellung von Metadaten, welche je nach Nutzungs-Szenario und Zielgruppe stark variieren können. Sie bestimmen letztendlich die Form und wie die Metadaten verwendet werden. Zur Klassifizierung und Strukturierung der Metadaten wird folgend eine Gliederung vorgeschlagen (Haynes 2004; Ludwig und Enke 2013):

- **Sichtbarkeit des Datenbestandes:** Metadaten können dazu dienen, einen Überblick des vorliegenden Datenbestandes zu geben, bzw. die Daten zu filtern. Typischerweise lassen sich durch diese Form der Metadaten direkte Rückschlüsse auf den Inhalt der angebotenen Informationen und Ressourcen ziehen.
- **Interpretierbarkeit der Daten:** Oftmals sind Daten nur in ihrem Kontext verständlich. Metadaten können z.B. die Rahmenbedingungen bei der Erhebung (Ort, Zeit, Sensoren etc.) beschreiben und auf die Bedeutung der Daten eingehen (Kategorien, Skalierung etc.).
- **Datenaustausch / Harvesting:** Der Austausch von Daten zwischen verschiedenen Systemen oder der Datenverbund von Netzwerkpartnern kann ebenso mit Hilfe von Metadaten erfolgen. So können z.B. Informationen durch „Zeitstempel“ aggregiert oder synchronisiert werden und andere zum Austausch erforderliche Angaben bereitgestellt werden.
- **Datenpflege und Verwaltung:** Unter diese Rubrik fallen Metadaten, die Auskunft hinsichtlich der Verwaltung und Pflege der Datenbestände geben, wie z.B. Autor, Version, eindeutige Datensatznummern zur Identifizierung, Änderungsdaten, Gültigkeitszeiträume, Zugriffsrechte, etc.
- **Darstellung und Präsentation:** Die effektive Nutzung von Daten erfordert meist zusätzliche Informationen und die Zusammenführung und Verknüpfung mehrerer kontextspezifischer Objekte, wie z.B. Multimedia-Objekte, die zu einem Text passen. Auch die nutzerspezifischen Angaben, wie z.B. die voreingestellte Sprache des Nutzers, gehören dazu. Für die Aufbereitung der Informationen zu einer gesamten Darstellung oder Präsentation der Datenbestände wird diese Form der Metadaten eingesetzt.

Metadaten können verschiedene Erscheinungsformen annehmen, wobei sie zum einen für den Nutzer sichtbar sein können, wie z.B. in einem Bibliothekskatalog oder zum

anderen nur für das System verborgen im Hintergrund. Drei Typen von Metadaten werden definiert, mit denen Objekte in einem digitalen Katalog beschrieben werden können (Oxford Digital Library 2009):

- **Deskriptive Metadaten:** Diese Form der Metadaten beschreibt den intellektuellen Kontext der Daten, wie z.B. Titel, Schlagwörter, Zusammenfassungen und dergleichen.
- **Administrative Metadaten:** Administrative Metadaten oder auch oft technische Metadaten genannt, dienen vornehmlich zur Datenpflege und kennzeichnen meist technische Aspekte der Daten, wie z.B. Dateigröße, Digitalisierungsart, Dateiformat und Metadaten zur Langzeitarchivierung (engl. Preservation Metadata).
- **Strukturelle Metadaten:** Informationen über den logischen Zusammenhang von einzelnen Informationseinheiten werden als strukturelle Metadaten bezeichnet. Dadurch lassen sich z.B. Bilder zu einem Artikel zuordnen oder Querverweise erstellen um so größerer Informationseinheiten aufzubauen.

Unabhängig vom jeweiligen Projekt gibt es Relationen, die generell von Bedeutung sind um die Beziehungen zwischen den Daten darzustellen (Ludwig und Enke 2013):

- **Objekte:** Metadaten finden Verwendung zur Beschreibung und Identifikation von Objekten, die während eines Arbeitsprozesses entstehen oder bearbeitet werden. Wichtig ist eine eindeutige und dauerhafte Identifikation der Objekte beispielsweise mit persistenten Identifikatoren.
- **Akteure:** Sie sind die Durchführer von Aktionen. Akteure können z.B. Personen, Gruppen und/oder Organisationen sein, die an der Entstehung oder Bearbeitung von Daten beteiligt waren. Aber auch technische Systeme können zu den Akteuren zählen, z.B. bei agentenbasierten Systemen.
- **Quellen:** Um den Entstehungskontext von Daten zu dokumentieren, werden z.B. Zeit, Ort, die Umgebungsparameter und weitere Umstände, unter denen sie erhoben wurden, festgehalten. Dazu gehören auch die Akteure.
- **Vorgänge:** Wichtig ist zu wissen, wie die Daten in ihrem Lebenszyklus chronologisch bearbeitet werden. Dazu gehört eine Dokumentation der Bearbeitungsschritte, welche Aktionen von welchen Akteuren angestoßen wurden und welche Programme z.B. zum Einsatz gekommen sind.
- **Ergebnisse:** Die aufbereiteten Ergebnisse sind für eine direkte Nutzung vorgesehen und daher von besonderer Relevanz für die Nachnutzung. Sie müssen durch Metadaten auffindbar sein und für eventuelle Präsentationen verknüpft und aufbereitet werden.

Es gibt mehrere Standards für die Beschreibung von Daten mit Metadaten. Sie hängen in erster Linie von der Art der Daten und ihrer Verwendung ab. Im wissenschaftlich-bibliothekarischen Bereich werden die Metadaten nach dem Regelwerk RAK-WB (Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken) vergeben (Popst 1993). Für Webseiten gibt es die Regeln der Dublin Core Metadata Initiative (Lynch 1998). Daneben gibt es noch fachspezifischen Regelwerke, wie die „Content Standards for Digital Geospatial Metadata“ (CSDGM) (Federal Geographic Data Committee 2017) für die Geowissenschaften oder fachliche Regelwerke für Metadaten der IEEE (Institute of Electrical and Electronic Engineers) (IEEE 2017). Hierzu sind auch die Normen der „International Organization for Standardization“ (ISO) (iso.org 2017) und INSPIRE für georeferenzierte Daten (vgl. Kapitel 2.6) zu erwähnen.

Benutzer, die ein Dokumentenmanagementsystem verwenden, benutzen oftmals selbst definierte Metadaten, die sich aus der Art der Dokumente und den Prozessen ergeben. Bei ökologischen Daten könnten das zum Beispiel folgende Metadaten sein: Probennummer, Probendatum, Bearbeiter, Ort, Zeit.

Metadaten können durch Auszeichnungssprachen beschrieben und maschinell verarbeitet werden. Auszeichnungssprachen (engl. markup languages, Abkürzung ML) sind maschinenlesbare Sprachen, die zur Gliederung und Formatierung von Text und anderen Daten verwendet werden (Witt und Metzinger 2010). Sie beschreiben die Eigenschaften, Zugehörigkeiten und Darstellungsformen einer Datenmenge. Sie sind somit auch eine Sprache für Metadaten (Haynes 2004). Der bekannteste Vertreter ist die Hypertext Markup Language (HTML), die als Basissprache im World Wide Web verwendet wird.

Nach Steigerwald (2013) ist die eXtensible Markup Language (XML) eine plattformübergreifende erweiterbare Auszeichnungssprache in Form von Textdateien, die hierarchisch strukturierte Daten beschreiben kann und die sich besonders für den Austausch zwischen Computersystemen eignet. Sie wurde vom World Wide Web Consortium (W3C 2017) erstmals 1998 spezifiziert und ist seitdem weltweiter Standard, um Daten maschinenlesbar zu beschreiben. Sie dient als Basis für andere aufgaben- und fachspezifische Auszeichnungssprachen, die auf dem gleichen Schema (XML-Dokumentenstruktur) aufbauen, sich aber in den zu beschreibenden Objekten und damit verbundenen Elementen und Attributen unterscheiden. Damit eignet sich XML besonders zur maschinellen und systemübergreifenden Verarbeitung von Metadaten-Schemata bei gleichzeitig noch guter Lesbarkeit der XML-Dokumente für den Menschen.

In Tabelle 2.6 werden einige allgemeine Metadaten-Formate und Auszeichnungssprachen (Markup Languages) vorgestellt, Tabelle 2.7 enthält Beispiele aus dem Umweltsektor.

Tabelle 2.6: Beispiele für Metadaten Standards und Auszeichnungssprachen vgl. Jensen et al. (2011)

Name	Beschreibung
BibTeX	Bibliografien in LaTeX-Dokumenten
Dublin Core	Standard zur Beschreibung aller Arten von Ressourcen
Encoded Archival Description (EAD)	Beschreibung von Archivgütern
Exchangeable Image File Format (Exif)	Metadaten in digitalen Bildern
FGDC (Federal Geographic Data Committee), ISO 19115	Geodaten
IPTC-IIM-Standard	Metadaten in Bilddateien
LOM (Learning Object Metadata)	Beschreibungen von Lernobjekten
MARC, MAB, MODS	Austauschformate für Bibliotheken
Metadata Encoding and Transmission Standard (METS)	XML Schema zur Beschreibung von digitalen Sammlungen von Objekten mit Metadaten
Meta-Object Facility (MOF)	spezielle Metadaten Architektur
MPEG-7, MPEG-21	Beschreibung von Multimedia Objekten
ONIX	bibliografische Daten im Buchhandel
Resource Description Framework (RDF) –	formale Sprache zur Bereitstellung von Metadaten im WWW
SDMX - Statistical Data and Metadata eXchange	Austausch statistischer Daten
XBMF Exchange Binary Broadcast and Metadata Format -	Containerformat für Metadaten im XML-Format
Extensible Metadata Plattform (XMP)	Auszeichnungstechnologie für Metadaten

Tabelle 2.7: Beispiele umweltrelevanter Auszeichnungssprachen

Name	Beschreibung
AgroXML	Datenaustauschsprache im Agrarsektor
GML	Geography Markup Language
GPX	GPS Exchange Format
KML	Keyhole Markup Language
OWL	Web Ontology Language
RDF	Resource Description Framework
SBML	Systems Biology Markup Language
SensorML	Datenaustauschsprache für Sensoren
WaterML	Datenaustauschsprache in der Hydrologie

2.8.6 Modellkopplung

In L-TER Projekten kommen häufig verschiedene digitale ökologische Modelle aus interdisziplinären Gebieten mit unterschiedlichen Datenstrukturen zusammen (vgl. Kapitel 2.2 und 2.3). Um im emergenten Sinne eine Kopplung der verschiedenen Modelle zu ermöglichen, muss eine gemeinsame Informationsverwaltung und ein gegenseitiger Informationsaustausch der Modelle realisiert werden. Die Kopplung von Modellen ist eine komplexe Aufgabe, die von verschiedenen Faktoren abhängt. Eine der größten Herausforderungen sind die oft unzureichenden Kopplungsschnittstellen und Kopplungsfähigkeiten von Modellen in der eingesetzten Software. Verschiedene Datenformate, verwendete Raum- und Zeitskalen und die Diskretisierung von Raum und Zeit müssen neben weiteren Metadaten dabei berücksichtigt werden. Eine Lösung für dieses Problem können spezielle Software-Frameworks, wie „OpenMI“ (OpenMI.org 2017), zur Modellkopplung auf der Basis von *Tensorobjekten* bieten (Notay 2015; Molkenhuth 2000; Gregersen et al. 2007).

Tensorobjekte sind in diesem Zusammenhang autonome Entitäten, die sich an die Anforderungen der gekoppelten Modelle flexibel anpassen können. Ein *Tensor Exchange Server* kann die zentrale Aufgabe der Kopplung der verschiedenen Modelle als Middleware übernehmen (Notay 2015). Zu seinen Funktionen zählen die Definition des Kopplungsmechanismus, die Kommunikation zwischen gekoppelten Modellen und die Anpassung dieser Informationen an die Anforderungen der Modelle (z.B. Zeit- und Ortsdiskretisierung), die durch die interne Anpassungsfähigkeit von Tensorobjekten gegeben ist. Diese Funktionsweise eines Tensor Exchange Server erfordert eine individuelle interne Anpassung der einzelnen gekoppelten Modelle und geht somit einen Schritt weiter als konventionelle Kopplungsstandards, wie „OpenMI“. Mit Tensor Exchange Servern lassen sich auch Szenarien basierte Prognosen für Zeitreihen aus verschiedenen Modellen realisieren (Li und Molkenhuth 2014).

3. Voraussetzungen einer L-TER VRE

Projekte in der ökologischen Langzeitforschung sind durch den Langzeitaspekt und Raumbezug in der Regel umfangreiche Projekte, mit großen Datenmengen (Medium- und Big Data) und vielen unterschiedlichen Untersuchungsobjekten auf verschiedenen Skalenebenen (vgl. Kapitel 2.3 und 2.8.2). Besonders der Zeitraum unterscheidet sie von anderen ökologischen Projekten. Im Sinne der ökologischen Langzeitforschung ist ggf. der Wissenstransfer über Generationen von Wissenschaftlern, Administratoren oder Betreibern hinweg sicherzustellen. Dadurch wird ein vergleichsweise langer Betriebszeitraum für VREs im Vergleich zu normalen Projekten in der Ökologie erforderlich, die sich oft nur über wenige Jahre in einem Förderungszeitraum erstrecken. Dies bringt große Herausforderungen hinsichtlich der Persistenz (Daten, Infrastruktur), Flexibilität (projektspezifische Eigenschaften), der eingesetzten ICT sowie des Faktors „Mensch“ und seinem sozialen und gesellschaftlichen Umfeldes mit sich. Die persistente Langzeitarchivierung (vgl. Kapitel 2.5.3) ist eine zentrale Eigenschaft von L-TER VREs. VRE-Konzepte für die ökologische Langzeitforschung sollten diese entscheidenden Aspekte daher genauer untersuchen und in einem speziellen Konzept verankern, welches die genannten Punkte berücksichtigt. Aus der Literaturrecherche (vgl. Kapitel 2) gingen keine speziellen Konzepte für VREs in der ökologischen Langzeitforschung hervor. Daher hat die vorliegende Arbeit das Ziel, die speziellen Aspekte und Anforderungen an virtuelle Forschungsumgebungen von Projekten in der ökologischen Langzeitforschung zu untersuchen und führt damit den Begriff der „L-TER VRE“ ein. Grundlage hierfür ist der Stand des Wissens zu allgemeinen VREs, Datenmanagement und ICT (vgl. Kapitel 2), die interdisziplinär in Verbindung gebracht werden mit den Zielen und Methoden (vgl. Kapitel 2.3) sowie der Daten (vgl. Kapitel 2.8) der ökologischen Langzeitforschung.

3.1 Einführung des Begriffes „L-TER VRE“ und weitere Definitionen

Aus den verschiedenen Definitionen von VREs (vgl. Kapitel. 2) wird folgende eigene Definition für die weitere Verwendung des Begriffes in Anlehnung an Fraser (2005) in dieser Arbeit gewählt:

Virtuelle Forschungsumgebungen (VREs) sind auf der Basis von ICT dynamische, webbasierte Arbeitsumgebungen für digitale Forschungsgemeinschaften zur Unterstützung der Gesamtheit oder einzelner Schritte des Forschungsprozesses. Sie fördert insbesondere das wissenschaftliche Arbeiten im Team, dessen Kommunikation und Organisation. Sie haben das übergeordnete Ziel, zu schnelleren und besseren Forschungsergebnissen zu gelangen, als es ohne ihren Einsatz möglich gewesen wäre. Eine VRE besteht aus VRE-Ressourcen,- Services und -Tools.

- **VRE-Ressourcen:** Forschungsprimärdaten, Ergebnisse der Forschung, Hardware-Infrastruktur, Software, Netze, Datenarchive, Personen.
- **VRE-Services:** Einzelne von der VRE prinzipiell unabhängige Dienste zur Unterstützung der Forschungsprozesse, disziplinübergreifend oder spezifisch in die Gesamtarchitektur ggf. in Form einer SOA eingebettet, wie z.B. komplette Infrastrukturen für CMS, Cloud-Services oder Remote-Desktop-Services. Sie sind funktional erweiterbar durch Service-spezifische VRE-Tools.
- **VRE-Tools:** Werkzeuge zur Bearbeitung einzelner Aufgaben innerhalb der Services oder der gesamten VRE (auch in der Form von Middleware und Schnittstellen). Als Beispiele: Tool für räumliche Dateninterpolation von georeferenzierten Daten, Tool als Middleware zur Kopplung einzelner Services für den Datenaustausch untereinander.

Diese allgemeine Definition von VREs wird im Folgenden für L-TER VREs um den Faktor der „Nachhaltigkeit“ erweitert, der den Langzeitaspekt von L-TER Projekten berücksichtigen soll. Unter „Nachhaltigkeit“ wird zunächst in diesem Zusammenhang die ursprüngliche Bedeutung einer „längeren Zeit anhaltenden Wirkung“ (Duden) von VRE verstanden, die im Weiteren differenziert wird:

L-TER VREs sind definiert als VREs, die im besonderen Maße nachhaltig dazu geeignet sind, die primären Aufgaben der langzeitökologischen Forschung unter der Herausforderung des stark variablen und mit Unsicherheiten behafteten Langzeitaspektes zu unterstützen. Zu einer L-TER VRE gehören die dynamischen Basis-Nachhaltigkeitsfaktoren der Persistenz, Flexibilität, ICT und der Faktor „Mensch“. Dies wirkt sich beispielsweise aus auf: Lebensdauer der VRE, gespeicherte Daten und Strukturen, Analysemethoden, Infrastrukturen, Langzeitarchivierung und Benutzer.

Open-Source-Software kann als „Software-Ökosystem“ im Sinne eines wirtschaftlichen Ökosystems betrachtet werden (vgl. Kapitel 2.4). Ökosysteme sind dynamische Systeme (vgl. Kapitel 2.3). Das „Rapid L-TER VRE“ Baukastenkonzept dient zur Realisierung eines dynamischen, technischen *Systems* auf Open-Source-Basis in Form einer L-TER VRE. Daher werden die fünf Kriterien zur Nachhaltigkeit von Ökosystemen (Renn et al. 2007) für eine nachhaltige L-TER VRE zunächst näher betrachtet (Tabelle 3.1).

Aus der Gegenüberstellung (Tabelle 3.1) geht hervor, dass sich die Nachhaltigkeitskriterien von Ökosystemen auf L-TER VREs anwenden lassen. Bei Ökosystemen sowie bei L-TER VREs gibt es ökosystem- und *projektspezifische (L-TER VRE-spezifische) Rahmenbedingungen*, die für den Start einer L-TER VRE zunächst durch ein Konzept vorgegeben sind und sich mit der Zeit dynamisch ändern können. Die projektspezifischen Rahmenbedingungen werden im Folgenden auch als *Flexibilität* einer L-TER VRE bezeichnet.

Tabelle 3.1: Nachhaltigkeitskriterien von Ökosystemen (Renn et al. 2007) im Vergleich zu L-TER VREs

Nachhaltigkeitskriterium	Ökosystem (Renn et al. 2007)	L-TER VRE
Persistenz	Wenn sich ein Ökosystem über die Zeit hinweg erhält.	Wenn sich eine L-TER VRE und Ihre Bestandteile über die Zeit hinweg erhalten.
Konstanz	Wenn ein Ökosystem trotz Umwelteinflüssen den ursprünglichen Zustand erhält.	Wenn eine L-TER VRE trotz Einflüssen den ursprünglichen Zustand erhält.
Resistenz	Wenn ein Ökosystem trotz Umwelteinflüssen nicht aus dem Gleichgewicht gerät.	Wenn eine L-TER VRE trotz Einflüssen nicht aus dem Gleichgewicht gerät.
Elastizität	Ausmaß der Ökosystemänderung nach starker Störung, wobei das Ökosystem wieder in den Ausgangszustand übergeht.	Ausmaß der L-TER VRE - Änderung nach starker Störung, wobei die L-TER VRE wieder in den Ausgangszustand übergeht.
Resilienz	Ausmaß der Störung, nach der ein Ökosystem wieder in den Ausgangszustand übergeht.	Ausmaß der Störung, nach der eine L-TER VRE wieder in den Ausgangszustand übergeht.

In dieser Arbeit wird unter den fünf Nachhaltigkeitskriterien der Schwerpunkt auf die Persistenz gelegt: Die L-TER VRE und ihre Bestandteile sollen für die Projektzeit generell zur Verfügung stehen und ggf. darüber hinaus die Forschungsdaten über längere Zeit persistent gespeichert werden. Wie konstant, resistent, elastisch und resilient ihre Eigenschaften sind, ist zweitrangig und abhängig von den Rahmenbedingungen (Flexibilität), der eingesetzten *ICT*, dem *Menschen* mit seinem *sozialen Umfeld*, sowie generell von der *Zeit*. Für L-TER VREs ist der Faktor „Menschen“ bezüglich einzelner Individuen und ihrer Gruppierungen gegenüber der L-TER VRE erheblich.

Die anderen Nachhaltigkeitskriterien (Konstanz, Resistenz, Elastizität und Resilienz) werden daher hier zur Vereinfachung der Darstellung nicht berücksichtigt, spielen jedoch eine bedeutende Rolle für L-TER VREs. Sie werden daher im Kontext noch näher erläutert. Diese Annahmen führen zu der Vereinfachung der Nachhaltigkeitsdefinition für L-TER VREs: Die Nachhaltigkeit einer L-TER VRE setzt sich zusammen aus den *dynamischen Basis-Nachhaltigkeitsfaktoren*: Persistenz, Flexibilität, *ICT* und Mensch (Abbildung 3.1).

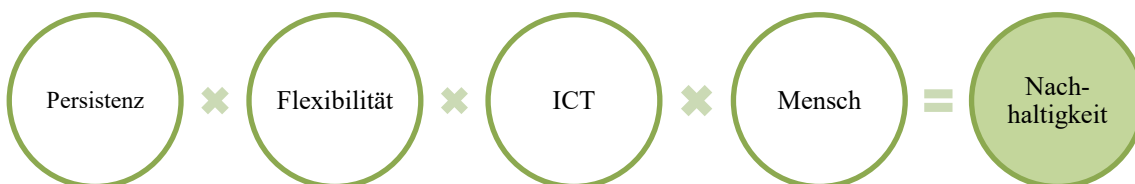


Abbildung 3.1: Dynamische Basis-Nachhaltigkeitsfaktoren einer L-TER VRE

Die Komponenten der dieser Nachhaltigkeitsfaktoren einer L-TER VRE gehen aus Abbildung 3.2 hervor.

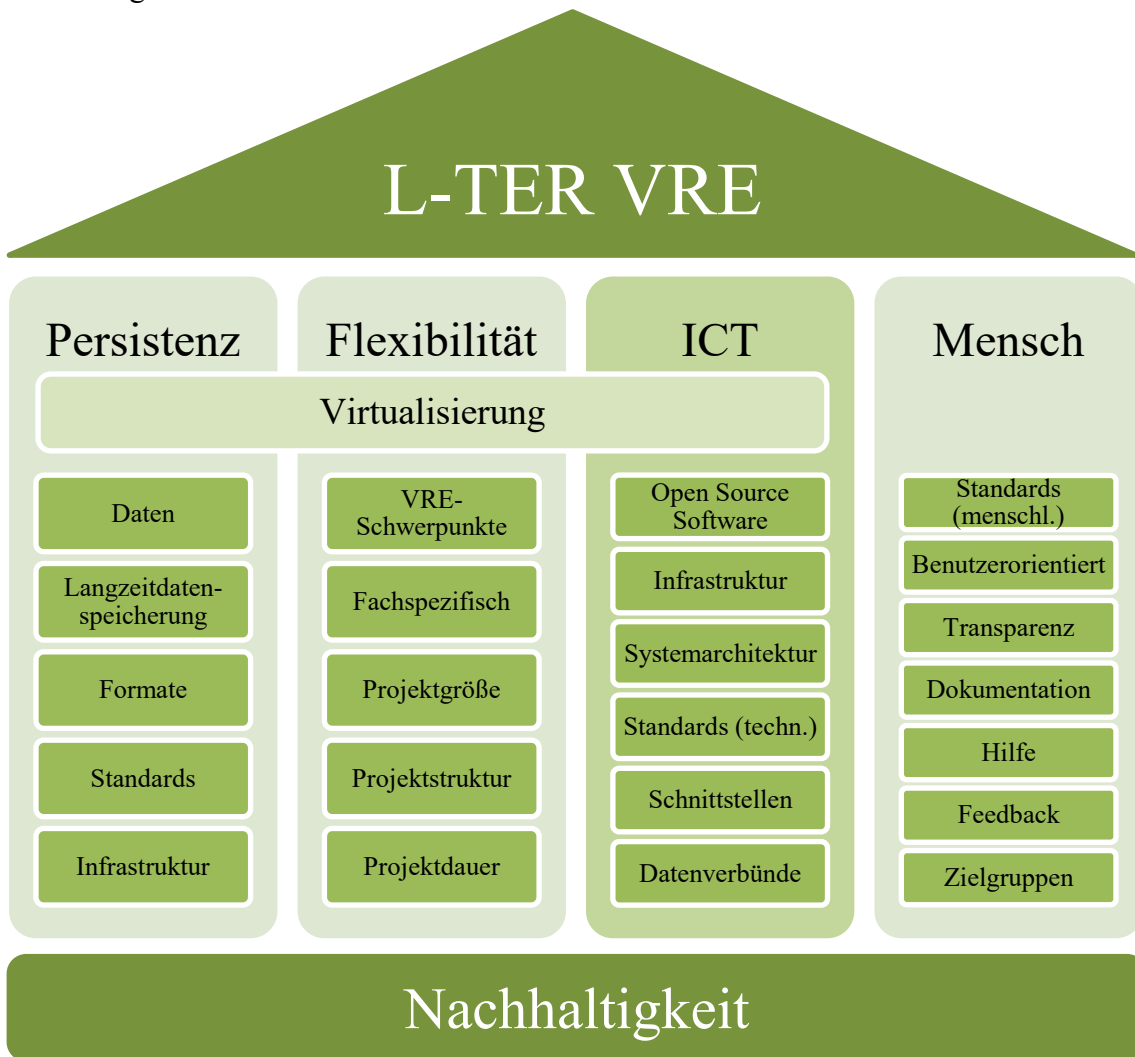


Abbildung 3.2: Einzelne Parameter der Basis-Nachhaltigkeitsfaktoren einer L-TER VRE

Daraus werden die einzelnen Definitionen für die Basis-Nachhaltigkeitsfaktoren einer L-TER VRE abgeleitet:

Persistenz einer L-TER VRE: Die Beständigkeit (Persistenz) der L-TER VRE insgesamt und ihrer Bestandteile (Ressource, Services, Tools) über den geplanten Projektzeitraum oder darüber hinaus.

Flexibilität einer L-TER VRE: Anpassungsfähigkeit der L-TER VRE an projektspezifische Rahmenbedingungen, die für den Start einer L-TER VRE zunächst durch ein Konzept vorgegeben sind und sich mit der Zeit dynamisch ändern können.

Nachhaltige ICT: ICT (mit deren Eigenschaften und Standards), die im besonderen Maße dazu geeignet sind, nachhaltige L-TER VREs in Persistenz, Flexibilität und in Bezug auf den Menschen mit seinem sozialen Umfeld zu unterstützen.

Faktor „Mensch“: *Der Mensch als „maßgeblicher“ Faktor sollte im Mittelpunkt einer L-TER VRE stehen. Bestandteile einer L-TER VRE (Ressourcen, Services, Tools) müssen letztendlich auf den Menschen und sein soziales Umfeld ausgerichtet sein, um persistente, flexible und auf einer nachhaltigen ICT basierende L-TER VREs erfolgreich nachhaltig umzusetzen.*

In Bezug auf den Faktor „Mensch“ können zusätzlich die Erfolgskriterien für allgemeine VREs (vgl. Kapitel 2.5.7) übernommen und um das Kriterium der „Nachhaltigkeit“ gemäß den vorangegangenen Definitionen erweitert werden. Sie werden jedoch nicht zur weiteren Bewertung in dieser Arbeit herangezogen

L-TER VREs können in L-TER Projekten als gemeinsame Basis für Datenverbünde und digitale Forschungsgemeinschaften dienen (Abbildung 3.3).

Digitale Forschungsgemeinschaft: *Eine strukturell dynamische Interessengruppe von Wissenschaftlern/innen, die digital durch ICT unterstützt (mit geeigneten Kommunikations- und Community-Werkzeugen) an gemeinsamen Zielen, über einzelne Institutionen und Forschungsprojekte hinaus, arbeitet. Hier stehen, neben den gemeinsamen Zielen, die Community und sozioökonomische Aspekte im Mittelunkt.*

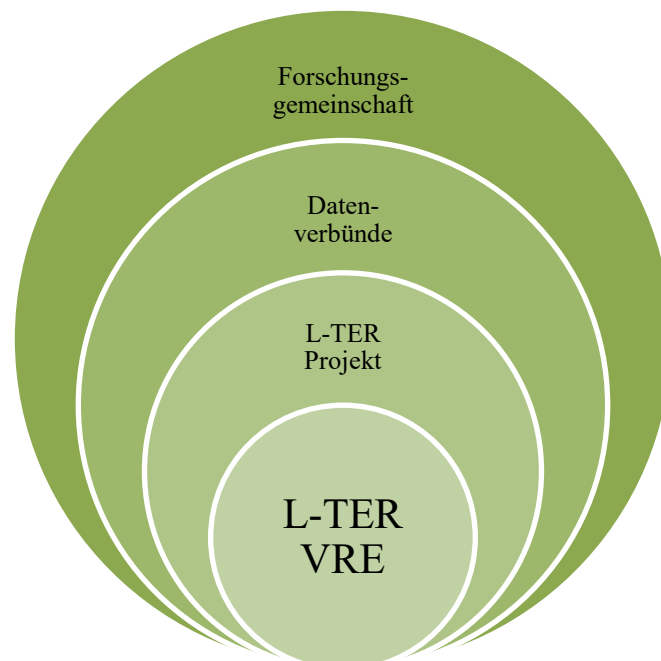


Abbildung 3.3: L-TER VREs als Basis für eine „digitale Forschungsgemeinschaft“

Eine Abgrenzung von L-TER VREs zu allgemeinen VREs zeigt Tabelle 3.2.

Tabelle 3.2: Abgrenzende Kriterien von L-TER VREs gegenüber allgemeinen VREs

Kriterium	L-TER VRE	VRE
Lebensdauer	nachhaltig ausgelegt auf einen möglichst langen Nutzungszeitraum, Generationen von Wissenschaftlern	wenige Jahre, nicht generationenübergreifend, nicht nachhaltig konzipiert
Daten	Ökologische Langzeitdaten: viele Zeitreihen, räumliche Daten und Sensordaten, Ort /Zeit Komponente -> Datenmenge, dynamische Strukturen erforderlich, Formatbeschreibung mit abspeichern (selbstbeschreibend)	dynamische Datenstrukturen weniger benötigt, dafür kurze Projektzeiträume, die vorgegebene Struktur bleibt meist unverändert, Formatbeschreibungen für derzeitige Standards (z.B. Aufbau einer WORD-Datei nur bedingt erforderlich)
Datenanalyse	Langzeitanalysen und –statistik, Trendanalysen, automatisierbar und kombinierbar durch Wokflows	kurzer Analysezeitraum, die Programmierung von Workflows rentiert sich weniger für kurze Projekte
Infrastruktur	modularer Aufbau, erweiterbar, skalierbar, dynamisch, nachhaltige Standards und Software, Virtualisierung	abhängig im Einzelfall
Speicherung	persistente, zuverlässige Langzeitspeicherung /-archivierung und Datensicherung	abhängig im Einzelfall
Benutzer	hohe Dynamik: viele und häufig wechselnde Benutzer, Administratoren, Gruppen, Institutionen, mehrere Generationen	oft kleinere Gruppen und wenige Institutionen mit wenig wechselnden Benutzern (relativ statisch)
Reports	(Auto)Report-Generierung	Aufwand der Programmierung zum Nutzen ist relativ groß, daher oft nur in Form von Publikationen zum Download
Datenverbünde	Umfangreiche Datenverbünde als wesentliches Merkmal, INSPIRE-konform, Austausch mit anderen Netzwerken und Institutionen	zentrale Daten stehen im Vordergrund, Datenverbünde im Hintergrund

3.2 Funktionale Voraussetzungen von L-TER VREs

Die folgenden Funktionalitäten für VREs leiten sich aus der Literaturrecherche zum Stand der Technik (vgl. Kapitel 2.5) ab und wurden durch die „Deutsche Initiative für Netzwerkinformation e.V. - AG Virtuelle Forschungsumgebungen“ (Deutsche Initiative für Netzwerkinformation e.V. 2011; Munson et al. 2016) klassifiziert. Sie werden auch für L-TER VREs übernommen.

Funktionalitäten von VREs allgemein:

1. gemeinsame Nutzung lokaler und externer Ressourcen wie z.B. Daten, Dienste/Software, Hardware,
2. Bereitstellung einer gemeinsamen Kommunikationsplattform und deren Nutzung,
3. Unterstützung bei der Erfassung und Erhebung von Daten vor Ort: z.B. Feldstudien, Beobachtungen, Fernsteuerung von Experimenten, Sensor-Monitoring,
4. Dokumentation und Organisation von Daten unter Berücksichtigung entsprechender Metadaten und Standards,
5. Weiterverarbeitung der Daten, sowie deren fachsystematische Analyse, Aufbereitung und Zusammenführung,
6. Publikation von Daten und Ergebnissen ggf. mit zitierfähigen Datenquellen durch persistente Identifikatoren.

Aufbauend auf den Funktionalitäten von allgemeinen VREs definiert die vorliegende Arbeit spezifische Funktionalitäten von L-TER VREs. Diese basieren auf den Anforderungen und Zielen der ökologischen Langzeitforschung – insbesondere des Langzeitaspektes und der Nachhaltigkeit (vgl. Kapitel 2.3), sowie den Definitionen zu L-TER VREs (vgl. Kapitel 3.1).

Spezifische Funktionalitäten von L-TER VREs:

1. nachhaltige Nutzbarkeit der L-TER VRE und ihrer Inhalte über lange Zeiträume und gegebenenfalls Generationen von Wissenschaftlern hinweg,
2. nachhaltiges Datenmanagement mit geeigneten Datenstrukturen und Metadaten,
3. nachhaltige Workflows für wiederkehrende Aufgabenstellungen,
4. Interoperabilität im (Meta-)Datenaustausch mit anderen Netzwerken und L-TER VREs,
5. nachhaltige und umfangreiche Dokumentationsmöglichkeiten zur L-TER VRE, verwendeten Datenformaten und Standards,
6. besondere Funktionalitäten für die Aufbereitung und Auswertung von Langzeitdaten (z.B. Zeitreihen) mit räumlichem und ökologischem Bezug,
7. Faktor „Mensch“, sein soziales und gesellschaftliches Umfeld wird berücksichtigt,
8. automatische Generierung von Reports, Metadaten, Kennzahlen.

3.3 Projektspezifische Voraussetzungen

Bei der Fragestellung, unter welchen Umständen eine VRE (allgemein) aufgebaut werden sollte, ist zunächst zu beachten, dass die Entwicklung einer VRE bisher meist viel Zeit benötigte und eine erhebliche Menge an Ressourcen beanspruchte. Für die Entscheidungsfindung sollte daher abgewogen werden, ob der *zu erwartende Mehrwert* der VRE für die geplanten Benutzergruppen den Einsatz der vorhandenen oder noch zusätzlich benötigten Ressourcen rechtfertigt. Nach der analysierten Literatur (vgl. Kapitel 2.5) liegt dieser Mehrwert besonders in einer digitalen Forschungsgemeinschaft mit Forschungsergebnissen, die ohne den Einsatz einer VRE nicht denkbar gewesen wäre. Insbesondere zählen hierzu die orts- und zeitunabhängige Nutzung und Bearbeitung des Datenbestandes, ein einfacher Austausch von Materialien und Ergebnissen, sowie Moderationsprozesse und Workflows. Hinzu kommen moderne Kommunikationsmöglichkeiten z.B. in Form von Foren und dynamischen Gruppen, wie sie in sozialen Netzwerken mit weiter Verbreitung angeboten werden. Auch die Einbeziehung der Öffentlichkeit in Projekte kann durch VREs erleichtert werden. Allgemein haben VREs die Eigenschaft, für unterschiedlichste Disziplinen eingesetzt werden zu können. Die spezifische Adaption an die Bedürfnisse der jeweiligen Benutzergruppe ist dabei der Schlüssel zur Realisierung des gesamten Potentials.

In Bezug zur ökologischen Langzeitforschung sollten insbesondere L-TER VREs aufgebaut werden, da es sich um Langzeitprojekte oder „größere“ Projekte handelt und wenn die hier angestellten Überlegungen zutreffen. Eine persistente Archivierung des Forschungsprozesses zur Nachvollziehbarkeit und Nachnutzung ggf. über Generationen von Wissenschaftlern hinweg soll dadurch gewährleistet und der Investitionswert an Ressourcen persistent gesichert werden.

Prinzipiell eignen sich auf lange Zeiträume ausgerichtet Projekte vermutlich besser für den Aufbau von L-TER VREs, da sich der anfängliche Aufwand eher amortisiert und das Erlernen der Bedienung einer L-TER VRE im Verhältnis zur Nutzungsphase eines Benutzers in einem effizienteren Verhältnis steht. Dafür muss jedoch der Mehraufwand bei der Planung für spezielle L-TER VREs abgewogen werden.

Erfüllen die genannten Vorteile und Möglichkeiten die Bedürfnisse und Anforderungen (Erfolgskriterien) des L-TER Projektes bzw. der Forschungsgemeinschaft (vgl. Kapitel 2.5.7) und die Ziele der L-TER (vgl. Kapitel 2.3), so sollte als Lösung die Entwicklung einer L-TER VRE in Betracht gezogen werden. Sind bereits schon Forschungswerkzeuge und Kommunikationsdienste im Einsatz, die angemessen sind oder einen großen Teil des Forschungsprozesses abbilden, kann es besser sein, vorhandene Ressourcen in die Verbesserung der alten Infrastruktur zu stecken. Die vollständige Integration aller Ressourcen, VRE-Services und -Tools in eine L-TER VRE ist technisch aufwendig und könnte eine Verkettung von Umstellungen nach sich ziehen, sodass die alte Infrastruktur zu Teilen oder im Ganzen angepasst oder ersetzt werden muss. Bevor Ressourcen für den Aufbau einer L-TER VRE zum Einsatz kommen, sollte gründlich geprüft werden, ob eine L-TER VRE wirklich benötigt wird. Andernfalls

sollten die Ressourcen für andere, relevantere Einsatzzwecke verwendet werden (Munson et al. 2016).

4. Das Baukastenkonzept „Rapid L-TER VRE“

Aus den erarbeiteten Grundlagen zum Stand der Technik (vgl. Kapitel 2) und daraus gewonnenen Erkenntnissen (vgl. Kapitel 3) wird im Folgenden ein innovatives Baukastenkonzept zur Planung, dem Aufbau und Betrieb von nachhaltigen L-TER VREs abgeleitet und vorgestellt.

Das *Baukastenkonzept* bezieht sich hier insbesondere auf die VRE-Services und -Tools, die je nach Anforderungen der Projekte und Forschungsgemeinschaften die „Bausteine“ eines „VRE-Baukastens“ der L-TER VRE bilden.

Das Konzept ist im Ansatz hinsichtlich einer *rapiden Entwicklung* von Prototypen mit einer schnellen (engl. rapid) Übergangsphase in den Produktivbetrieb (regulärer Betrieb des Systems) ausgelegt und trägt daher den Namen „Rapid L-TER VRE“. Die schnelle Entwicklung von L-TER VREs ist zwar keine Voraussetzung, aber eine erhebliche Eigenschaft des Konzeptes, wie in der Einleitung (Kapitel 1) beschrieben.

Damit ergänzt das Konzept die bereits bestehenden Konzepte für VREs durch einen *nachhaltigen Ansatz*, der durch den Langzeitaspekt bezüglich L-TER erforderlich ist und die schnelle Entwicklung von L-TER VREs.

Das Konzept besteht aus zehn Punkten und verbindet somit verschiedene grundlegende Aspekte, die für die erfolgreiche Realisierung einer L-TER VRE erforderlich sind. Konkrete Softwareempfehlungen im Anhang sollen Zeit bei der Suche nach geeigneter und untereinander kompatibler Open-Source-Software einsparen. Zunächst werden die einzelnen Bestandteile des Konzeptes einleitend kurz erläutert und anschließend in den Unterkapiteln näher beschrieben. Abbildung 4.1 zeigt die Bestandteile des „Rapid L-TER VRE“ Baukastenkonzeptes.

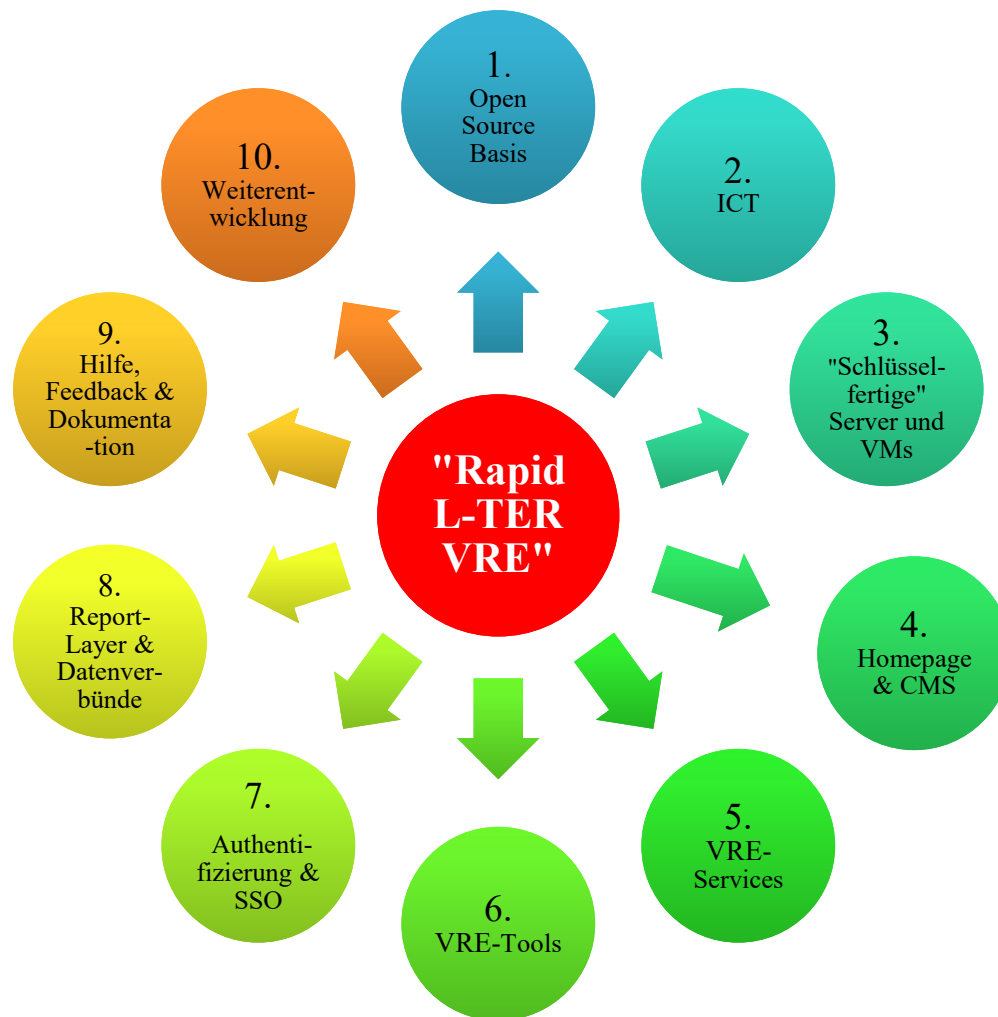


Abbildung 4.1: Bestandteile des „Rapid L-TER VRE“ Baukastenkonzept

Das Konzept für die schnelle Realisierung nachhaltiger L-TER VREs basiert auf den folgenden Punkten:

1. Einsatz von Open-Source-Software (FLOSS) für die Basis-Bestandteile einer L-TER VRE (vgl. Kapitel 2.4).
2. ICT in Form einer webbasierten Informationsstruktur, die auf Webtechnologien und offenen Standards nachhaltig aufbaut (vgl. Kapitel 2.7).
3. „Schlüsselfertige“ Server und Virtualisierungstechnologien zur schnellen Realisierung, Reduzierung des administrativen Aufwandes, Langzeitarchivierung und Weiterentwicklung von L-TER VREs (vgl. Kapitel 2.7.3).
4. Homepage und modular aufgebautes CMS (Open-Source) als zentraler Einstiegspunkt der L-TER VRE und Basis-Software mit erweiterbaren VRE-Tools sowie zur Einbindung weiterer VRE-Services (vgl. Kapitel 2.7.5).
5. Baukasten mit L-TER VRE-Services (vgl. Kapitel 3.1) als einzelne, von der VRE prinzipiell unabhängige Dienste zur Unterstützung der Forschungsprozesse, disziplinübergreifend oder spezifisch in die Gesamtarchitektur (ggf. im Rah-

- men einer SOA) eingebettet (vgl. Kapitel 2.7.8.), wie z.B. ein Datenbank-Service oder WebGIS-Service.
6. Erweiterbarkeit durch VRE-Tools als Werkzeuge für Forschungsprozesse innerhalb der VRE-Services, bzw. VRE-übergreifend (VRE selbst als Service betrachtet) (vgl. Kapitel 3.1).
 7. Umsetzung einer einheitliche „Einmalanmeldung“ von Benutzern in der L-TER VRE durch zentrale Authentifizierung und SSO (vgl. Kapitel 2.7.9).
 8. Anwendung eines „Reportlayers“ mit interoperablem Datenaustausch (engl. Harvesting), basierend auf offenen Standards (vgl. Kapitel 2.7.10), sowie maschinell unterstützter Datenaufbereitung von Daten und Metadaten in Form von verschiedenen zielgruppenorientierten Berichten (engl. Reports).
 9. Benutzerhilfe, -Feedback und technische Dokumentation im Fokus des Langzeitaspektes bei der Realisierung und Entwicklung von L-TER VREs.
 10. Einteilung nach „Basis-“ und „erweiterten Funktionalitäten“ (VRE-Tools und -Services) hinsichtlich einer schnellen Realisierung von L-TER VREs. Weiterentwicklung durch modularen Aufbau des Systems für den zukünftigen Stand der Technik mit innovativer ICT und neuen fachspezifischen Methoden (Langzeitaspekt).

Die folgenden Unterkapitel erläutern näher die einzelnen Bestandteile des „Rapid L-TER VRE“ Baukastenkonzeptes. L-TER VREs können nach diesem Konzept schrittweise umgesetzt werden, wie das vorgestellte Life-Cycle-Modell aus Kapitel 2.5.3 beschreibt.

4.1 Einsatz von Open-Source-Software

Umweltgüter, wie z.B. Luft und Wasser, sind in der Regel *öffentliche Güter* (Rogall 2002), die daher meist mit öffentlichen Mitteln z.B. durch L-TER Projekte geschützt werden sollen. Zum Schutz der Umwelt (und damit verbunden zur Erhaltung der menschlichen Gesundheit) stehen vermutlich weniger Gelder zur Verfügung, als für gewinnbringende Projekte aus der Wirtschaft. Daher muss mit den vorhandenen Ressourcen besonders sparsam unter Ausnutzung möglichst vieler Synergien (vgl. „Open-Data“) umgegangen werden. Die Philosophie von „Open Source“ ist eine Philosophie der „freien“ und „transparenten“ Software. Daher sollte sie insbesondere auch bei L-TER VREs zur Schonung der Ressourcen beitragen. Nach dem Open-Source-Prinzip (vgl. Kapitel 2.4) kann eine weltweite Entwicklergemeinschaft an einer speziellen Komponente (bzw. Problem) viel effizienter und mit mehr „Manpower“ arbeiten, da meist auch der Nutzerkreis von kleinen, unspezifischen Modulen größer ist, als der eines L-TER Projektes. Weiterhin kann die jeweilige fachliche Ausrichtung der Programmierer bei kleinen Teil-Problem meist eine bessere Wirkung entfalten.

Derzeit existiert kein „VRE-Software-Framework“, das „out-of-the-box“ sofort einsetzbar oder für alle Anforderungen geeignet wäre („one size fit them all“) (vgl. Kapitel 2.4, 2.5.1). Wie bereits in der Einleitung (Kapitel 1) erwähnt, sind komplette Eigenentwicklungen von großen Softwareprojekten in Form von nachhaltigen VREs aus Gründen der Komplexität und des dazu nötigen Ressourcenbedarfs nicht unbedingt zu empfehlen. Bei proprietärer Software ist die Abhängigkeit vom Anbieter sehr groß und individuelle Anforderungen lassen sich kaum umsetzen. Zudem sind sie oft teuer in der Anschaffung oder Miete. Daher sollten L-TER VREs – insbesondere bei den Basis-Services und -Tools – mit Open-Source-Software aufgebaut werden, bei der das Know-how aus der gesamten Entwicklergemeinschaft einfließt, der Quelltext offen liegt und die Software kostenlos verfügbar ist (vgl. Kapitel 2.4).

Der Einsatz von Open-Source-Software ist daher der erste Bestandteil des „Rapid L-TER VRE“ Baukastenkonzeptes, der zur Nachhaltigkeit und „schnellen“ Entwicklung beiträgt.

Zuerst sollte ein Überblick erfolgen, welche relevante Open-Source-Software für den jeweiligen Einsatzzweck und die einzelnen VRE-Services und -Tools in Frage kommt. Dazu empfiehlt es sich, die einzelne Software zunächst isoliert betrachtet, praktisch zu testen und im Anschluss auf die möglichst „nahtlose“ Einbindung in eine L-TER VRE zu überprüfen. Häufig lässt sich allgemein ausgerichtete Software in ihrem Einsatzzweck flexibel und individuell an die Anforderungen anpassen und konfigurieren.

Gründe, die in der Regel für den Einsatz von „nachhaltiger“ Open-Source-Software für L-TER VREs sprechen:

- Meist vergleichbarer technischer Stand wie proprietäre Software,
- Kosteneinsparung bei der Anschaffung,
- offener Quellcode,
- offene internationale Standards, Schnittstellen und Datenformate,
- Nachhaltigkeit und Flexibilität durch Unabhängigkeit von kommerziellen Herstellern,
- oft große, spezialisierte Entwicklergemeinschaft,
- oft schnelle Entwicklungszyklen,
- Open Source als „Ökosystem“ für Weiterentwicklungen z.B. in Form von „Open-Source VRE-Frameworks“ (vgl. Kapitel 2.4).

Bei der Auswahl und dem Einsatz von Open-Source-Software für L-TER VREs muss folgendes beachtet werden:

- Software muss auf dem Stand der Technik sein,
- einzelne Bausteine müssen sich technisch nahtlos und ggf. vom Layout einfach in die L-TER VRE integrieren lassen,

- große, spezialisierte und gut organisierte Entwicklergemeinden sind meist von Vorteil für die gesamten Softwareeigenschaften,
- schnelle Entwicklungszyklen sind ein Indiz für die Aktualität der Software, ggf. aber auch für Fehler,
- Kompatibilität mit anderer Software, Versionen und Komponenten ist zu gewährleisten.

Problematisch sind folgende Aspekte:

- erhöhte Sicherheitsrisiken sind durch weite Verbreitung der Software gegeben,
- Gefahr der Auswirkung von Sicherheitslücken in einzelnen Modulen auf das gesamte System,
- ggf. Abhängigkeit von der Community (relativ, da offener Quellcode),
- intensiv in der Administration durch häufige Updates.

Konkrete Softwareempfehlungen im Anhang in Kapitel C, D und F sollen Zeit bei der Suche nach geeigneter und untereinander kompatibler Open-Source-Software für VREs einsparen. Bei der Auswahl der Empfehlungen wurden die hier genannten Aspekte für nachhaltige Open-Source-Software berücksichtigt.

4.2 ICT-Basis für L-TER VREs

Für das Konzept der L-TER VREs wird eine webbasierte Informationsstruktur gewählt, die auf Webtechnologien und offenen Standards nachhaltig aufbaut. Die in Kapitel 2.7 beschriebene ICT ergänzen das Konzept. Insbesondere Technologien wie Cloud Computing, Virtualisierung, CMS, Datenbanken, GIS und Langzeitdatenspeicherung bilden die Basis. Je nach Anforderungen werden nicht unbedingt immer alle Technologien benötigt, wie z.B. Remote Desktop, Screen- und Application Sharing, HPC und Grid-Computing, Dataspace- und Semantic-Web-Modelle, sowie SOA. Die Klassen von relevanten ICT-Basistechnologien für L-TER VREs sind in Tabelle 4.1 mit Beispielen aufgeführt (vgl. Kapitel 2.7):

Tabelle 4.1: ICT Klassen mit Beispielen für L-TER VREs

L-TER VRE Technologie-Klasse	Beispiele
Authentifizierung und SSO	SAML (Security Assertion Markup Language), LDAP (Verzeichnisserver)
Cloud Computing	PaaS – Platform as a Service, SaaS – Software as a Service, z.B. „OwnCloud“
CMS	Joomla, Typo3, Drupal, Plone, Wordpress
Dataspace- und Semantic-Web-Modelle	RDF, OWL, INSPIRE

L-TER VRE Technologie-Klasse	Beispiele
Datenbanken und DBMS	SQL, NoSQL, GeoNetwork als Metadatenbanksystem, GIS
Dateneinbettung	Mash-up-Technologien, iFrame, HTML5
Direkte Webtechnologien	HTML, http, https
Hochleistungsrechnen und GRID-Computing	OpenMP, OpenMPI
Langzeitdatenspeicherung	DOI, DNA, spezielle Formate (z.B.: PDF-A),
Metadaten und Auszeichnungssprachen	XML, Dublin Core, SensorML, WaterML, KML, GML
Programmiersprachen	PHP, Python, C, Perl, Java, R
Remote Desktop, Screen- und Application-Sharing, Videokonferenz	Windows Remote-Desktop, VNC, X-Windows, Skype, HTML5 (browserbasiert), Adobe Connect
Server	Betriebssysteme (OS), OS Services
Social Networks	Facebook, Twitter (basiert u.a. auf Web 2.0)
Virtualisierung	VM-Ware, Oracle Virtual Box
Web 2.0 (und weitere Versionen)	Java-Script, jQuery, AJAX

4.3 „Schlüsselfertige“ Server und Virtualisierung

Um schnelle und nachhaltige Erfolge bei der Realisierung von L-TER VREs und deren Prototypen zu ermöglichen, sieht das Konzept den Einsatz von „schlüsselfertigen“ Servern und Technologien zur Virtualisierung (vgl. Kapitel 2.7.3) vor. Die „schlüsselfertigen“ Server sind (auf der Basis von Open-Source-Software) vorinstallierte Server mit Betriebssystem und einer spezifischen Software (VRE-Service), die in Form z.B. eines *Images* (Datenträgerabbild) oder einer virtuellen Maschine (VM) einsatzbereit im Internet z.B. bei dem Anbieter „Turnkeylinux“ (turnkeylinux.org 2017) kostenlos zur Verfügung stehen. „Linux“-Betriebssysteme (Open Source) haben sich bei Web-Servern als Standard etabliert (Europäische Kommission 2006). Die Images werden auf die eigene Hardware übertragen oder in die Virtualisierungssoftware importiert. Dazu ist weniger Zeit und ggf. auch etwas weniger fachspezifisches Know-how im Vergleich zur manuellen Einrichtung und Konfiguration eines Servers erforderlich. Die „schlüsselfertigen“ Server oder VMs stellen dann die einzelnen VRE-Services für die L-TER VRE bereit. Durch Virtualisierung wird erreicht, dass die Systeme schneller realisiert werden können, unabhängiger von der eingesetzten Hardware sind, Hardware eingespart werden kann, sich besser in der Administration handhaben lassen und vorteilhaft für die Weiterentwicklung sind.

Für die schlüsselfertigen Server ist die Grundkonfiguration in der Regel bereits vorgenommen. Jedoch müssen individuelle Anpassungen der Servereinstellungen und die Konfiguration des Betriebssystems (z.B. die IP-Adresse oder die Firewall), sowie der

speziellen Software, vorgenommen werden. Die Betriebssysteme und Basissoftware als Grundlage für die verschiedenen schlüsselfertigen Server von ein und demselben Anbieter basieren meist auf dem gleichen Grundsystem (Betriebssystem mit Serverkomponenten), das sich nur in der VRE-Service-Software (Anwendung) unterscheidet. Dadurch ist die Konfiguration und Pflege der Systeme einheitlich.

Für einige VRE-Services, wie z.B. „Remote-Desktop-Services“, bieten sich Betriebssysteme, wie „MS Windows“, „iOS“, „Linux“ mit „X-Windows“ oder „Android“ an. Dies betrifft etwa Anwendungsfälle, in denen Benutzer auf „MS Windows“ kompatible Anwendungen auf einem Windows Server als VRE-Service zugreifen sollen.

Für VRE-Services empfiehlt es sich, unterschiedliche physikalische oder virtuelle Server zu verwenden. Dadurch werden die einzelnen VRE-Services hinsichtlich ihrer Modularisierbarkeit besser voneinander getrennt und bei Ausfall eines VRE-Services oder Servers werden andere weniger oder gar nicht beeinträchtigt. Auch bei der Konfiguration hat dies Vorteile, da die einzelnen VRE-Services oftmals verschiedene Konfigurationen benötigen. Andererseits ist dadurch der Hardwarebedarf größer, da mehrere physikalische oder virtuelle Server zum Einsatz kommen, die jeweils Ressourcen für ein eigenes Betriebssystem benötigen. Eine Lösung für dieses Problem bieten spezielle Technologien zur Virtualisierung mit einem gemeinsamen Overhead für das Betriebssystem (vgl. Kapitel 2.7.4). Bei der Planung der Infrastruktur sollte daher darauf geachtet werden, welche VRE-Services sich z.B. in Bezug auf Sicherheit und Performance gut kombinieren lassen und welche besser auf separaten Servern laufen. Dabei sind Ausfallsicherheit, Hardwareressourcen und Zeitbedarf für die Implementation und Pflege gegeneinander abzuwiegen.

Für die „Langlebigkeit“ (Persistenz) von L-TER VREs hat die Virtualisierung eine besondere Bedeutung. Bei Hardwareaustausch und neuen Hardware-Technologien können die L-TER VREs mit minimalem Aufwand auf ein neues System wechseln und einzelne Entwicklungsstadien als „SnapShot“ abgespeichert werden. Für die Skalierung einer L-TER VRE, bedingt durch schwankende Nutzerzahlen, wird ein Lastenausgleich oder eine dynamische Anpassung der Systemperformance durch Virtualisierung vereinfacht. Auch bei der Datensicherung bietet die Virtualisierung Vorteile für die Erstellung von Backups. Für die Weiterentwicklung bietet die Virtualisierung den Vorteil von standardisierten, lauffähigen „Gesamtsystemen“ (Laufzeitumgebung) mit Speicherung verschiedener Entwicklungsschritte in Form von „SnapShots“. Eine Liste von einsetzbarer Software zur Virtualisierung und schlüsselfertigen „Linux“-Servern finden sich im Anhang C (Tabelle C.3) und F.

4.4 Homepage und CMS

Die Homepage (Website) eines Forschungsprojektes steht mit ihrer zentralen Startseite im Mittelpunkt des gesamten Internetauftrittes zum Projekt – sowohl in der Außendarstellung (Internet), also auch intern (Intranet). Sie sollte daher als erster Punkt bei der

Realisierung einer L-TER VRE mit Hilfe eines *Content-Management-Systems* (CMS) als Basis und dessen optionalen Erweiterungen implementiert werden (vgl. Kapitel 2.7.5). Viele dieser Erweiterungen sind ebenfalls Open-Source-Software. Bei manchen lohnt sich die Aufwendung eines meist geringen Betrages für Erweiterungen mit kostenpflichtigen Lizenzen. Dies widerspricht nicht unbedingt der Philosophie für L-TER VREs, Open-Source-Produkte für die *Basis-Services* und -Tools zu wählen, da es sich hier um untergeordnete Bereiche (nicht die Basis) handelt und dadurch wenig Anschaffungskosten sowie Abhängigkeiten insgesamt entstehen. CMS-Erweiterungen werden folgend im Kontext von L-TER VREs auch als VRE-Tools bezeichnet, da hier das CMS den VRE-Service laut Definition darstellt.

CMS sind das „Standard-Werkzeug“ zur Implementierung und für den Betrieb von Homepages, von denen aus Verweise zu allen weiteren VRE-Ressourcen, -Services und -Tools erfolgen. Ihre Flexibilität ermöglicht es meist ohne Programmieraufwand, andere ICT-Systeme einzugliedern, so dass die resultierenden komplexen Systeme wie „aus einem Guss“ erscheinen. Dies erfordert in der Regel einen zentralen Login durch „Single Sign-on“-Funktionen, der alle Systeme übergreifend miteinander verbindet. Das CMS, die VRE-Services und -Tools müssen dabei die gleichen Standards für Authentifizierung und SSO unterstützen (vgl. Kapitel 2.7.9). Auch individuelle Datenbankanwendungen (vgl. Kapitel 2.7.6) lassen sich mit VRE-Tools meist ohne zusätzlichen Programmieraufwand bewerkstelligen. Durch die große Anzahl an vorhandenen, flexiblen Erweiterungen für die einzelnen CMS, sind individuelle VRE-Tools ebenso ohne Programmierung realisierbar. Sollten spezielle VRE-Tools ohne zusätzliche Programmierung nicht möglich sein, sind diese Systeme aufgrund des gegebenen Frameworks und der eingesetzten Standards offen für vielseitige Anpassungen. „Generatoren“ in Form von CMS-Erweiterungen können den Aufwand für die Programmierung *individueller* VRE-Tools durch automatisierte Anwendungsroutinen erheblich vereinfachen. Der Einsatz von modular aufgebauten Open Source CMS auf dem Stand der Technik ist somit eine effektive Basis zur Implementation und Betrieb einer L-TER VRE.

Die Grundfunktionen eines CMS zur Erstellung von Homepages wurden bereits erläutert (vgl. Kapitel 2.7.5). Darüber hinaus hat eine Homepage für L-TER Projekte beispielsweise die folgenden besonderen Funktionen:

- Einstiegspunkt für Benutzer und Öffentlichkeit (Intranet bzw. Internet),
- allgemeine Informationen zum Projekt,
- Informationen zu Betreibern und Benutzern,
- rechtliche Informationen zur Benutzung, Datenschutz, Data-Policies
- zentraler Zugang zu allen Ressourcen der VRE,
- Community und „Social Networks /-Media“,
- globale Suchfunktionen,
- Berichte für die Öffentlichkeit,
- Suchmaschinenoptimierung,
- Linksammlungen und Kataloge,

- Übersicht zu Publikationen,
- Statistiken und Kennzahlen, Reports,
- Hilfe und Dokumentation für Benutzer und Entwickler,
- Öffentliche und interne Umfragen,
- Akquise von wissenschaftlichem Personal,
- Projektmanagement der L-TER Projekte mit Meilensteinen,
- Langzeitpersistenz.

Die Auswahl eines geeigneten CMS ist bei derzeit ca. 250 – 300 verschiedenen Systemen (ks-webagentur.de 2017) keine leichte Entscheidung. Abhängig sollte die Auswahl von Einsatzzweck, Vorkenntnissen der Entwickler, verfügbaren Erweiterungen, der Community und der Systemarchitektur sein. Auch Sicherheitsaspekte, Skalierbarkeit und Lastverteilung spielen eine wichtige Rolle. Die vorgestellten „Top5“ Open-Source-CMS (vgl. Tabelle C.5 im Anhang C) haben den größten Marktanteil (Breitenstrom et al. 2013), wobei sie jeweils einen leicht unterschiedlichen Schwerpunkt aufweisen. Prinzipiell sind die vorgestellten Systeme jedoch recht ähnlich im Funktionsumfang. Was ein CMS als Basisfunktion mitbringt, muss in einem anderen CMS oft zusätzlich als Erweiterung installiert werden. Sicherheitsanalysen der Studie zur Sicherheit von CMS des „Bundesamtes für Sicherheit in der Informationstechnik“ (Breitenstrom et al. 2013) zeigen klar, dass ein Großteil der Schwachstellen in den Erweiterungen der CMS zu finden ist. Die Schwachstellen der Kernkomponenten werden meist schnell durch regelmäßige Update-Zyklen geschlossen. „WS-Security“ (Protokoll für Sicherheit in Webservices) wird von keinem der CMS unterstützt (Breitenstrom et al. 2013). Eine Lastverteilung wird bei allen PHP-basierten CMS serverseitig realisiert.

Als Ergebnis von über 10 Jahren praktischer Erfahrungen des Autors mit dem CMS „Joomla“, sowie den Studien zu dieser Arbeit, findet sich im Anhang D ein „VRE-Tool Baukasten“ in Form einer Liste mit ca. 30 der besonders geeigneten „Joomla“ (joomla.org 2017) Erweiterungen (über 5.000) für den Einsatz in einer L-TER VRE. Diese Empfehlungen können sehr hilfreich sein, da die Anzahl der verfügbaren Erweiterungen unübersichtlich groß ist und es viele ähnliche Erweiterungen mit gleichem Schwerpunkt gibt. Die Suche nach geeigneten Erweiterungen und deren praktischer „Eignungstest“ sind ansonsten zeitaufwendig.

4.5 VRE-Services

Laut der Definitionen (vgl. Kapitel 3.1) bestehen L-TER VREs aus VRE-Services, -Tools und -Ressourcen, wobei in diesem Abschnitt näher auf die VRE-Services und das „Baukastenkonzept“ eingegangen werden soll. VRE-Services machen die individuellen Eigenschaften einer L-TER VRE aus, die vorher durch die Zieldefinitionen des Projektes festgelegt wurden (vgl. Kapitel 2.5.7). Services können multidisziplinär oder fachspezifisch sein und sollten sich in eine Gesamtarchitektur eingliedern. Sie können modular, je nach Anforderungen in die L-TER VRE nachträglich eingebunden werden.

Modulare und serviceorientierte Architekturen (SOA) sowie „Cloud Computing“ unterstützen die Einbindung der Services, indem z.B. verteilte Dienste durch verschiedene Domänen bereitgestellt werden können (Elsayed 2012). Die Aufgaben der einzelnen VRE Services leiten sich aus den grundlegenden Anforderungen an VREs (vgl. Kapitel 2.7), den technischen Anforderungen an L-TER VREs (vgl. Kapitel 2.5.10), den weiteren Anforderungen an L-TER VREs (vgl. Kapitel 3) und den projektspezifischen Anforderungen des L-TER Projektes ab.

Das „Baukastenkonzept“ für „Rapid L-TER VREs“ stellt dazu einen zentralen „Baukasten“ mit VRE-Services und Beispielsoftware bereit, der nach Aufgabenbereichen der einzelnen VRE-Services gruppiert ist (Tabelle 4.2). Dieser Baukasten soll als Grundlage für individuelle L-TER VREs dienen. Eine schnelle Implementierung der VRE-Services kann durch den Einsatz „schlüsselfertiger“ Server und durch virtuelle Maschinen erfolgen (vgl. Kapitel 4.3). VRE-Services können durch servicespezifische VRE-Tools im Funktionsumfang erweitert werden.

Dazu finden sich im Anhang C, D und F als Bestandteil des Konzeptes konkrete Softwareempfehlungen zu VRE-Services, CMS und CMS-Erweiterungen sowie „schlüsselfertigen“ Servern.

Tabelle 4.2 gliedert die Hauptgruppen für VRE-Services mit Beispielen.

Tabelle 4.2: Aufgabenbereiche für VRE-Services in L-TER VREs mit Aufgaben und Beispielen

Aufgabenbereiche	Aufgaben und Beispiele für L-TER VRE-Services
Administratoren	Server (z.B. „Linux“), Virtualisierung (z.B. „Virtual Box“), Dokumentation, Backup, Sicherheit und Überwachung des IT Systems: Firewall, Serverstatus, Datenbankstatus, Fehlermeldungen, Angriffsmeldungen, Updateverwaltung, Leistungsdaten, etc. Administrationsoberflächen („Backend“) der jeweiligen VRE-Services
Authentifizierung, SSO	Zentrale Benutzerverwaltung und Authentifizierung, SSO (z.B. SAML), Verzeichnisserver (z.B. LDAP), CMS
Cloud	Zentrale Datenspeicherung, Software und Infrastrukturen aus der Cloud, z.B. „OwnCloud“
Datenverbände	Automatisierter (Meta-)Datenverbund mit internen und externen Netzwerken (Standards, Datenformate, Schnittstellen, Markup-Languages), z.B. „GeoNetwork“
Datenbank-Management-Systeme (DBMS)	Zentrale Datenbanken in einem DBMS, z.B. SQL („MySQL“, „PostgreSQL“, NoSQL (z.B. „MongoDB“))
Datenmanagement und Langzeitarchivierung	Metadatenkataloge (z.B. „GeoNetwork“), Langzeitdatenspeicherung, Suche, Pflege, Datenstrukturen, Semantic Web, Persistente Identifikatoren, automatisierte Formatkonvertierung in Archivformate (z.B. PDF-A), automatisierte Metadatenerzeugung
Datenverarbeitung, -aufbereitung und -analyse	Echtzeit-Monitoring, pre- und post-processing, statistische Auswertungen, Datenvisualisierung, Statistik-Frameworks (z.B. R-Statistics)
E-Learning	E-Learning Plattformen (z.B. „Moodle“), interaktive Lernanwendungen, Plattformen für Lehre und Studierende, Wissensvermittlung, Wissensüberprüfungen, Multimedia-Tutorials

Aufgabenbereiche	Aufgaben und Beispiele für L-TER VRE-Services
Entwickler	Dokumentation, Benutzerfeedback, Quellcode-Repositoryen (z.B. „CodeGIT“), Virtualisierung, Bug-Tracking (z.B. „BugZilla“),
Geografische Informationssysteme (GIS)	Räumliche Datenverarbeitung, Geografische Karten, -Daten, -Analysen, z.B. Geodienste: WCS, WMS, WFS, WPS, WTS, CSW
Gruppen- und Rechtemanagement	CMS oder jeweiliger VRE-Service
Kommunikation und Web 2.0	E-Mail, Messenger, Videokonferenzen, Boards, Umfragen, Bewertungen, Wiki
Modellierung und Simulation	Software für Modellierung und Simulation mit Metadaten (Parameter, Verfahren, Ergebnis, etc.), z.B. „Netlogo“, „R-Statistics“
Projektmanagement und Teamwork	Gruppenkalender, Aufgaben, Termine, Meilensteine, Stellenbörsen, Angebote für wissenschaftliche Arbeiten, Finanzen, Personal, Evaluation, Wiki (z.B. „MediaWiki“), Dokumentation und Hilfe, Quellcode-Repositoryen
Publikationsmanagement	Liste und Archiv der Publikationen mit geeigneter Struktur und Suche
Reporting	Berichte für spezielle Zielgruppen aufbereitet, Entscheidungsunterstützung, z.B. ein automatisierter Report der Handlungsempfehlungen für Entscheidungsträger.
Remote-Services	Ferngesteuerter Zugriff auf andere Computer und Software für Teamwork-Unterstützung und spezielle, schwer zu integrierende Software, z.B. „RDP“, „X-Window“, HTML5
Virtualisierung	Virtualisierung der Infrastruktur, Virtuelle Maschinen, z.B. „Virtual Box“, „VM-Ware“
Webserver / Betriebssystem	Betriebssysteme (z.B. „Linux“), Sicherheitskonzept, Serveradministration, Loadbalancing, Web-/Datenbankserver
Social-Media, -Networks, -Communities	Social-Media-Plattformen, Communities, dynamische Gruppen, digitale Forschungsgemeinschaften, Nachrichten, Timelines, Bewertungen, Umfragen, Bilder, Videos, Events, Benutzerprofile, Integration von anderen Social-Networks wie z.B.: „Facebook“, „Twitter“, „LinkedIn“, „JomSocial“ für „Joomla“ CMS
Website und CMS	Basis-Software für die VRE (z.B. „Joomla“): Bereitstellung von Websites, Suche, Suchmaschinenoptimierung, Einbindung weiterer VRE-Services und -Tools in die VRE

4.6 VRE-Tools

VRE-Tools stellen laut Definition (vgl. Kapitel 3.1) die funktionellen Erweiterungen zur Bearbeitung des Forschungsprozesses innerhalb einzelner VRE-Services dar. Die VRE-Services sind mit spezifischen Basis-Tools (Kernwerkzeuge) ausgestattet, die durch VRE-Tools (Erweiterungen) modular ergänzt werden können. Für die VRE-Services steht in der Regel eine Vielzahl an allgemeinen Erweiterungen bereit, die für den konkreten Einsatz in L-TER VREs als VRE-Tools individuell und flexibel zum

Einsatz kommen können. Die Aufgabenbereiche der VRE-Tools ergeben sich aus den jeweils übergeordneten VRE-Services.

Da CMS die Basis für L-TER VREs bilden, sind CMS-Erweiterungen (VRE-Tools) von besonderer Bedeutung. Das gleiche gilt für Wikis. Sie können sehr flexibel an verschiedenen Stellen innerhalb einer L-TER VRE die Projektarbeit im Team unterstützen. Das „Joomla“ CMS und „MediaWiki“ verfügen aufgrund der weiten Verbreitung und der großen Entwicklergemeinschaft über zahlreiche Erweiterungen. Diese können als VRE-Tools von Nutzen sein, wie z.B. der „Visual-Editor“ (WYSIWYG) oder „R-Statistic“, die nicht standardmäßig zu den Kernkomponenten der Open-Source-Software „MediaWiki“ gehören. Im Rahmen des „Rapid L-TER VRE“ Baukastenkonzeptes finden sich im Anhang ergänzend Listen und Beispieltabellen zum „Joomla“ CMS (Anhang D) und zu „MediaWiki“ (Anhang C, Tabelle C.8), die den Baukasten mit VRE-Services um spezifische VRE-Tools erweitern.

VRE-Tools können in L-TER VREs (als gesamtheitlicher Service betrachtet) zum Beispiel auch in Form von *Middleware* eingesetzt werden, um VRE-Ressourcen, -Services und -Tools untereinander zu verbinden, sofern diese keine oder unzureichende Kernfunktionalitäten oder Erweiterungen derart aufweisen. Middleware (aus dem Englischen übersetzt, Zwischenanwendung) bezeichnet in der Informatik Programme, die anwendungsneutral zwischen anderen Programmen vermitteln und Daten über gemeinsame Schnittstellen (engl. Interfaces) austauschen (Etzkorn 2017). Sie treten dabei meist nicht offensichtlich in Erscheinung, wobei ihre Infrastruktur und Komplexität dem Anwender „verborgen“ bleiben.

Das Angebot an verfügbaren Erweiterungen der verschiedenen Services ist jedoch nicht speziell hinsichtlich L-TER VREs ausgelegt. Dennoch lassen sich viele Basisfunktionalitäten einer L-TER VRE direkt oder durch kleine Änderungen im Quelltext bzw. spezifische Konfigurationen der Tools umsetzen. Auch mit servicespezifischen Scriptsprachen lassen sich gegebenenfalls individuelle VRE-Tools auf einfache Weise realisieren. Wo dies nicht gelingt, ist die Entwicklung eigener VRE-Tools erforderlich. In den folgenden Bereichen sind aufgrund der Individualität, des Umfangs und der fachlichen Ausrichtung von VREs, beispielsweise wenige Erweiterungen zu erwarten, die als L-TER VRE-Tool geeignet erscheinen:

- Datenaufbereitung, Standardisierung, Datenauswertung
- Austausch von Daten mit anderen VRE-Services,
- VRE-übergreifende Indexierung und Suche von Ressourcen,
- Kopplung von Modellen und Datenstrukturen z.B. mit Tensor-Objekten und OpenMI,
- Reportfunktionen zur Berichterstattung,
- Langzeitarchivierung von Daten, Formaten und Umgebungen.

In diesen Bereichen entsteht meist ein großer zeitlicher Aufwand an individueller Programmierung. Für VRE-Tools in Form von Middleware eignen sich spezielle

Frameworks, wie z.B. COBRA (corba.org 2017). Als Software für die Programmierung von VRE-Tools kommen z.B. Script- oder Programmiersprachen, wie JAVA(Script), PERL, PYTHON, PHP oder C++ zum Einsatz, um die gewünschten VRE-Tools zu implementieren. Die Programmierung der VRE-Tools sollte nach Möglichkeit nicht direkt in andere Komponenten eingreifen, um Updates dieser einfacher einspielen zu können und den Quellcode übersichtlich zu halten. Standardisierte Schnittstellen und spezifische Software-Architekturen (z.B. SOA, REST, Modularisierung) (vgl. Kapitel 2.7.8) fördern dabei die Nachhaltigkeit der L-TER VRE durch eine einheitliche Struktur.

4.7 Authentifizierung und Einmalanmeldung (SSO)

L-TER VREs haben auf Grund ihrer längeren Nutzungszeit und Interdisziplinarität im Gegensatz zu normalen VREs meist mit einer relativ „großen“ Benutzeranzahl und „wechselnden“ Benutzern zu rechnen. Dazu sind verschiedene Zugriffsrechte auf einzelne Bereiche bzw. verschiedene (Teil-)Projekte nötig. Auch die Anzahl der eingebunden Institutionen ist meist größer, als bei anderen Projekten. Daher muss eine L-TER VRE diesen Ansprüchen einer kooperierenden und Institutionen übergreifenden Authentifizierung mit SSO und abstimmbaren Zugriffsrechten genügen. Das Konzept baut daher auf eine zentrale Authentifizierung mit SSO.

Die Authentifizierung von Benutzern in der VRE sollte einfach, aber sicher erfolgen. Da VREs aus unterschiedlichen Komponenten bestehen, ist eine einheitliche Authentifizierung mit Einmalanmeldung (SSO) für alle VRE-Services eine besondere Herausforderung – insbesondere da nicht alle VRE-Services die gleichen Standards aufweisen und jede Software (VRE-Service) meist eine eigene Authentifizierung als Kernkomponente beinhaltet. Bei der Planung einer VRE muss daher darauf geachtet werden, dass sich alle VRE-Services zentral über eine SSO-Authentifizierung technisch verbinden lassen. In diesem Zusammenhang können Verzeichnisdienste (wie z.B. nach dem LDAP-Standard) eine zentralisierte Benutzerverwaltung erheblich unterstützen.

Bei der konkreten Auswahl von Software-Komponenten für die L-TER VRE ist darauf zu achten, dass diese bereits über entsprechende Standards zur zentralen Authentifizierung und SSO verfügen. Eine nachträgliche Aufrüstung mit diesen Mechanismen kann komplex und zeitaufwendig sein, sofern keine Erweiterungen (VRE-Tools) dafür vorhanden sind.

In Verbindung mit den Verzeichnisdiensten ist SSO ein entscheidender Faktor, um einzelne VRE-Services miteinander zu kombinieren und dadurch aus der Sicht des Anwenders „monolithische“ VREs und Systeme – wie aus einem „Guss“ – zu erzeugen. Er erfordert viel spezifisches Know-how u.a. im Bereich der digitalen Kryptographie und erheblichen Aufwand in der Konfiguration. Im Anhang C findet sich ein Überblick zu oft genutzten Verzeichnisdiensten (Tabelle C.6) und SSO-Produkten (Tabelle C.7).

4.8 Report-Layer

Im potentiell langen Nutzungszeitraum von L-TER VREs müssen vielfach verschiedene Daten und Metadaten für unterschiedliche Nutzergruppen und (Teil-)Projekte, regelmäßig und nach Abruf, aufbereitet werden. Daraus lassen sich z.B. Berichte (engl. Reports) zu Forschungsergebnissen erstellen oder der Erfolg einer L-TER VRE kann durch die Hilfe von Kennzahlen (z.B. Anzahl der Datensätze oder der Publikationen) evaluiert werden (vgl. Kapitel 2.5.7). Um diese Aufgaben zusammenzufassen, sieht das Konzept einen „Report-Layer“ vor, der die Bedeutung dieser Aufgaben für die Nachhaltigkeit von L-TER VREs verdeutlicht.

Der Vorteil einer VRE liegt unter anderem darin, dass die Informationen einer VRE im „Ganzen“ und im Zusammenhang schneller durch Automatismen ausgewertet, bewertet und zusammengestellt werden können, als in verteilten und autarken Systemen. Dabei ist eine zielgruppengerechte Aufbereitung und Darstellung der Daten sowie Ergebnissen von besonderer Bedeutung (Ruchter 2002).

Die Aufbereitung der Ergebnisse kann z.B. nach den folgenden Zielgruppen erfolgen: Benutzer, Administratoren, Entwickler, Betreiber, Förderer, Öffentlichkeit, Entscheidungsträger, Datenverbände (Abbildung 4.2).

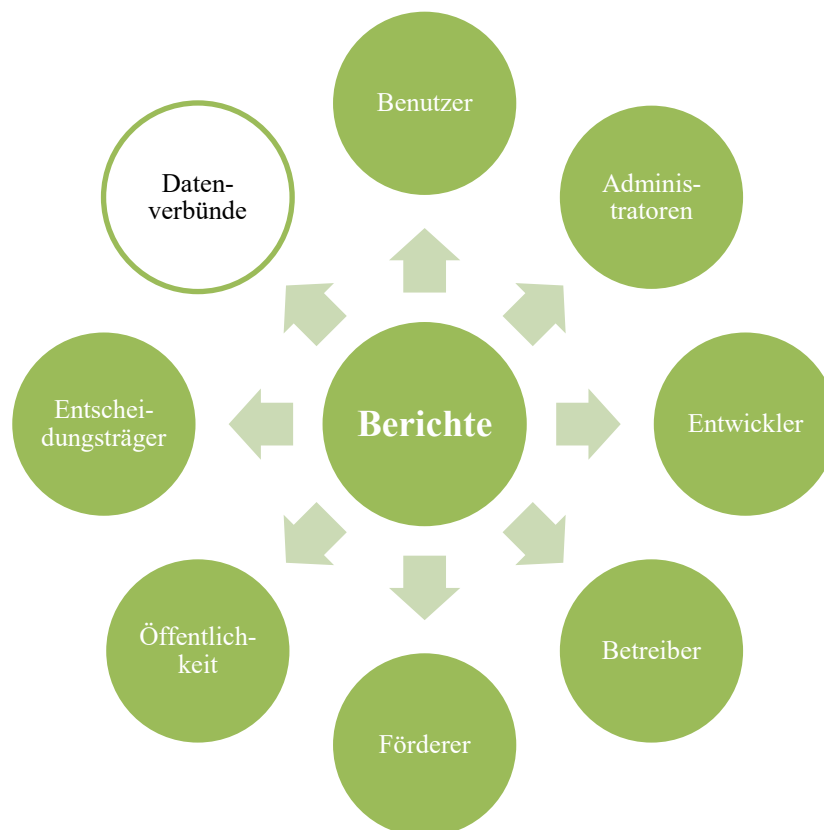


Abbildung 4.2: Report-Layer erstellt Berichte für verschiedene Zielgruppen und Datenverbände

Der „Report-Layer“ ist ein übergeordneter Service der gesamten L-TER VRE, der auch aus mehreren zusammengefassten VRE-Services und deren -Tools bestehen kann (Abbildung 4.3). Er bietet Unterstützung bei der manuellen und automatischen Erstel-

lung von zielgruppenorientierten Kennzahlen, Visualisierungen und Berichten sowie für den interoperablen Datenaustausch in Datenverbänden (vgl. Kapitel 2.7.10).

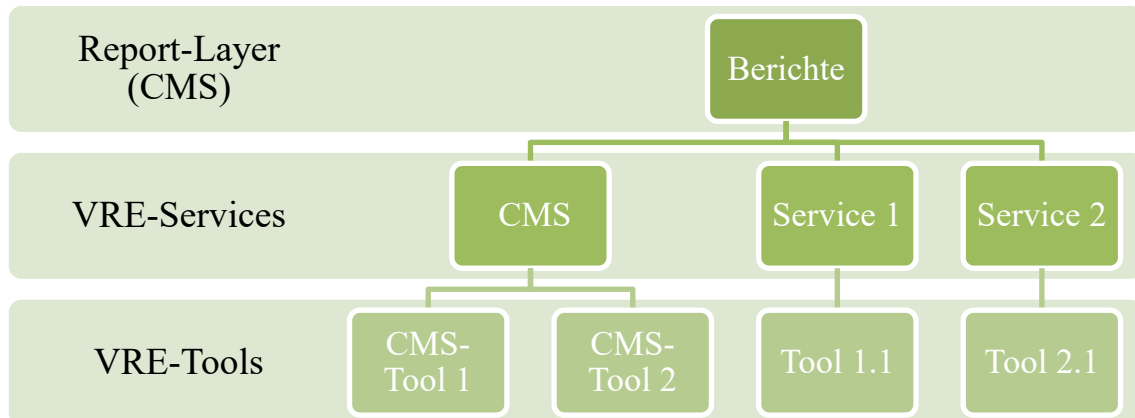


Abbildung 4.3: Report-Layer stellt Berichte aus einzelnen VRE-Services und-Tools

Bei „einfachen“ Sachverhalten sollte die Zusammenfassung in Form von Kennzahlen und die Berichtserstellung automatisch erfolgen. Diese einfache Funktionalität ist in vielen verfügbaren VRE-Services nativ und konfigurierbar durch den Administrator oder Anwender integriert. Viele dieser VRE-Services beinhalten bereits derartige Tools für ihren eigenen Kontext (z.B. CMS: Nutzeranzahl, Anzahl der Beiträge und deren Bewertungen). Bei „komplexeren“ Forschungsberichten kann unter weiterem programmiertechnischen Aufwand ein (teil-)automatisierter Prozess implementiert werden, der je nach Anforderungen durch den Programmierer, Administrator oder Benutzer konfiguriert werden kann.

Auch Datenverbände im weltweiten Netzwerk können einen erheblichen Erfolgsfaktor für die L-TER VRE und die Forschungsgemeinschaft darstellen, da die Wahrnehmbarkeit außerhalb des Projektes zunimmt. VREs sind der zentrale Sammelpunkt für Daten aus unterschiedlichen Quellen und Institutionen. Dabei müssen die Daten nicht zwangsläufig direkt in einer VRE gespeichert sein. Es genügt ggf. eine zentrale Metadatenbank zu pflegen, die auf die Quelle der eigentlichen Daten verweist. Dadurch können verschiedenen „Datenlieferanten“ die „Hoheit“ über ihre Datensätze behalten und die Nutzungs- sowie Zugriffsrechte können gewahrt werden. Um eine derartige Metadatenbank für L-TER VREs aufzubauen, ist das OAI-PMH Protokoll (vgl. Kapitel 2.7.10) geeignet, da hier offene Schnittstellen und Standard Web-Technologien zum Einsatz kommen (Rusch-Feja 2002). Zum Beispiel baut die Software „GeoNetwork“ auf diesem Standard auf und kann als Metadaten-Management-System für INSPIRE-konforme (vgl. Kapitel 2.6), interoperable Datenverbände genutzt werden. Auf diese Weise können Metadaten-Kataloge der Datenbestände in weltweiten Netzwerken von großen und verzweigten Hierarchien einfach zusammengeführt und aufgebaut werden. Dabei ist die Abbildung hierarchischer Struktur möglich, wie z.B. Landesnetzwerke, kontinentale Netzwerke, Netzwerke von Weltverbänden.

Zu den übergeordneten Aufgaben des Report-Layers zählen:

- Zielgruppengerechte Aufbereitung von Informationen,
- Zusammenfassung von Metadaten,
- Zusammenfassung von Daten,
- Erstellung von Kennzahlen aus den Metadaten,
- Visualisierung von Metadaten und Daten,
- Erstellung eines Report-Layouts,
- Unterstützung bei der manuellen oder automatischen Erstellung von Reports,
- Textgenerierung für wiederkehrende Reportbausteine,
- Auswertung von Feedback, Umfragen, Bug-Meldung zur VRE selbst,
- interoperabler Datenaustausch in Datenverbänden.

Der Report-Layer umfasst (Meta-)Daten, Kennzahlen und Visualisierungen von:

- der VRE selbst und von ihren Ressourcen,
- den eigentlichen Forschungsdaten,
- einzelnen oder mehreren Projekten.

Beispiele für Parameter des Report-Layers:

- Anzahl: Benutzer, Datensätze, Projekte, Veröffentlichungen, Kommunikation (Nachrichten),
- Themenschwerpunkte und Relationen,
- spezifische Mittelwerte von Messdaten,
- spezifische Kennzahlen von Projekten.

Für die Implementation des Report-Layer eignen sich z.B. CMS, welche die verschiedenen VRE-Services und deren Report-Tools bündeln. Unterstützend wirken Middleware mit einer SOA, sowie Methoden zum Parsing von HTML-Dateien (Grune und Jacobs 2004), die vordefinierte Teile aus einer Website anhand ihrer Struktur extrahieren, um sie an einer anderen Stelle der Website oder in einer extra Datenbank wieder einzufügen.

4.9 Benutzer Hilfe, Feedback und technische Dokumentation

Das Wissen über die L-TER VRE (und das L-TER Projekt) sollte für die jeweilige Zielgruppe nachvollziehbar und nachhaltig weitergegeben werden. Insbesondere bei L-TER VREs ist aufgrund des langen Nutzungszeitraumes mit wechselnden Benutzern, Administratoren und Entwicklern auf eine nachhaltige Benutzerhilfe, -Feedback und technische Dokumentation zu achten.

Die Akzeptanz durch die Benutzer ist eine der Voraussetzungen für den Erfolg von VREs (vgl. 2.5.7). Dazu gehört es, den Sinn und die Benutzung einer L-TER VRE zu

verstehen. Neben Einweisungen und Schulungen der Benutzer sollte der Zweck und die Ziele einer L-TER VRE auf der Homepage vermittelt werden. Eine kurze Übersicht mit Beschreibung der angebotenen einzelnen VRE-Services und -Tools hilft besonders Einsteigern bei den ersten Schritten. Wo eine intuitive Bedienung nur schwer umsetzbar ist, kann eine direkte Hilfe innerhalb des Kontextes erfolgen. Weitere Hilfe für Benutzer sollte in L-TER VREs durch Benutzerhandbücher, FAQs, Hilfe-Foren und Online- oder Ticket-Support integriert sein.

Für den wissenschaftlichen Umgang mit Daten müssen allgemeingültige und verbindliche Regeln definiert werden, um eine „gute wissenschaftliche Praxis“ auch in der Realität einzuhalten und den „offenen Zugang“ zu Daten zu gewährleisten (*open access*). Diese sollten in Form von *Data Policies* (vgl. Kapitel 2.5.8) in der L-TER VRE im System abgebildet und angewendet werden. Sie regeln die Aufnahme, die Bereitstellung und den Zugriff auf Daten in der L-TER VRE.

Die administrativen Aufgaben in einer L-TER VRE sind durch die vielen technischen Komponenten umfangreich. Administratoren sollten ihre Arbeit so einfach wie möglich dokumentieren können, damit sie selbst und andere Administratoren die Arbeitsschritte an der VRE nachvollziehen können. Das ist nützlich, um bei gleichen oder ähnlichen Problemen auf effektive Lösungen zurückzugreifen, dient der Arbeitsdokumentation im „betrieblichen“ Sinne und kann als Wissensbasis genutzt werden. Ein Benutzerhandbuch ist auch für Administratoren unumgänglich.

Für Entwickler ist besonders die technische Dokumentation entscheidend für die nachhaltige Weiterentwicklung einer L-TER VRE. Auch Feedbacks der Benutzer geben den Entwicklern Unterstützung z.B. welche Fehler auftreten, welche Bedienung vereinfacht werden muss oder welche VRE-Services und -Tools unzureichend sind oder fehlen. Dazu tragen Bugtracking-, Umfrage- und Bewertungssysteme bei. Auf der Seite der Programmierer sind „Quellcode-Repositoryen“ ein wichtiges Werkzeug für die Projektverwaltung und Entwicklung einzelner Softwarebausteine im Team.

„Wikis“ mit multimedialen Inhalten können als „Team-Komponente“ für die Erarbeitung gemeinsamer Forschungsdokumente genutzt werden und für die Dokumentation genutzt werden. Ihrer Herkunft entsprechend sind sie speziell für den Aufbau von Wissensarchiven entwickelt (vgl. Wikipedia als Enzyklopädie). Sie kommen auch im Umweltbereich zum Einsatz, z.B. in einem Wiki zur Bürgerbeteiligung im Klimawandel, welches durch das BMU gefördert wurde und bei dem das UBA der Projektträger war (KlimaScout 2017). Funktionalitäten von Wikis umfassen eine teamorientierte Entwicklung von wissenschaftlichen Inhalten mit Nachvollziehbarkeit von Änderungen im Allgemeinen und durch einzelne Bearbeiter. Zahlreiche spezifische Erweiterungen können den Funktionsumfang noch erheblich steigern. Wikis sind somit ein wichtiges, individuelles Werkzeug für den Einsatz in VREs.

Dieser Punkt des Konzeptes erfordert, neben der aufwendigen initialen Erstellung der Dokumentationen, Hilfen und technischen Systeme, eine kontinuierliche Pflege

durch Administratoren, sowie die Bereitschaft von Personal für den Support. Dies ist mit erheblichem Zeitaufwand verbunden und beeinflusst die Betriebskosten.

4.10 Weiterentwicklung

Die vorangegangenen Punkte des „Rapid L-TER VRE“ Konzeptes berücksichtigen durchweg den Aspekt der Nachhaltigkeit bei der Entwicklung von L-TER VREs. Als letzter Punkt und Bestandteil des Konzeptes steht die nachhaltige, kontinuierliche *Weiterentwicklung* von L-TER VREs im Fokus.

Der Entwicklungsprozess wird nach dem „Rapid L-TER VRE“ Konzept bei der Planung in „Basis-“ und „erweiterte“ VREs aufgeteilt, um die Entwicklung von Prototypen und Basis-VREs bis zum Produktivbetrieb schnell zu realisieren (Abbildung 4.4).

Zunächst werden in der Basis-VRE die Grundfunktionalitäten (CMS mit Basis-Services und -Tools) geplant und implementiert, die ggf. den Status „Quo“ des Projektes als L-TER VRE gebündelt wiedergeben, sowie *schnell zu realisieren* sind. Schnell zu realisieren sind in der Regel alle „schlüselfertigen“ VRE-Services und deren Tools (vgl. Kapitel 4.3, 4.5, 4.6), die in den Funktionalitäten den geplanten Anforderungen entsprechen. Andernfalls sollten die VRE-Services oder -Tools erst in der erweiterten VRE realisiert werden, sofern sie nicht essentiell für die L-TER VRE sind. Danach kann die Basis-VRE in den Produktivbetrieb übergehen. Die L-TER VRE sollte zunächst nicht zu „komplex“ erscheinen, damit die „Akzeptanz“ der Benutzer langsam wachsen kann. Bewährt sich die Basis-VRE im praktischen Einsatz, erfolgt eine *schrittweise Weiterentwicklung* der L-TER VRE, indem bereits existierende VRE-Services und -Tools verbessert werden und gegebenenfalls neue hinzukommen. Zusätzlich sollen sich durch den iterativen, kontinuierlichen Prozess der Weiterentwicklung, innovativere L-TER VREs mit Hilfe von neuen fachspezifischen- und ICT-Trends entwickeln.

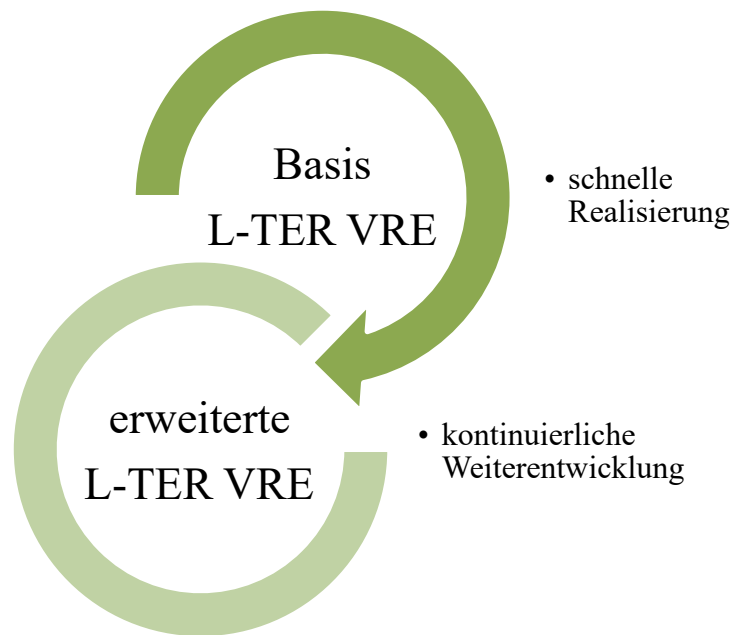


Abbildung 4.4: Aufteilung in „Basis-“ und „erweiterte“ L-TER VRE für schnelle Realisierung und kontinuierliche Weiterentwicklung

Über die bereits beschriebenen Bestandteile des Konzeptes hinaus sind hinsichtlich einer nachhaltigen Weiterentwicklung von L-TER VREs folgende Aspekte zu berücksichtigen:

- auf offene Standards und Open-Source-Software der Zukunft bauen (zukünftiger SdT ungewiss),
- regelmäßige Weiterentwicklung zur Anpassung an neue Standards und Anforderungen,
- nachhaltige Systemarchitektur z.B.: durch Modularisierung und SOA,
- bei Auswahl der VRE-Services und -Tools auf spätere Kompatibilität achten,
- bei der Auswahl von VRE-Services darauf achten, dass später geplante VRE-Tools nach Möglichkeit schon in der Open-Source-Gemeinde zur Verfügung stehen,
- Virtualisierungstechnologien für Testumgebungen mit „Snap-Shots“ bei der Weiterentwicklung nutzen,
- Feedback der Benutzer, Administratoren, Entwickler, Betreiber, Förderer einbeziehen,
- detaillierte Dokumentation von Entwicklungsschritten und -Ergebnissen, sowie Anpassung der allgemeinen Dokumentationen und Handbücher,
- neue fachspezifische- und ICT-Trends verfolgen und relevante integrieren,
- ggf. Open-Source-Community für eigenes L-TER-VRE Framework mit VRE-Services und -Tools gründen.

Durch den iterativen Prozess der nachhaltigen Weiterentwicklung sollen sich L-TER VREs mit erweitertem Funktionsumfang auf Basis von neuen fachspezifischen- und ICT-Trends entwickeln. Tabelle 4.3 zeigt dazu Beispiele, welche zusätzlichen

Technologien und Funktionalitäten für erweiterte VREs Frage kommen können (Tabelle 4.3).

Neben der Administration und dem Support ist die kontinuierliche Weiterentwicklung ein großer Kostenfaktor bei L-TER VREs. Über die lange Nutzungsphase von L-TER-VREs gerechnet, summieren sich die Kosten zur Weiterentwicklung und können die Kosten übersteigen, die zur Erstellung einer Basis-VRE erforderlich sind. Gleichzeitig stehen jedoch durch Förderer meist weniger finanzielle Mittel für die Weiterentwicklung zur Verfügung, als zum Aufbau. Daher sollte ein solides Finanzierungsmodell für die Weiterentwicklung bereits in der Planung ausgearbeitet werden. Vorzugsweise besteht ein interdisziplinäres Entwicklerteam aus einem festen „Kern“, der die einzelnen Entwicklungsstadien der L-TER VRE möglichst über einen langen Zeitraum begleitet. Änderungen der Team-Struktur sollten mit Übergängen und nicht in zu großem Ausmaß erfolgen.

Tabelle 4.3: Beispiele für L-TER VRE Erweiterungen (SdT), gruppiert nach Klassen

Klasse	Beispiele
Datamining/ Softcomputing	Künstliche Intelligenz, Neuronale Netze, Genetische Algorithmen, Partikelschwarmoptimierung, Fuzzy Logic, Wavelets, Agentenbasierte Systeme, Zelluläre Automaten
Hochleistungsrechnen (HPC)	Anbindung an Hochleistungsrechner für besonders rechenintensive Prozesse, wie z.B. für Modellierung und Prognosen, Methoden zur Steigerung der Parallelisierungseigenschaften von Prozessen
Langzeitdatenhaltung	Speichermedien, z.B. DNA-Codierung, Fehlerkorrekturverfahren, „BlockChain“-Technologie
Mobile Apps	Mobil optimierte Websitedarstellung und Kommunikation, Unterstützung von Feldmessungen
Modellierung, Simulation, Prognose	Einbindung von spezifischen Frameworks, Erweiterungen, Schnittstellen zur Modellkopplung und Funktionalitäten, wie z.B. „Netlogo“
Monitoring	Automatisierte Datenübernahme, Echtzeit-Monitoring, innovative Sensortechnologien, funkbasierte Sensornetzwerke, Alarmsysteme für Messwertüberschreitungen, kostengünstige Minicomputer (Raspberry Pi, etc.) für spezielle Aufgaben, Echtzeit-Video-Signalverarbeitung
Visualisierung	Augmented Reality (AR), Virtual Reality (VR), Wortwolken (engl. Tag Clouds), Heatmaps, Beziehungsstrukturen, besondere Graphen und Diagramme

5. Anwendungsbeispiele für „Rapid L-TER VRE“

Das „Rapid L-TER VRE“ Baukastenkonzept wird nachfolgend auf verschiedene existierende L-TER Projekte im Umfeld des Fachgebietes „Environmental Informatics“ der „BTU Cottbus-Senftenberg“ konzeptionell, sowie in der Form von Prototypen angewendet und dessen projektspezifische Anwendbarkeit (Flexibilität) bewertet. Aus der Bewertung wird anschließend in Kapitel 6 die Gesamtbewertung der Anwendungsbeispiele vorgenommen.

Um die projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Baukastenkonzeptes zu überprüfen, wurden L-TER Projekte mit verschiedenen Rahmenbedingungen ausgewählt, welche folgend gruppiert wurden: *Projekt-Größenordnungen*, *-Strukturen*, *-Schwerpunkte*, *-Zeiträume*. Diese haben unter anderem Einfluss auf die eingesetzten Technologien. Die spezifischen Eigenschaften der ausgewählten L-TER Projekte decken ein kleines bis mittleres Spektrum an Rahmenbedingungen „weitgehend“ ab (Abbildung 5.1).

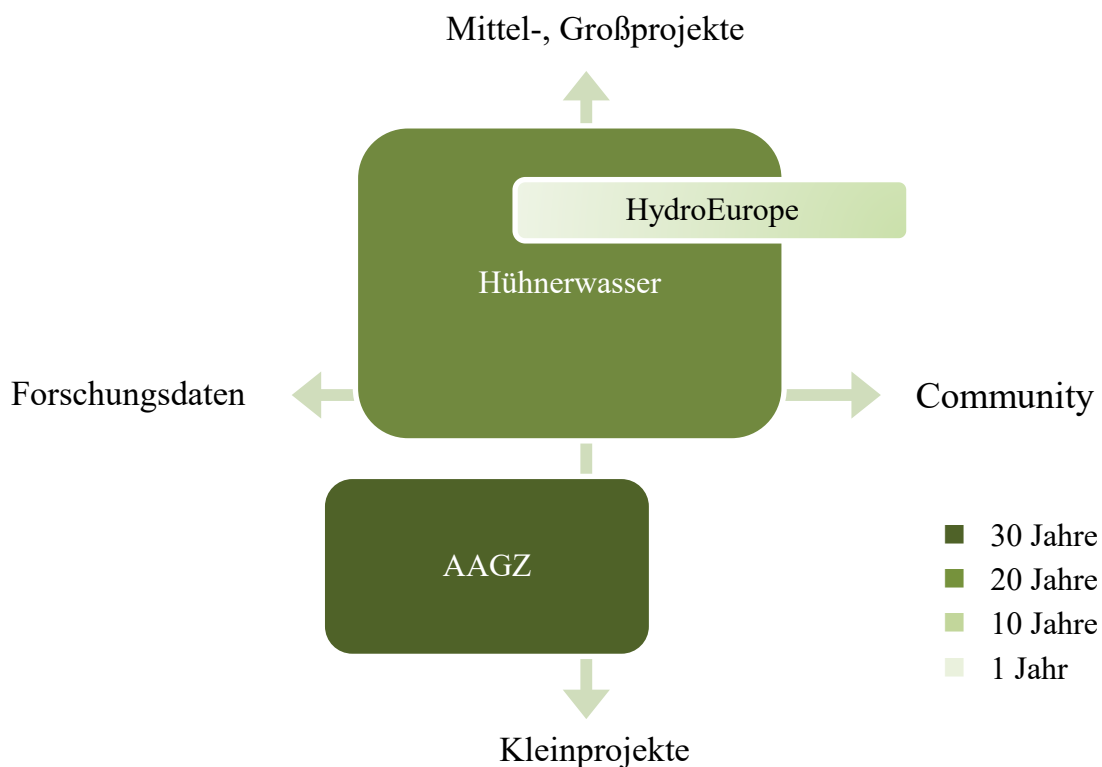


Abbildung 5.1: Projektspezifische Eigenschaften (Größe, Schwerpunkt, Zeitraum) der Anwendungsbeispiele von L-TER Projekten

Die Größenordnung der Projekte ist abhängig von: der Benutzeranzahl, Anzahl der Untersuchungsobjekte, dem Datenvolumen, der Anzahl an Teilprojekten, geografische Ausdehnung des Projektes und der Teilnehmer sowie der eingesetzten ICT-Infrastruktur. Verschiedene Strukturen ergeben sich bei: Datenstrukturen, den teilnehmenden Institutionen und dem interoperablen Datenaustausch in Netzwerken. Neben den rein „fachlichen“ Schwerpunkten, werden bei der Planung der L-TER VREs zwei VRE-spezifische Schwerpunkte gesetzt, die sich wahlweise auf die Forschungsdaten im

Sinne eines eher „technisch“ orientierten Systems und/oder „Communities für Forschungsgemeinschaften“ beziehen. Die Projektzeiträume sind durch die allgemeinen Rahmenbedingungen vorgegeben.

Die jeweilige projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes wurde anhand der Kriterien Projektgröße, -Schwerpunkt und -Zeitraum überprüft. Daraus ließ sich bewerten, ob das Konzept prinzipiell für die Entwicklung einer „flexiblen“ L-TER VRE geeignet ist.

L-TER VREs, deren Prototypen und Weiterentwicklungen sollen sich schnell realisieren lassen (vgl. Kapitel 1). Um das „Rapid L-TER VRE“ Baukastenkonzept hinsichtlich einer „schnellen“ Realisierung zu überprüfen, wurden die Entwicklungszeit für Basis-VREs (Prototypen) als Kriterium verwendet. Die zeitlichen Einflüsse in den anderen Lebensphasen einer L-TER VRE wurden hier nicht berücksichtigt, da im Rahmen dieser Arbeit L-TER VREs im Produktivbetrieb über einen spezifisch langen Zeitraum hinweg weitgehend nicht praktisch untersucht werden konnten. Der Einfluss der Zeit auf die einzelnen Lebensphasen einer L-TER VRE wird jedoch im Rahmen der Gesamtbewertung (vgl. Kapitel 6.1.4) theoretisch bewertet.

Die folgenden L-TER Projekte wurden als Anwendungsbeispiele gewählt:

Das Projekt „Überwachung des Grundwasserpfades der Altablagerung Großziehten“ (AAGZ) steht stellvertretend für „kleinere“ L-TER VREs mit relativ „wenigen“ Benutzern und beteiligten Institutionen sowie einem (einzigem) Hauptprojekt mit eher strukturierten Forschungsdaten im Umfang von „Small Data“. Der Schwerpunkt liegt hier eher auf dem Forschungsdatenmanagement, als auf einer Community.

Das Projekt „Hühnerwasser“ erfordert eine relativ „große“ L-TER VRE mit interdisziplinären Teilprojekten. Kooperationen mit vielen beteiligten Instituten bedingen eine relativ „hohe“ und „wechselnde“ Anzahl an Benutzern. Besonderer Fokus liegt auf dem Datenmanagement, der Datenauswertung, sowie einer digitalen Forschungsgemeinschaft, gestützt durch moderne Web 2.0 Kommunikationstools. Die Forschungsdaten haben vorwiegend unterschiedliche Datenstrukturen und den Umfang von „Medium Data“.

Stellvertretend für eine „artverwandte“ L-TER VRE steht das Projekt „HydroEurope“ aus dem Bereich der „kollaborativen“ und internationalen Lehre im Wasserwesen und der Hydroinformatik. Hier ist der Langzeitbezug nicht bei den einzelnen Projekten, sondern in der Kontinuität der Lehre zu sehen. Das Projekt umfasst eine relativ konstante Nutzeranzahl von ca. 100 wechselnden Benutzern pro Jahr.

5.1 L-TER VRE für die Nachsorge der „Altablagerung Großziethen“

Die Nachsorgephase einer ehemaligen Mülldeponie (Altablagerung) ist gesetzlich für 30 Jahre laut Deponieverordnung (DepV) vorgeschrieben³. In dieser Zeit erfolgen Renaturierung, ökologische Langzeitstudien und Kontrollen des Grundwassers auf chemische Verunreinigungen durch die Altablagerung, um die Gesundheit und Sicherheit der angrenzenden Bevölkerung zu gewährleisten. Für die Nachsorge der Altablagerung in Großziethen (AAGZ) in Brandenburg südlich der Berliner Stadtgrenze (Abbildung 5.2) ist das Umweltamt des Landkreises „Dahme-Spreewald“ (Brandenburg) zuständig. Ihm obliegt seit 1997 das Monitoring an derzeit ca. 20 Messpunkten in verschiedenen Lagen und Sektoren der Altablagerung (Abbildung 5.3). Jährlich werden zwei Messkampagnen durch externe Labore im Frühjahr und Herbst durchgeführt, bei der das Grundwasser auf einschlägige anorganische und organische Substanzen, sowie einige physikalische/chemische Parameter untersucht wird – insgesamt ca. 35 Parameter.



Abbildung 5.2: Luftaufnahme der Altablagerung Großziethen (LDS Umweltamt 2005)

5.1.1 Aktueller Stand des Projektes

Für die Analyse der Schadstoffausbreitung im Grundwasser ist die „BTU Cottbus-Senftenberg, Lehrstuhl Environmental Informatics“ beauftragt. Das Monitoring mittels Sensoren in Echtzeit ist aufgrund der relativ langsamen Schadstoffausbreitung und der

³ Deponieverordnung vom 27. April 2009 (BGBl. I S. 900), die durch Artikel 2 Absatz 23 des Gesetzes vom 20. Juli 2017 (BGBl. I S. 2808) geändert worden ist

Gesetzesgrundlagen nicht erforderlich. Die geostatistischen Auswertungen zur Schadstoffausbreitung wird aus den Analysedaten der Labore manuell mit ESRI ArcGIS vorgenommen und als Bericht sowie in Form von gedruckten Karten und Animationen der Konzentrationsverläufe an den Auftraggeber übergeben. Die Datenspeicherung der Messdaten erfolgt mittels Excel-Tabellen für die einzelnen Parameter und Messstellen. Derzeit existieren keinerlei Komponenten für das digitale Datenmanagement oder web-basierte Erweiterungen. Die Kommunikation der Projektpartner erfolgt ganz klassisch über E-Mail und per Telefon oder bei Projekttreffen an einem Ort.

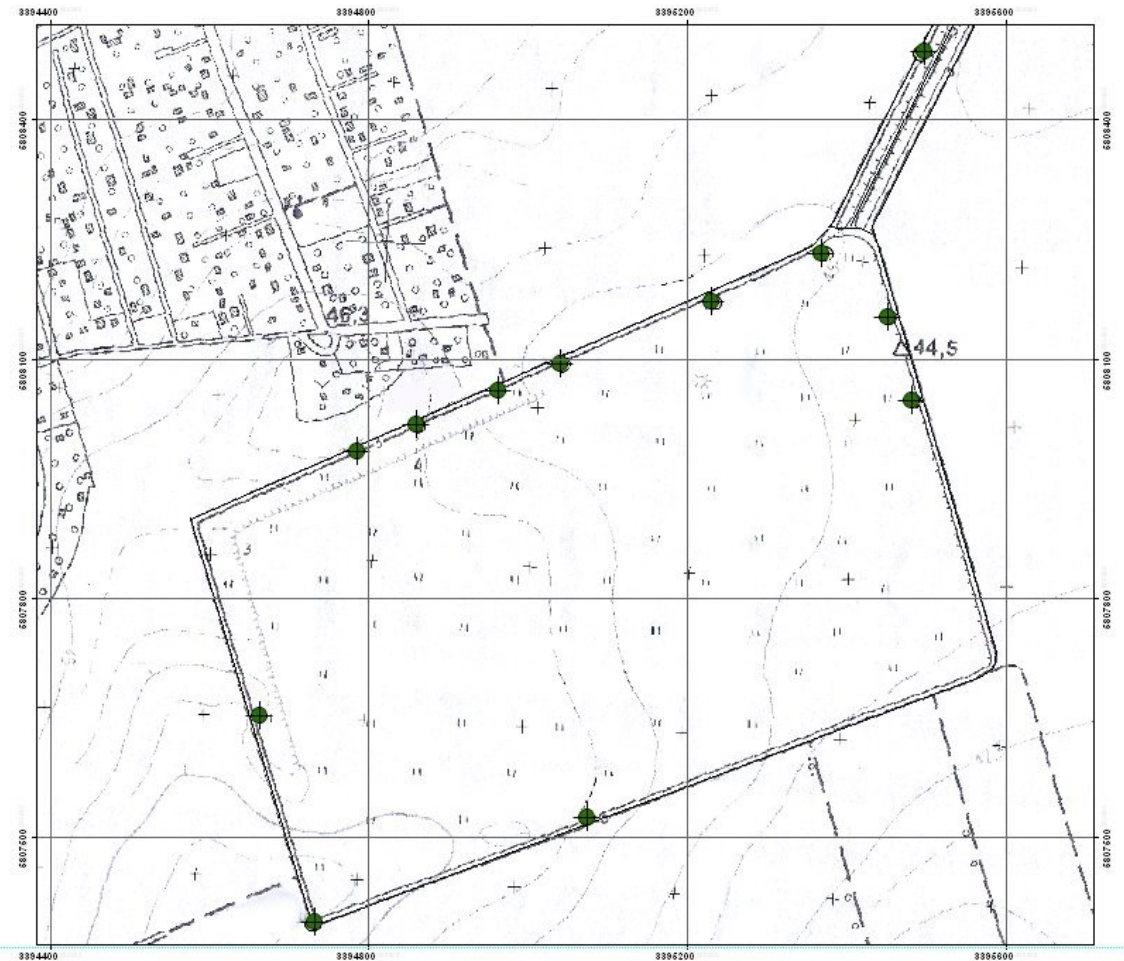


Abbildung 5.3: Altablagerung Großziethen - Messpunkten des Oberpegels nahe angrenzender Siedlung

Es sammeln sich zunehmend Messdaten, digitalen Karten, Bilder, Videoanimationen der Konzentrationsverläufe, Gutachten, Grenzwerte und Hintergrunddaten über die Zeit der Nachsorgepflicht an. Das Datenvolumen für das Projekt insgesamt beträgt ca. 5 GB pro Jahr in den letzten 20 Jahren (1997 bis 2017). Diese inhomogenen und strukturierten Daten stammen von verschiedenen Projektpartnern wie z.B. Ämtern, Laboren, Baugesellschaften, externen Experten, der Universität und anderen Institutionen. Sie müssen aufbereitet und zusammengefügt werden, damit ein Gesamtbild des Gefahrenpotentials der Altablagerung entsteht und die Öffentlichkeit adäquat über den Umweltzustand informiert werden kann. Das öffentliche Interesse ist bei der Altablagerung in Großziethen besonders hoch, da der minimale Abstand zur umliegenden Bebauung ca.

30-50 m beträgt (Abbildung 5.2) und der Ort im Grundwassereinzugsgebiet der ca. 3 km entfernten Berliner Stadtgrenze liegt. Mit einem Gefahrenpotential der Altablagerung ist auch für zukünftige Anwohner im Einzugsgebiet der Altablagerung zu rechnen. Tabelle 5.1 zeigt die Rahmenbedingungen des Projektes in der Übersicht.

Tabelle 5.1: Rahmenbedingungen des Projektes „Altablagerung Großziethen“ für eine L-TER VRE

Klasse	Beschreibung
Benutzerkreis	relativ kleiner institutionen-übergreifender und statischer Nutzerkreis von ca. 10 Personen
Zielgruppen	Projektgruppe, Verwaltung und Öffentlichkeit
Institutionen	Landkreis Dahme Spreewald (Auftraggeber), COSA GmbH (Projektplanung / -überwachung), Hafemeister Erd- und Tiefbau GmbH (Baudienstleistungen), BTU Cottbus-Senftenberg (wissenschaftliche Begleitung und praktische Lehre)
Projektanzahl	ein Hauptprojekt
Daten <ul style="list-style-type: none"> ▪ Datenmanagement ▪ Datenvolumen ▪ Datenstrukturen ▪ Datenaufbereitung ▪ Datenauswertung 	Daten aus Laborauswertungen von ca. 20 Messstellen, 35 Parametern, zwei Messungen im Jahr. Pro Parameter vier Karten für Oberpegel, Unterpegel und Saison im Jahr, Jahresbericht, GIS-Projektdateien mit Vektor- und Rasterdaten, Sonstige (z.B. Gesetzestexte und Grenzwerttabellen) kleine Datenmengen in unstrukturierter und strukturierter Form, insgesamt ca. 5 GB pro Jahr. Die Datenaufbereitung findet derzeit manuell statt. Die Datenauswertung erfolgt derzeit mit „ESRI Arc GIS Desktop“
Aufgaben	Monitoring, Visualisierung und Bewertung von Schadstoffen und Parametern im Grundwasserpfad unter einer Altablagerung mit Berichten für die Öffentlichkeit
VRE-Schwerpunkt	Forschungsdaten
Zeitlicher Bezug	30 Jahre im Rahmen der Nachsorge-Verpflichtung 1997-2037. Ggf. generationenübergreifend, da Gefahrenpotential für die Nachwelt.
Geographischer Bezug	Großziethen, Brandenburg

5.1.2 Zielsetzung der L-TER VRE

Unter den gegebenen Rahmenbedingungen soll eine L-TER VRE für die Nachsorgephase der Altablagerung „Großziethen“ aufgebaut werden. Die bisherigen Daten, Forschungsprozesse und Ergebnisse werden dazu zunächst in einem Basis-VRE Prototyp mit Basisfunktionalitäten zusammengefasst. Nach erfolgreicher Umsetzung kann in einem zweiten Schritt die VRE zur zentralen Umwelt-Plattform der AAGZ mit erweiterter Funktionalität für alle Beteiligten im Projekt und zur Information der Öffentlichkeit ausgebaut werden. Die Automatisierung von Workflows und Berichten für verschiedene Zielgruppen steht dabei im Vordergrund. Die gewonnen Daten und Metadaten sollen optional im Datenverbund mit Hilfe eines Harvesting-Interfaces automatisch an regionale oder (inter)nationale übergeordnete Einheiten wie z.B. das Umweltbundes-

amt und dessen Umweltdatenbanken übermittelt werden können, wie z.B. die GDI (geoportal.de).

Basis L-TER VRE Funktionalitäten für die „AAGZ“:

- Website mit allgemeinen CMS-Funktionalitäten,
- Gruppen und abstimmbare Zugriffsrechte,
- zentrale SQL-Datenbank mit GIS Erweiterung für die Speicherung der Monitoring-Daten,
- vordefinierte und individuelle Abfrage der Datenbank,
- (Meta)Daten-Management / -Harvesting,
- Management zusätzlicher Hintergrundinformationen (Bilder, Presseartikel, Koordinaten),
- Wiki für Hilfe, Dokumentation und Teamarbeit,
- Suche nach Kategorie, Zeit, Ort, Schlüsselwort, Volltext, Medientyp,
- Datenschema für Monitoring-Daten und Ergebniskarten,
- Übersicht und Download der Daten der Karten und jährlichen Reports,
- zentrale Benutzerverwaltung, Authentifizierung und SSO,
- Animationen der zeitlichen Verläufe einzelner gemessener Parameter,
- ausgewählte Informationen für die Öffentlichkeit,
- allgemeine Informationen für die Öffentlichkeit zum Projekt, Betreiber, beteiligte Institutionen,
- Persistente Identifikatoren (PI) für Datensätze,
- Management der GIS-Projektdateien und speziellen Metadaten,
- einfaches WebGIS.

Erweiterte L-TER VRE Funktionalitäten für die „AAGZ“:

- Automatisierte räumliche Interpolation der Konzentrationen
- Social-Media-Community,
- Report Layer zur automatischen Erstellung von Jahresberichten für verschiedene Zielgruppen,
- Automatisierung des Forschungsprozesses und der Workflows allgemein,
- Daten Pre- und Postprocessing (z.B. Plausibilitätsprüfung, Unterscheidung von Messwerten mit „0“ Wert und nicht vorhandenen Messungen),
- Zeitreihenanalyse,
- Prognose der Konzentrationen mit Szenarien,
- Erweiterung auf weitere Forschungsbereiche der „AAGZ“ (z.B. Nachnutzung und Vegetation der Abdeckung),
- generationenübergreifende Langzeitdatenspeicherung,
- Mehr Informationen für die Öffentlichkeit,
- KI Modul für die Gasprognose (Köwitsch 2016),

- digitale Verbindung aller relevanter Sensoren und Daten mit der VRE,
- Publikationsverzeichnis,
- Öffentliche Umfragen zur Nachnutzung,
- Suchmaschinenoptimierung.

Die Hauptdatenbasis für das Projekt bilden die weitgehend strukturierten Messwertreihen des Grundwasserkörpers in Form von Excel-Tabellen, sowie die daraus manuell erzeugten kartografischen Visualisierungen (JPG Bildformat, PDF-A Langzeitarchivierungsformat) (Abbildung 5.4) und Animationen (WMV-Videoformat) der Konzentrationsverläufe ausgewählter Substanzen. Hinzu kommen inhomogene Daten unterschiedlicher Herkunft und Formate, wie z.B. Grenzwerttabellen, Berichte, Gutachten, Protokolle.

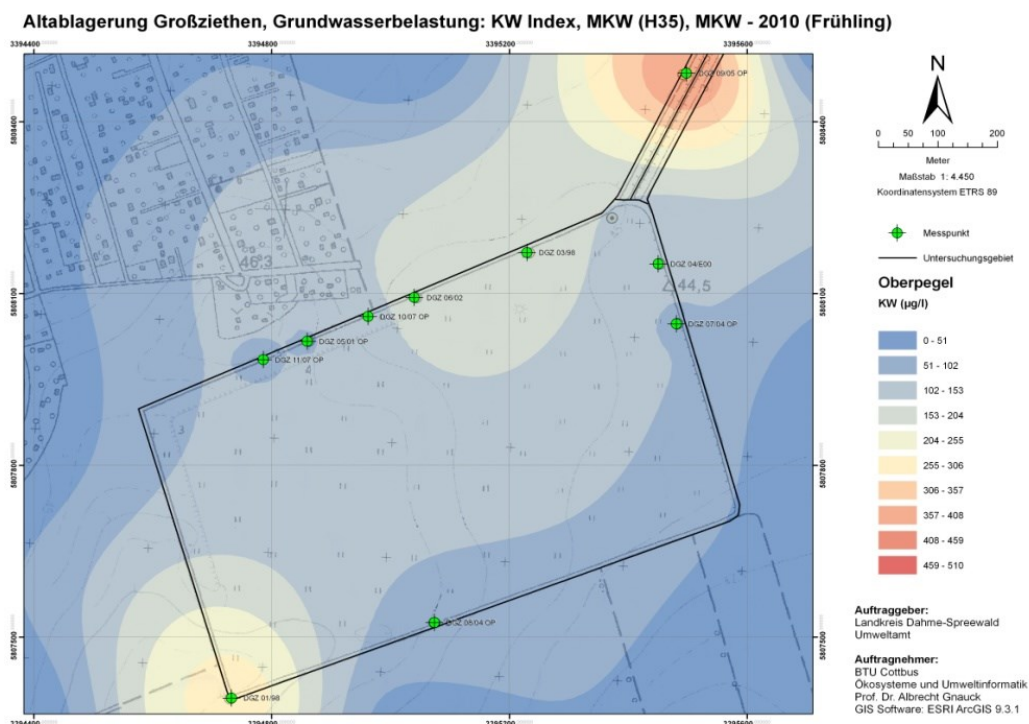


Abbildung 5.4: Konzentrationsverlauf für Kohlenwasserstoffe des Unterpegels der Altblagerung Großziethen im Frühjahr 2010

Tabelle 5.2 zeigt die unterschiedlichen Datentypen für das Projekt mit einer kurzen Beschreibung.

Tabelle 5.2: Datentypen für das Projekt „AAGZ“

Datentyp	Beschreibung
Foto, Video	Foto- und Videodateien: Projektbilder (JPG, BMP, etc.), Zeitraffer-Animation der Konzentrationsverläufe im WMV-Format
Gesetze, Richtlinien, Grenzwerttabellen	Unterschiedliche Textformate: Umweltgesetze, Richtlinien, Grenzwerttabellen, internationale Standards
Interaktive Ressourcen	Interaktive Karten (WMS), Google Maps Layer
Karten	PDF / PDF-A, JPG Bilddateien, ESRI ArcGIS Formate
Messwertreihen	Tabellen- / Datenbankdateien: Messwerte der einzelnen Untersuchungs-

Datentyp	Beschreibung
	kampagnen im MS-Excel-Format
Software	Spezifische Software, wie z.B. Simulations- und Modellierungsmodelle, Mobile Applikationen (Apps)
Studien, „best practice“ Lösungen	Textdokumente: Berichte, Studien, Gutachten, Veröffentlichungen in herkömmlichen Dokumentformaten wie MS-Office oder PDF / PDF-A

5.1.3 Konzept und Prototyp

Mit Hilfe des „Rapid L-TER VRE“ Baukastenkonzeptes wird ein Entwurf für die schnelle Realisierung der L-TER VRE erstellt und in der Praxis in Form eines Prototypen umgesetzt. Die Vorgehensweise für die Implementierung der Prototypen erfolgt nach dem Life-Cycle-Modell für VREs, wobei die einzelnen Module der L-TER VRE in eigenen Zyklen iterativ durchlaufen werden.

Die Kurzform des Entwurfes stellt Tabelle B.1 im Anhang B als *komprimiertes* „Rapid L-TER VRE“ Konzept dar. Sie zeigt die einzelnen Bestandteile des Konzeptes aus Kapitel 4 mit den jeweiligen Parametern, die folgend nach den 10 Punkten des Konzeptes beschrieben werden:

Die L-TER VRE wird auf der Basis der Open-Source-Software (SdT) aus Anhang B, C und D implementiert. Diese baut auf den Web- und ICT-Basistechnologien sowie internationale Standards auf, die für das vorliegende Konzept der L-TER VRE nach den vorangegangenen Betrachtungen geeignet erscheinen. Dazu gehören ein CMS als Basis-System, Datenbanken für CMS und Forschungsdaten, Authentifizierung und SSO, sowie Datenverbünde.

Für eine schnelle Installation und Entwicklung der L-TER VRE werden „schlüssel-fertige“ Server-Komponenten von www.turnkeylinux.org verwendet (vgl. Anhang F), die mit der Software „Virtual Box“ von „Oracle“ auf der Basis von virtuellen Maschinen betrieben werden. Dies vereinfacht den administrativen Aufwand, reduziert benötigte Hardwareressourcen und ermöglicht Sicherungspunkte („SnapShots“) in Entwicklungs- und Betriebsphase.

Als Einstieg in die L-TER VRE (Internet und Intranet) wird zuerst eine zentrale Homepage mit der Basis-Software „Joomla“ als CMS aufgebaut. Neben den allgemeinen Informationen zum Projekt (für die Öffentlichkeit im Internet) findet sich auf der Homepage der Zugang zum geschützten Bereich (Intranet) für die Projektteilnehmer („Frontend“) und die Administratoren („Backend“) der L-TER VRE. Durch die beschriebenen VRE-Tools lässt sich das CMS gemäß den Anforderungen in Tabelle 5.1 ergänzen. So können z.B. für die erweiterte L-TER VRE Community-Funktionen über die umfangreichen „Joomla“ CMS-Erweiterungen „Community Builder“ (Open Source) oder „JoomSoical“ (kostenpflichtige Lizenz) eingesetzt werden.

Der „Baukasten“ für VRE-Services (vgl. Tabelle 4.2) bildet die Basis zur Auswahl von geeigneten VRE-Services und der eingesetzten Software (vgl. Anhang C). Die

Einbindung der Services erfolgt im CMS durch Links auf die Homepage des entsprechenden VRE-Services oder die Einbettung mittels „iFrame“ (HTML-Technologie). Folgende VRE-Services (Tabelle 5.3) sollen nach den Anforderungen (Tabelle 5.1) zum Einsatz kommen. Die erweiterten VRE-Services sind durch „*“ gekennzeichnet:

Tabelle 5.3: Basis- und erweiterte VRE-Services für die „AAGZ-VRE“

Kategorie	Software („Schlüsselfertig“: T)
CMS	„Joomla“(T)
Metadatenmanagement und Datenverbünde	„GeoNetwork“ (T)
Wiki	„MediaWiki“ (T)
Web-Server / Betriebssystem	„nginx“oder „Apache“ Webserver unter „Linux“ OS (T)
Datenbankserver	„Linux“ OS mit: „mySQL“ für VRE-Services, „PostgreSQL“ für Messdaten (T)
Authentifizierung	„OpenLDAP“ (T)
Community*	„Joomla“ mit Erweiterung „JomSocial“ (kostenpflichtige Lizenz), (T)
Statistik*	„R-Server“

Die Einbindung von VRE-Tools erfolgt durch die Installation servicespezifischer Erweiterungen oder durch Programmierung, sofern keine geeigneten Erweiterungen gefunden werden. Das CMS und das Wiki sind besonders flexible Services, wodurch individuelle Erweiterungen mit VRE-Tools möglich werden. Für diese beiden VRE-Services kommen daher (beispielhaft für andere VRE-Services und -Tools) die folgenden zusätzlich gewählten Erweiterungen zum Einsatz:

Tabelle 5.4: Verwendete Erweiterungen (VRE-Tools) zum „Joomla“ CMS für die „AAGZ-VRE“

Name	Beschreibung
AdCustomCSS	Layout und Design
ArtData	Visualisierung und Datenaufbereitung
FAQ Book	Hilfesystem
iSEO (kostenpflichtige Lizenz)	Suchmaschinenoptimierung
JomSocial (kostenpflichtige Lizenz)	Community, Gruppen, Benutzerhomepage, Terminkalender, Social Media, Web 2.0
K2	Artikel- und Kategorienmanagement
NoNumber ReReplacer	Text, Übersetzungen
OS PDF Indexer	PDF Verschlagwortung
OSE Anti Hacker (kostenpflichtige Lizenz)	Website-Sicherheit
Shmanic LDAP	Zentrale Authentifizierung
XCloner	Backup

Zur Auswahl dient die Übersicht zu Software für „Joomla“ CMS-Erweiterungen im Anhang D und für „MediaWiki“ Erweiterungen in Tabelle C.8 im Anhang C. In Tabelle 5.4 sind die verwendeten VRE-Tools des „Joomla“ CMS für die L-TER VRE in der Übersicht dargestellt. Tabelle 5.5 zeigt die ausgewählten Erweiterungen für „MediaWiki“.

Tabelle 5.5: Verwendete Erweiterungen (VRE-Tools) zu „MediaWiki“ für die „AAGZ-VRE“

Name	Beschreibung
Visual Editor	WYSIWYG Editor mit zahlreichen Optionen
Google Charts	Erzeugung von Diagrammen
R / Shiny	Einbindung von „R“ Scripten

Da nur wenige Personen anfänglich einen internen Zugang erhalten sollen, muss keine aufwendige und institutionenübergreifende Authentifizierung durchgeführt werden. Trotzdem ist eine benutzerfreundliche Anmeldung mit SSO mit Hilfe von LDAP Verzeichnisserver auch im Prototyp für den Test und für mehr als einen Service mit jeweils eigener Authentifizierung nützlich. Zu Testzwecken der übrigen L-TER VRE kann jedoch auf diese Funktionalitäten verzichtet.

Berichte sollen im Report-Layer für die Zielgruppen (Benutzer, Administratoren, Öffentlichkeit) durch die schrittweise Verbesserung von geeigneten Tools (teil-) automatisiert erzeugt werden können. Insbesondere ist die unterstützende „Generierung“ von jährlichen Abschlussberichten interessant, die auch der Öffentlichkeit zugänglich gemacht werden. Dazu sind VRE-Tools zu installieren und zu programmieren, die Kennzahlen ermitteln und mit Textbausteinen nach einer spezifischen Logik verknüpfen. Der maschinelle Datenverbund zu weiteren Institutionen kann bei Bedarf optional durch das OAI Protokoll für Metadaten über die bereits implementierte Funktion des Data-Harvesting von „GeoNetwork“ erfolgen. Dies wurde im Rahmen der Arbeit nur teilweise mit einem Grunddatenbestand getestet.

Hilfe, Feedback und Dokumentation zur L-TER VRE wird durch die Erweiterungen (VRE-Tools) des CMS (z.B. FAQ, Umfragen, Bewertungen), sowie durch das gruppenorientierte Wiki („MediaWiki“) realisiert.

Für die Weiterentwicklung der L-TER VRE sind die Funktionalitäten aus der Liste „Erweiterte L-TER VRE Funktionalitäten für die AAGZ“ geplant. Insbesondere die folgenden Bereiche sind für die Weiterentwicklung der L-TER VRE relevant: Langzeitdatenspeicherung, Visualisierungstechnologien, Datamining, Monitoring, Sensornetzwerke, Modellierung, Simulation, Prognose, KI Methoden. Das erweiterte Konzept für die L-TER VRE wurde nicht als Prototyp realisiert.

Die GIS Karten als Ergebnis des Forschungsprozesses sind von besonderer Bedeutung in der VRE. Für sie soll exemplarisch ein Metadatenmodell erstellt werden, um sie im Datenmanagement der VRE zu integrieren. Für das Datenmanagement wird ebenfalls „GeoNetwork“ eingesetzt, das für unterschiedliche Datentypen verschiedene Metadatenbeschreibungen zulässt. Diese können mit bereits integrierten Vorlagen auf internationalen Standards basieren, oder individuell beschrieben werden. Tabelle 5.6 zeigt das erstellte Metadatenschema auf der Basis von ISO 19115:2003/19139 für die Karten im PDF-A Datenformat.

Tabelle 5.6: Metadatenchema für die Hauptdatentypen auf Basis der ISO 19115:2003/19139 Norm für die „AAGZ“

Rubrik und Metadatum	Beispiel
Datenbeschreibung	
Titel	DGZ Blei 2011 (Frühling) Unterpegel
Datum	01.01.2011
Kurzbeschreibung	Karten zur Schadstoffausbreitung...
Kontaktinformation	Adresse, Website, Kontaktperson
Datenpflege	keine Aktualisierung
Datenqualität	Unsicherheiten, Rahmenbedingungen,
Schlüsselwörter	Altablagerung, Grundwasser
Nutzungsbedingungen	Keine kommerzielle Nutzung
Sprache	Deutsch
Zeichensatz	UTF-8
Maßstab	1:5,000
Koordinaten und Referenzsystem	50.3556, 13.4656, ETRS 89 UTM Z33N
Zeit / Zeitperiode	1990-2010
Metadatenbeschreibung	
Persistenter Dateiname	914ae19e-c3d1-44ba-97fd-1c1a20bdd
Sprache	Deutsch
Zeichensatz	UTF-8
Kontaktinformation	Adresse, Website, Kontaktperson

Die Suche nach den Daten in der L-TER VRE kann auf verschiedenen Wegen erfolgen. Zum einen steht die Suche in Kategorien (z.B. Bilder, Karten, Zeitreihen) aus dem Menü der Homepage von „GeoNetwork“ heraus zur Verfügung. Zum anderen die Variante über einen einfachen oder erweiterten Suchdialog, in den Keywords, Suchbegriffe, sowie geografische und zeitliche Filter eingegeben werden können. Die Ergebnisliste kann nach mehreren unterschiedlichen Kriterien sortiert werden.

In der Rubrik „Was?“ (vgl. Abbildung 5.5) sind Suchwörter einzugeben und gegebenenfalls mit einem logischen Operator zu verknüpfen. Die Suche bezieht sich dabei auf den Titel, den Abstract oder die Keywords. Auch eine Volltextsuche ist durchführbar. Jedoch kann „GeoNetwork“ derzeit nicht in den Dokumenten (z.B. PDF) an sich suchen – dazu sind andere VRE-Tools einsetzbar. Über weitere Einstellungen lässt sich auch die Suchgenauigkeit konfigurieren oder direkt nach einem speziellen Kartentyp suchen (digitale Karten, analoge Karten, interaktive Karten, herunterladbare Karten). Mit dem Suchfilter „Wo?“ erlaubt die Suche geografische Eingrenzungen. Über eine Auswahlbox wird die betreffende Region ausgewählt. Eine lokale Suche kann auch mittels einer vom Map-Server bereitgestellten vordefinierten, eigenen oder durch den Benutzer generierten Karte mit der Maus oder über Koordinateneingabe gestartet werden. Logische Operatoren sind hier ebenso möglich, wie bei den Suchbegriffen. Die Option „Wann?“ bietet Selektionsmöglichkeiten für einen bestimmten Zeitpunkt oder eine Zeitspanne, in der die Daten erzeugt oder publiziert wurden. In den allgemeinen Suchoptionen wird eingestellt, in welchem Katalog und in welcher Kategorie gesucht werden soll und wie die Ergebnisliste ausgegeben wird. Dazu lassen sich folgende

Sortiermöglichkeiten einstellen: Sortierung alphabetisch nach Titel, nach Änderungsdatum, Bewertung, Popularität und nach Relevanz. Die Anzahl der Ergebnisse pro Seite und die Darstellungsform (Kurzform oder erweiterte Ansicht) lassen sich ebenfalls konfigurieren.

Die Übersichtsliste der Suchergebnisse zeigt zum jeweiligen Datensatz eine kurze Beschreibung, weitere Schlüsselwörter, ein Vorschaubild des Datenmaterials (soweit vorhanden) sowie ein Logo des Datenanbieters an (Abbildung 5.5). Mit einem Klick auf den Button Metadaten sind die vollständigen Informationen zum Datensatz abrufbar. Je nach Zugriffsrechten des Benutzers können nur die Metadaten eingesehen werden oder die Ressourcen vom lokalen „GeoNetwork“-Server bzw. von verlinkten Quellen heruntergeladen werden. Administratoren und Autoren ist es in dieser Darstellung möglich, auch über das Frontend (Benutzeransicht) die Daten zu pflegen.

Optional kann „GeoNetwork“ ebenso für die Umsetzung der „Infrastructure for spatial information in Europe (INSPIRE)“ Richtlinie (INSPIRE 2007) genutzt werden, da diese Funktionalitäten bereits integriert sind. Während der Dateneingabe und der Metadatenbeschreibung schlägt ein (semantic web, RDF-konformes) INSPIRE Thesaurus Begriffe für die Verschlagwortung vor (Abbildung 5.6).

The screenshot displays the GeoNetwork web interface. At the top, the header includes the site name 'Altablagerung Großziethen virtuelle Forschungsumgebung' and logos for 'Dahme-Spreewald' and 'b-tu Brandenburgische Technische Universität Cottbus'. A navigation bar contains links for 'Home', 'Administration', 'Kontakt', 'Links', 'Über', and 'Hilfe', along with a language dropdown set to 'Deutsch' and a user login 'Benutzer: admin admin' with an 'Abmelden' button.

The main search area features a search box with 'blei 2010' entered. Below it, a 'WO?' section shows a world map with a highlighted region in Europe. A search button labeled 'Suche' is visible, along with options for 'Zurücksetzen' and 'Erweitert'. A sidebar on the left lists various resource categories like 'Andere Ressourcen', 'Applikationen', 'Audio/Video', etc.

The search results section, titled 'FINDEN SIE INTERAKTIVE KARTEN, GIS DATENSÄTZE UND WEITERE RÄUMLICHE DATEN', shows one result: 'DGZ BLEI 2010 FRÜHLING UP'. It includes a 'Logo' button, a 'Kurzübersicht' (summary) of the data, and a small map thumbnail. The summary text reads: 'Karten zur Grundwasserbelastung relevanter Stoffe der Altablagerung (Deponie) Großziethen (Bei Berlin) Deutschland'. Below this, 'Schlüsselwörter' (keywords) are listed: 'Altablagerung, Deponie, Großziethen, Berlin, Deutschland, Umweltüberwachung, Boden, Grundwasser, Umwelt, Boden'. At the bottom of the result, there are buttons for 'Metadaten', 'Erstellen', 'Bearbeiten', 'Löschen', and 'Other actions', with the user 'Besitzer: admin' indicated.

Abbildung 5.5: „GeoNetwork“ Suchergebnis für „Blei 2010“ mit Vorschau

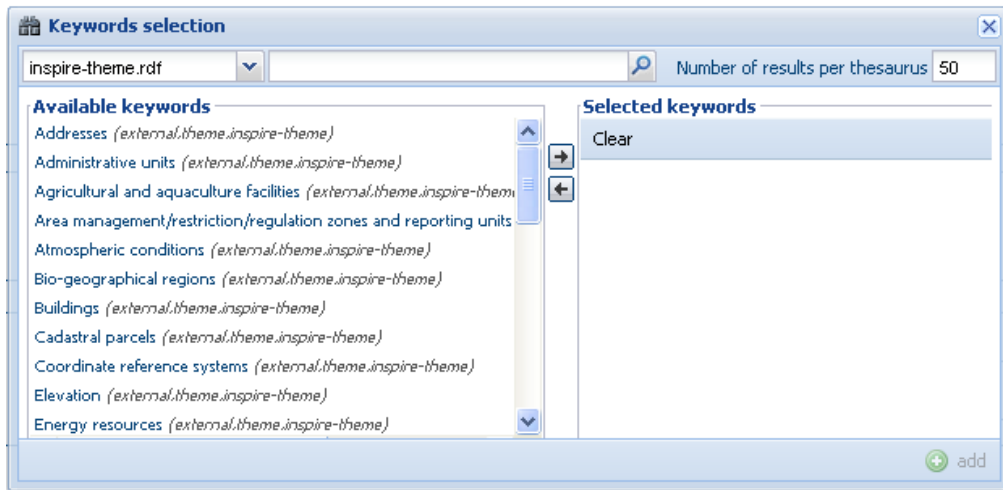


Abbildung 5.6: „GeoNetwork“ Schlüsselwörter für Metadatenbeschreibung mit INSPIRE Thesaurus

Für die Suche nach Umweltinformationen auf Basis von Vorgaben der INSPIRE Richtlinie sind folgenden Kriterien vorgesehen: Anhang (I, II, III), Datentyp, Servicetyp oder Themengebiet (Abbildung 5.7). Auf der Basis von „Geonetwork“ sind dadurch leicht und schnell INSPIRE-konforme Datenmanagementsysteme für den Umweltbereich umzusetzen.

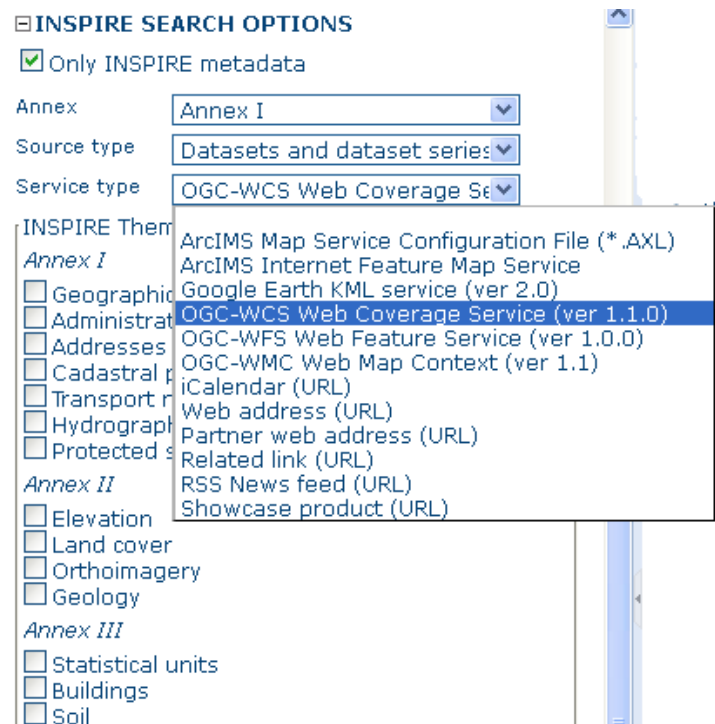


Abbildung 5.7: INSPIRE-konforme Suchoptionen: Anhang, Datentyp, Servicetyp, Themengebiet

5.1.4 Bewertung der Konzeptanwendbarkeit

Das „Rapid L-TER VRE“ Konzept bewährte sich für die gegebenen projektspezifischen Rahmenbedingungen einer relativ „kleinen“ L-TER VRE mit „wenigen“ Benutzern und beteiligten Institutionen sowie einem Hauptprojekt mit eher strukturierten Forschungs-

daten im Umfang von „Small Data“. Der Schwerpunkt im Bereich Forschungsdatenmanagement und die allgemeinen Aufgaben und Zielsetzungen des L-TER Projektes konnten realisiert werden.

Für die L-TER VRE der AAGZ fasst das komprimierte „Rapid L-TER VRE“ Konzept die wichtigsten Parameter für den Aufbau zusammen (vgl. Tabelle B.1 im Anhang B).

Die projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes wurde anhand der klassifizierten Kriterien (vgl. Kapitel 3.1) überprüft. Hinzu kamen die übrigen Kriterien zur Nachhaltigkeit (Persistenz, ICT, Mensch), schnelle Realisierung (Entwicklungszeit) und zur Virtualisierung (Größe, Struktur, Schwerpunkte, Dynamik) (vgl. Kapitel 3.1). Die Zusammenfassung stellt Tabelle 5.7 dar.

Somit konnte eine *nachhaltige* L-TER VRE, mit den gegebenen Rahmenbedingungen und Anforderungen, als Prototyp *schnell* und gut umgesetzt, sowie im Alpha-Stadium (ohne Benutzer) getestet werden.

Der Aufbau einer L-TER VRE für das Projekt über den Prototypen hinaus erscheint sinnvoll, da:

- bisher wenig ICT im Projekt vorhanden ist,
- es sich um eine kleine Benutzergruppe mit nicht ständig wechselnden Benutzern handelt,
- ein kleiner bis mittlerer Umfang an Komplexität der VRE zu erwarten ist,
- der Langzeitaspekt durch mindestens 30 Jahre Nachsorgeverpflichtung gegeben ist,
- das Projekt über Mittel verfügt, die für die Planung, den Aufbau und den Betrieb zum Teil genutzt werden können,
- eine Gefahrenabwehr für die Bevölkerung von besonderer Wichtigkeit ist,
- die Automatisierung nach Amortisationsphase zur Zeitersparnis führt,
- eine Community den Forschungsprozess belebt,
- durch die VRE-Basis es einfacher wird, neue Funktionalitäten nach dynamischen Anforderungen zu generieren.

Tabelle 5.7: Projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes für „AAGZ“

Klasse	Kriterium	Bemerkungen	Konzeptanwendbarkeit
Größe	Nutzerzahl	Vereinfachte Authentifizierung und SSO möglich, wenig dynamische Änderungen	✓
	Anzahl Untersuchungsobjekte	Forschungsdatenmanagement gut realisierbar mit „GeoNetwork“, „Small Data“	✓
	Datenvolumen	„Small Data“ (im Bereich von 5 GB / Jahr) lassen sich gut digital verarbeiten	✓
	Projektanzahl (Teilprojekte)	Komponenten sind durchweg sehr gut für ein Hauptprojekt geeignet	✓
	Umfang der ICT-Infrastruktur	Kleine Anzahl von VRE-Services und Tools vereinfacht die Komplexität des Gesamtsystems, weniger Systemperformance erforderlich	✓
	Geographische Ausdehnung (Untersuchungsgebiet, Benutzer)	PostgreSQL (Geo-)DBMS (Geo-)DBMS nützlich, Einhaltung regionaler Gesetze, keine Fremdsprachenübersetzung und Zeitunterschiede	✓
Struktur	Institutionen	Institutionenübergreifende Authentifizierung nur bedingt erforderlich	✓
	Datenstrukturen	Individuelle Strukturen erforderlich, „PostgreSQL“, Metadaten-Schema	✓
	Datenverbünde	Nicht unbedingt erforderlich, aber schnell realisierbar mit „GeoNetwork“ und nach INSPIRE, Harvesting mit zwei Servern getestet	✓
Schwerpunkte	Fachliche Schwerpunkte	Programmieraufwand erforderlich: Interpolation, Kartenerzeugung, Animation, Datenbanken	✓
	VRE-Schwerpunkte	Eignet sich für Schwerpunkt auf Forschungsdaten, individuelle Programmierung erforderlich – insbesondere bei automatischen Berichten, „GeoNetwork“ für Forschungsdatenmanagement	✓
Entwicklungszeit	Basis-VRE (Prototyp)	Schnelle Realisierung des Prototyps in wenigen Tagen	✓
Virtualisierung	Größe, Anzahl, Schwerpunkte, Dynamik	Größe: VM Dateigröße relativ „niedrig“ (ca.4 GB Basissystem) + 100 GB Datenspeicher Anzahl: Kleine Anzahl an VMs (ca. 4 Stk.) Schwerpunkte: VMs für Forschungsdatenmanagement vorhanden Dynamik: VMs benötigen wenig dynamische Eigenschaften (Load balancing, Hardwareperformance, Speicherplatz, Backup, Archivierung)	✓

5.2 L-TER VRE für BTU Sonderforschungsprojekt „Hühnerwasser“

Das Sonderforschungsprojekt „SFB/TRR 38 - Hühnerwasser“ (BTU Cottbus-Senftenberg 2017) untersuchte einen natürlichen Rekultivierungsbereich (Abbildung 5.9) im Braunkohletagebau der Niederlausitz (Abbildung 5.8). Es bietet vielfältige Möglichkeiten zur Erforschung der Initialphase der Ökosystementwicklung nach einem genau definierbaren "Punkt Null".

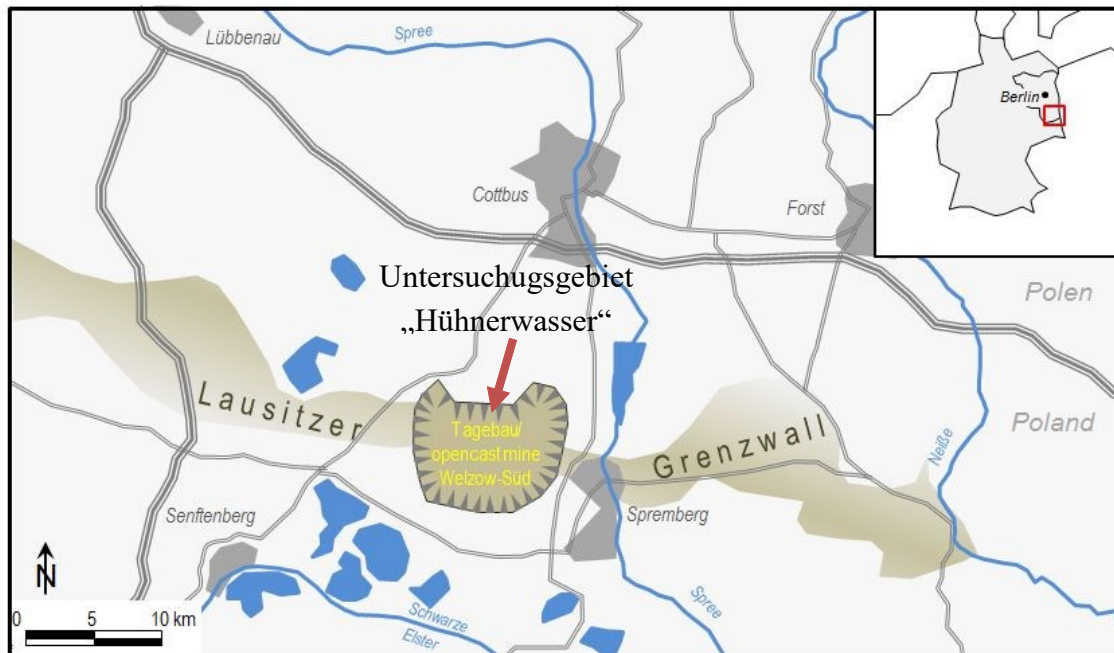


Abbildung 5.8: Lage des Untersuchungsgebiets „Hühnerwasser“ im Tagebau Welzow-Süd, bei Cottbus (BTU Cottbus-Senftenberg 2017)

Der Ablauf der Primärsukzession wird seit dem Startpunkt durch keine bewussten anthropogenen Lenkungsmaßnahmen beeinflusst. So kann die Selbstorganisation von Ökosystemen sowie die damit verbundene Struktur- und Musterbildung in terrestrischen und aquatischen Teilsystemen untersucht werden. Durch bauliche Maßnahmen auf der rund sechs Hektar großen Fläche entwickelte sich ein eigenständiger, lokaler Grundwasserkörper, der die Untersuchung von Feedbacks zwischen beispielsweise hydrologischen und geomorphologischen oder botanischen Strukturelementen ermöglicht. In einem entstandenen Teich können die Interaktionen zwischen terrestrischen und aquatischen Teilsystemen betrachtet werden.

Die entstandenen Bedingungen im künstlichen Wassereinzugsgebiet „Hühnerwasser“ sind weltweit einzigartig und bieten die Basis für vielfältige Forschungsprojekte.

Seit 2005 wird die Eigenentwicklung (Sukzession) des Hühnerwassereinzugsgebietes durch ein umfangreiches Umweltmonitoringprogramm und eine digitale Forschungsplattform in Form einer Website im Intranet und eingeschränkt im Internet begleitet. Die Daten stehen allen beteiligten Forschungsvorhaben der Hühnerwasserplattform zur Verfügung. Im Rahmen des Monitorings (Abbildung 5.10) werden meteorologische, hydrologische und bodenkundliche Umweltparameter sowie Depositionsda-

ten an vielen verschiedenen Punkten erfasst. Ergänzend zu den permanent installierten Einrichtungen finden weitere Einzelmessungen und -Beprobungen zur Vegetation, Bodenfauna und Limnologie im Feld und Labor statt. Luftbilder werden kontinuierlich im Frühjahr, Sommer und Herbst mit einer Drohne aufgenommen und ausgewertet.



Abbildung 5.9: Künstliches Wassereinzugsgebiet „Hühnerwasser“ – Primärsukzession
Luftbilder 2007, 2009, 2010 (BTU Cottbus-Senftenberg 2017)

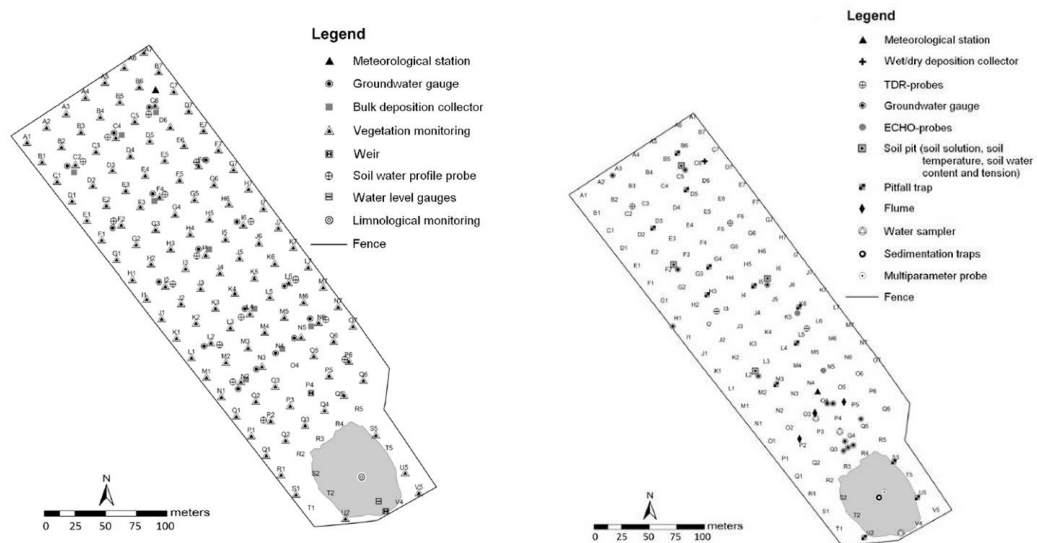


Abbildung 5.10: Permanente und zusätzliche Monitoreinrichtungen (BTU Cottbus-Senftenberg 2017)

5.2.1 Aktueller Stand des Projektes

Im Rahmen des Projektes sind in 12 Jahren folgende Datenvolumina mit inhomogenen Datenstrukturen angefallen: Insgesamt ca. 515 GB Datenvolumen, davon sind ca. 190 GB Bilder (vierteljährliche Luftbilder, tägliche Webcamaufnahmen sowie sonstige Fotos) und ca. 63 GB Daten aus dem laufenden Monitoringbetrieb (Messdaten von Dataloggern, von manuellen Messungen, jährlichen Vegetationserhebungen, aus dem Labor). Der Rest sind Rohdaten der Messwerte, Daten aus abgeschlossenen Projekten

sowie Studien- und Abschlussarbeiten. Im Test befindet sich derzeit ein kleines Sensornetzwerk, welches Daten per Mobilfunk an einen entfernten Server sendet.

Die Zusammenführung der Daten erfolgt über eine zentrale Speicherung der Zeitreihen in Form von MS Excel-Tabellen auf einem Netzwerkspeicher (NAS), wo auch die übrigen Daten in einer Ordnerstruktur gespeichert werden. Das Pre- und Postprocessing bei der Aufbereitung, sowie statistische Auswertungen der Daten, findet teils mit Scripten automatisiert, auf einzelnen PCs der Benutzer mit Statistikprogrammen (MS Excel, R, SPSS) statt. Ein Datenbank-Management-System für die Forschungsdaten befindet sich im Aufbau. Metadaten (gemessene Parameter, Methodik, Sensortyp, Dauer der Messungen, räumliche-, zeitliche Auflösung etc.) werden in einem Datenkatalog derzeit als PDF-Dokument zur Weitergabe an potentiell interessierte Datennutzer bereitgestellt. Weitere Metadaten nehmen meist die Form von bezeichnenden Datei- und Ordnernamen an oder werden automatisch durch die verwendeten Standardprogramme erzeugt, wie das EXIF-Format bei Kameraaufnahmen. Ein Metadaten-Management-System kommt nicht zum Einsatz. Derzeit wird an einer webbasierten statistischen Auswertung von verschiedenen Zeitreihen und Parametern gearbeitet, die Benutzer individuell auf die hinterlegten Forschungsdaten anwenden können. Dabei kommt ein „R-Statistic“-Server in Verbindung mit der Software „Shiny“ zum Einsatz, die für „R“ ein Webinterface bereitstellt, welches sich z.B. über „iFrames“ beliebig in Websites einbinden lässt.

Die Forschungsplattform für das Projekt basiert auf einem CMS weitgehend ohne speziell geeignete VRE-Erweiterungen (Tools). Sie dient zur Information über das Projekt in der Öffentlichkeit mit geschütztem separatem Zugang für die Wissenschaftler. Einige Messwerte lassen sich in Echtzeit verfolgen und Bilder sowie Videos können in einer Galerie betrachtet werden. Zugriff auf die Forschungsdaten als Download haben nur authentifizierte Nutzer. Die zentrale Authentifizierung erfolgt über den Verzeichnisdienst (LDAP) der BTU Cottbus-Senftenberg. Für wissenschaftliche Beiträge wird erfolgreich ein Wiki im Intranet verwendet. Darüber hinaus findet sich ein Verzeichnis der Publikationen auf der Basis von zentralen Datenbanken. Einige der vom Rechenzentrum der BTU Cottbus-Senftenberg bereitgestellten Services wurden jedoch zwischenzeitlich eingestellt.

Die Kommunikation der Projektmitglieder erfolgt hauptsächlich auf herkömmlicher Basis durch E-Mail, Telefon oder bei Meetings. Ein einfaches Web-Tagebuch für besondere Einträge zu Forschungsaktivitäten auf der Fläche steht ebenso zur Verfügung wie ein webbasierter Terminkalender für gemeinsame Ereignisse. Tabelle 5.8 zeigt die Rahmenbedingungen des Projektes in der Übersicht.

Tabelle 5.8: Rahmenbedingungen des Projektes „Hühnerwasser“ für eine L-TER VRE

Klasse	Beschreibung
Benutzerkreis	Benutzerkreis bestehend aus einem festen Kern und wechselnden Personen mit einer relativ konstanten Anzahl von insgesamt ca. 100 Personen aus verschiedenen Einzelprojekten und unterschiedlichen Institutionen.
Zielgruppen	Wissenschaftler, Studierende, Öffentlichkeit
Institutionen	Verschiedene Lehrstühle und Institutionen der BTU Cottbus-Senftenberg, Geoforschungszentrum (GFZ) Potsdam, TU München, TU Berlin, Leuphana Lüneburg, Universität Hohenheim, Senckenberg Museum für Naturkunde Görlitz, Helmholtz Zentrum für Umweltforschung UFZ
Daten <ul style="list-style-type: none"> ▪ Datenmanagement ▪ Datenvolumen ▪ Datenstrukturen ▪ Datenaufbereitung 	Insgesamt ca. 515 GB in 12 Jahren, DBMS im Aufbau, kein Metadaten-Management-System, inhomogene Daten aus hundert von Sensoren, Datenloggern, Messungen, Beobachtungen, Laborauswertungen, Luftbildern und sonstigen anfallenden Daten, Speicherung der Zeitreihen in MS Excel-Tabellen auf NAS in Verzeichnisstruktur, teils automatisiertes Pre- und Postprocessing, sowie statistische Auswertung am PC.
Aufgaben	Forschungsplattform, Monitoring, Datenhaltung, Datenauswertung, Visualisierung, Information für die Öffentlichkeit, Community
VRE-Schwerpunkt	Forschungsdaten und Community
Zeitlicher Bezug	Bisher 12 Jahre, im Rahmen der Ökosystemforschung so lange, wie möglich (> 30 Jahre)
Geographischer Bezug	Relativ kleines Gebiet

5.2.2 Zielsetzung der L-TER VRE

Für das interdisziplinäre L-TER Großprojekt „Hühnerwasser“ soll eine moderne L-TER VRE aufgebaut werden, welche die vorhandenen ICT-Strukturen schrittweise ablöst. Dazu gehören insbesondere der Aufbau eines Datenbank-Management-Systems (DBMS) und Metadaten-Management-Systems (MMS) für die Forschungsdaten, eine automatisierte Datenaufbereitung sowie die interaktive Auswertung der Daten über eine webbasierte Oberfläche. Der große Umfang des Projektes erfordert zudem viel organisatorischen Aufwand, der weitgehend zentral über die L-TER VRE abgewickelt werden kann. Einen besonderen Mehrwert soll die bisherige Struktur durch eine neue online Forschungs-Community mit „Facebook-Charakter“ erhalten sowie durch die verstärkte Einbeziehung von mobilen Endgeräten, Sensornetzwerken und automatischen Reports für unterschiedliche Zielgruppen. Die bisherigen Wiki-Funktionen sind ausbaufähig, um eine solide „Wissensbasis“ für Forschung, Lehre und die L-TER VRE selbst zu bilden.

Basis L-TER VRE Funktionalitäten für das Projekt „Hühnerwasser“:

- Website mit allgemeinen CMS-Funktionalitäten,
- Gruppen und abstimmbare Zugriffsrechte,
- zentrale Benutzerverwaltung, Authentifizierung und SSO,
- eigene Social-Media-Community mit neuen Kommunikationsformen,
- zentrales DBMS (SQL/noSQL) für die Speicherung von Forschungsdaten,
- Datenmanagement und Metadatenmanagement,
- Wiki mit speziellen Erweiterungen für Forschung, Lehre und die VRE selbst,
- allgemeine Informationen zum Projekt für die Öffentlichkeit, Betreiber, beteiligte Institutionen,
- Projektmanagement (Projekte, Meilensteine, Termine, Kontakte, etc.),
- aktuelle Messwerte auf der Website,
- Publikationsverzeichnis,
- Galerie für multimediale Inhalte,
- Screen-Sharing, Videokonferenzen,
- Remote-Desktop-Services für den Zugriff auf Windows Anwendungen.

Erweiterte Funktionalitäten:

- automatisierte Datenaufbereitung,
- interaktive Auswertung der Forschungsdaten über Web-Oberfläche,
- Aufbau von benutzerspezifischen Workflows,
- Tools für Modellierung und Simulation,
- Reports und Visualisierungen für verschiedene Zielgruppen,
- Aufbau und Einbindung von Sensornetzwerken,
- Optimierung und VRE-Apps für mobile Endgeräte (z.B. Augmented Reality und Dateneingabe im Gelände),
- Diskussionen zu neuen Forschungsaktivitäten (Brainstorming),
- Datenverbünde mit anderen Forschungsnetzwerken.

5.2.3 Konzept und Prototyp

Die Anwendung des „Rapid L-TER VRE“ Konzeptes erfolgt, wie in Kapitel 5.1.3 beschrieben. Die Kurzform des Konzeptes stellt Tabelle B.2 im Anhang B dar. Sie zeigt die einzelnen Bausteine des „Rapid L-TER VRE“ Konzeptes (vgl. Kapitel 4) mit den jeweiligen Parametern. Die speziellen Erweiterungen für das „Joomla“ CMS sind in Tabelle 5.10 gelistet. Abbildung 5.11 zeigt die den Prototypen der VRE mit der Community-Homepage und zugehörigem Menü.

Folgende VRE-Services (Tabelle 5.9) sollen nach den Anforderungen (Tabelle 5.8) zum Einsatz kommen. Die erweiterten VRE-Services sind durch „*“ gekennzeichnet:

Tabelle 5.9: Basis- und erweiterte VRE-Services für „Hühnerwasser“

Kategorie	Software („Schlüsselfertig“: T)
CMS	„Joomla“ (T)
Metadatenmanagement und Datenverbünde*	„GeoNetwork“ (T)
Wiki	„MediaWiki“ (T)
Web-Server / Betriebssystem	„nginx“oder „Apache“ Webserver unter „Linux“ OS (T)
Datenbankserver	„Linux“ OS mit: „mySQL“ für VRE-Services, „PostgreSQL“ für Messdaten (T)
Authentifizierung	„OpenLDAP“ (T)
Community	„Joomla“ mit Erweiterung „JomSocial“ (kostenpflichtige Lizenz) (T)
Projektmanagement	„Colabtive“ (T)
Cloud-Speicher	„OwnCloud“ (T)
Statistik	„R-Server“

Für das „Joomla“ CMS kamen die folgenden, zusätzlich gewählten Erweiterungen zum Einsatz (Tabelle 5.10):

Tabelle 5.10: Verwendete Erweiterungen (VRE-Tools) zum „Joomla“ CMS für „Hühnerwasser“

Name	Beschreibung
AdCustomCSS	Layout und Design
ArtData	Visualisierung und Datenaufbereitung
Community Polls	Umfragen
FAQ Book	Hilfesystem
iSEO (kostenpflichtige Lizenz)	Suchmaschinenoptimierung
JomSocial (kostenpflichtige Lizenz)	Community, Gruppen, Benutzerhomepage, Terminkalender, Social Media, Web 2.0
JS Jobs (kostenpflichtige Lizenz)	Jobbörse
K2	Artikel- und Kategorienmanagement
NoNumber ReReplacer	Text, Übersetzungen
OS PDF Indexer	PDF Verschlagwortung
OSE Anti Hacker (kostenpflichtige Lizenz)	Website-Sicherheit
Shmanic LDAP	Zentrale Authentifizierung
XCloner	Backup

Zur Auswahl dient die Übersicht zu Software für „Joomla“ CMS-Erweiterungen im Anhang D und für „MediaWiki“ Erweiterungen in Tabelle C.8 im Anhang C. In Tabelle 5.10 sind die verwendeten VRE-Tools des „Joomla“ CMS für die L-TER VRE in der Übersicht dargestellt. Tabelle 5.11 zeigt die ausgewählten Erweiterungen für „MediaWiki“.

Tabelle 5.11: Verwendete Erweiterungen (VRE-Tools) zu „MediaWiki“ für „Hühnerwasser“

Name	Beschreibung
Visual Editor	WYSIWYG Editor mit zahlreichen Optionen
Google Charts	Erzeugung von Diagrammen
R / Shiny	Einbindung von „R“ Scripten

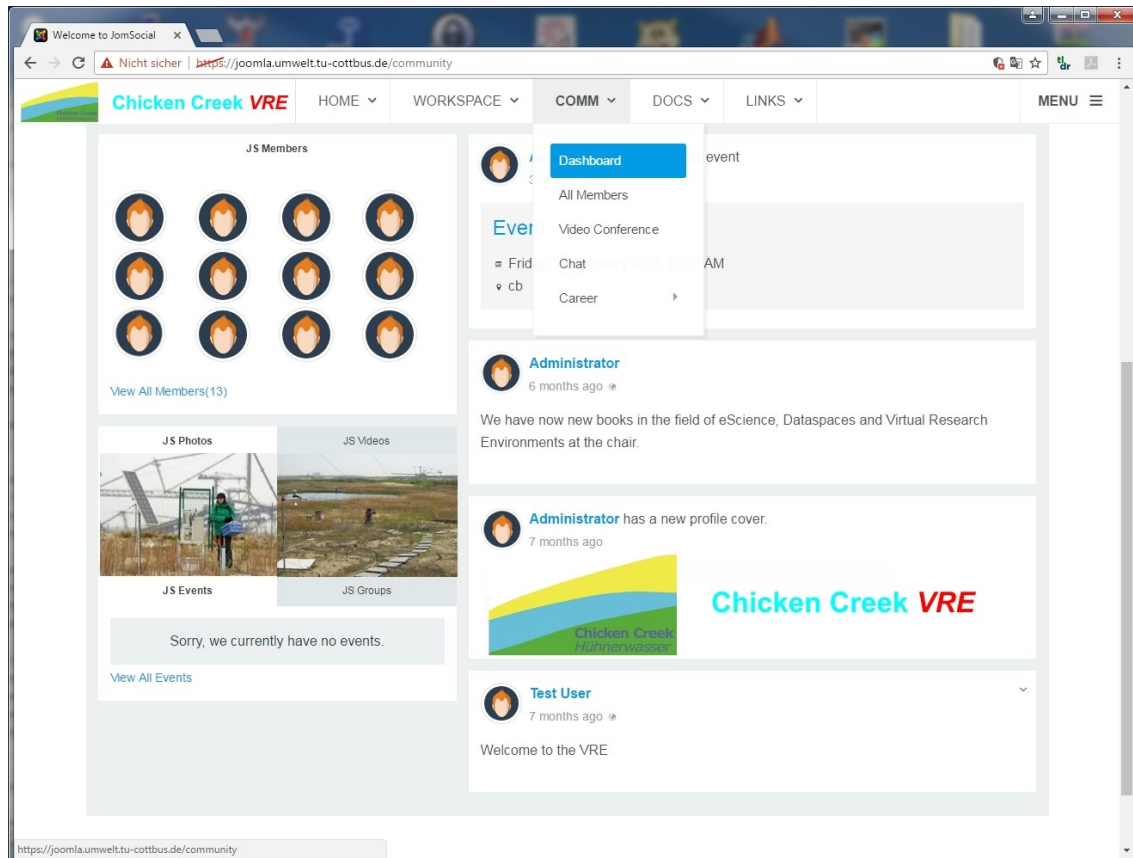


Abbildung 5.11: Prototyp „Hühnerwasser“ L-TER VRE – Community Homepage

5.2.4 Bewertung der Konzeptanwendbarkeit

Das „Rapid L-TER VRE“ Konzept bewährte sich für die gegebenen Rahmenbedingungen einer relativ „großen“ L-TER VRE mit interdisziplinären Teilprojekten, Kooperationen mit vielen beteiligten Instituten, relativ hohe und wechselnde Anzahl an Benutzern. Besonderer Fokus lag auf dem Datenmanagement, der Datenauswertung, sowie einer digitalen Forschungscommunity. Die Forschungsdaten hatten vorwiegend unterschiedliche Datenstrukturen und den Umfang von „Medium Data“. Die allgemeinen Aufgaben und Zielsetzungen des L-TER Projektes konnten realisiert werden.

Für die L-TER VRE „Hühnerwasser“ fasst das *komprimierte* „Rapid L-TER VRE“ Konzept die wichtigsten Parameter für den Aufbau zusammen (vgl. Tabelle B.2 im Anhang B).

Die projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes wurde anhand der klassifizierten Kriterien (vgl. Kapitel 3.1) überprüft. Hinzu kamen die übri-

gen Kriterien zur Nachhaltigkeit (Persistenz, ICT, Mensch), schnelle Realisierung (Entwicklungszeit) und zur Virtualisierung (Größe, Struktur, Schwerpunkte, Dynamik) (vgl. Kapitel 3.1). Die Zusammenfassung stellt Tabelle 5.12 dar.

Somit konnte eine *nachhaltige* L-TER VRE, mit den gegebenen Rahmenbedingungen und Anforderungen, als Prototyp *schnell* (vgl. Kapitel 6.3) und einsatzfähig umgesetzt sowie im Alpha-Stadium (ohne Benutzer) getestet werden.

Tabelle 5.12: Projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes für „Hühnerwasser“

Klasse	Kriterium	Bemerkungen	Konzeptanwendbarkeit
Größe	Benutzerzahl	Authentifizierung mit LDAP Verzeichnisservern und SSO erforderlich, viele dynamische Änderungen, höherer Administrationsaufwand, Faktor „Mensch“ hat größeren Einfluss	✓
	Anzahl Untersuchungsobjekte	Forschungsdatenmanagement gut realisierbar mit „GeoNetwork“, „Medium Data“, Strukturierungsmöglichkeiten erforderlich, Erweiterte Suche- und Filterfunktionen	✓
	Datenvolumen	„Medium Data“ lassen sich gut digital verarbeiten	✓
	Projektanzahl (Teilprojekte)	Die Komponenten sind allgemein gut für verschiedene Teilprojekte geeignet, jedoch ist der Aufwand zur Konfiguration bei Teilprojekten größer. Teilprojekte erfordern ggf. Anpassung oder Weiterentwicklung der VRE	✓
	Umfang der ICT-Infrastruktur	Viele VRE-Services und Tools erhöhen die Komplexität des Gesamtsystems, mittlere Systemperformance erforderlich	✓
	Geographische Ausdehnung (Untersuchungsgebiet, Benutzer)	(Geo-)DBMS, Einhaltung internationaler Gesetze und Standards, kulturelle Unterschiede, zeitliche Unterschiede, mindestens englische Übersetzung	✓
Struktur	Institutionen	Institutionenübergreifende Authentifizierung erforderlich, Hürden bei institutionellen Einrichtungen (z.B.: Verwaltungen und Rechenzentren)	✓
	Datenstrukturen	Individuelle Strukturen erforderlich, „PostgreSQL“ (Geo-Daten), ggf. NoSQL DBMS für diverse Zeitreihen, Metadaten-Schema, Kopplung von Modellen, Schnittstellen	✓
	Datenverbünde	In der erweiterten L-TER VRE vorgesehen, schnell realisierbar mit „GeoNetwork“ und nach INSPIRE	✓
Schwerpunkte	Fachliche Schwerpunkte	Viel Programmieraufwand erforderlich: Datenaufbereitung, -Auswertung, Workflows, Statistik, Datenbanken für Zeitreihen, VRE-Tools	✓
	VRE-Schwerpunkte	Eignet sich für Schwerpunkt auf Forschungsdaten, individuelle Programmierung erforderlich, „GeoNetwork“ für Forschungsdatenmanagement,	✓

Klasse	Kriterium	Bemerkungen	Konzeptanwendbarkeit
		gut geeignet für Communities	
Entwicklungszeit	Basis-VRE (Prototyp)	Schnelle Realisierung des Prototyps in ca. einer Woche	✓
Virtualisierung	Größe, Anzahl, Schwerpunkte, Dynamik	Größe: VM Dateigröße relativ „mittel“ (ca.4 GB Basissystem) + 100 GB für Daten reserviert Anzahl: Mittlere Anzahl an VMs (ca. 8 Stk.) Schwerpunkte: VMs für Forschungsdatenmanagement und Community Service („Joomla“ CMS mit „JomSoical“ Erweiterung) vorhanden Dynamik: VMs benötigen mittlere dynamische Eigenschaften (Load balancing, Hardwareperformance, Speicherplatz, Backup, Archivierung) Script für VM Administration erfolgreich getestet	✓

Die Evaluation des Projektes „Hühnerwasser“ kommt prinzipiell zu ähnlichen Ergebnissen, wie im ersten Projekt (vgl. Kapitel 5.1.4). Die folgenden Unterschiede und Bemerkungen sind zu ergänzen:

- Durch die größere Anzahl an Services und Softwarekomponenten ist die Anbindung an Verzeichnisserver (LDAP) zur Benutzerauthentifizierung inklusive SSO-Funktionalitäten mit größerem Aufwand verbunden, da nicht alle Komponenten mit diesen Basisfunktionalitäten ausgestattet sind.
- „MediaWiki“ lässt sich durch zahlreiche Erweiterungen für Multimediale Inhalte zu einem umfangreichen Werkzeug für die Erstellung von Berichten und Arbeiten in Teamarbeit mit Versionskontrolle erweitern.
- Eine Social-Media-Community ist leicht zu implementieren, erweitert die Kommunikationsmöglichkeiten und „belebt“ durch die verstärkte zwischenmenschliche Kommunikation den Forschungsprozess.
- Nach Aufbau einer Basis-VRE wird auf Grund des Projektumfangs vermutlich eine Vollzeitstelle für die Administration und die Weiterentwicklung benötigt.
- Eine enge Kooperation mit dem Rechenzentrum der Universität ist dringend erforderlich, um die ICT-Struktur der VRE ggf. in die Gesamtstruktur der BTU einzubinden, sicherheitsrelevante Maßnahmen abzustimmen und gemeinsame Synergien zu nutzen.
- Alternativ zu einer eigenen ICT-Struktur kann es aus organisatorischen, personellen, finanziellen oder rechtlichen Gründen sinnvoll sein, einen externen Dienstleister für die VRE ausgründen oder zu beauftragen.

Für das Projekt „Hühnerwasser“ sollte eine VRE im Produktivbetrieb aufgebaut werden, da es sich um ein interdisziplinäres Großprojekt handelt, bei dem konstante

Benutzergruppen über längere Zeit hinweg die VRE nutzen. Das „Rapid L-TER VRE“ Konzept eignet sich hierfür. Eine digitale Forschungs-Community analog zu „Facebook“ bringt einen deutlichen Mehrwert zu den momentan vorhandenen Möglichkeiten der Kommunikation im Projekt. Die alte ICT kann schrittweise abgelöst werden, sobald eine Basis-VRE bereit steht, die den bisherigen Funktionsumfang abdeckt. Schwierig wird es bei diesem umfänglichen Projekt, allen Anforderungen der beteiligten Disziplinen und beteiligten Gruppen gerecht zu werden. Ein hoher Aufwand für die Administration und die Weiterentwicklung einer derart umfangreichen VRE wird viele Ressourcen benötigen – wobei aber die Vorteile überwiegen sollten und die Wettbewerbsfähigkeit des Forschungsprojekts für weitere externe finanzielle Förderungen gestärkt wird.

5.3 L-TER VRE für Projekt „HydroEurope“ in der Lehre

Das „HydroEurope“-Projekt (hydroeurope.org 2017) entstand im Rahmen der europäischen Ausbildung im Wasserwesen durch eine enge Partnerschaft der akademischen Ausbildung mit der Wirtschaft und öffentlichen Institutionen. Es unterscheidet sich von anderen Master-Ausbildungsprogrammen durch eine teamorientierte und intensive praktische Komponente in Verbindung mit vertieften theoretischen Grundlagen. An sechs verschiedenen europäischen Standorten der Partnerinstitutionen arbeiten die Studierende, Wissenschaftler und andere Projektteilnehmer derzeit in mehreren Teams zusammen an gemeinsamen Aufgaben im Wasserwesen. Im Durchschnitt handelt es sich um ca. zehn Gruppen mit je sechs bis zehn Teilnehmer an sechs Institutionen. Dabei findet die Lehre in Form des integrierten Lernens (engl. blended learning) statt, bei der die Teilnehmer drei Monate online und zwei Wochen an einem Ort zusammenkommen.

Zu den Aufgaben der Studierenden gehören unter anderem die Einarbeitung, Anwendung und Bewertung branchenüblicher Software, die zur Lösung praxisbezogener Übungsszenarien verwendet wird. Dabei fallen viele Daten unterschiedlichster Form und Projektdateien der einzelnen Softwareprogramme an. Über die einzelnen Ergebnisse müssen die örtlich verteilten Teammitglieder miteinander diskutieren und online zusammenarbeiten, bevor alle Teilnehmer sich für zwei Wochen zum Projektabschluss an einem Ort treffen. Das Endresultat ist eine von jedem Team und Jahrgang erstellte „Team-Website“ („Google Sites“), auf der alle Ergebnisse öffentlich präsentiert werden.

Das Projekt eignet sich besonders in Hinblick auf Untersuchungen über das örtlich verteilte Arbeiten im Team und die Kommunikation unter den Projektteilnehmern. Die Zielgruppe besteht vornehmlich aus Studierenden, die den Umgang mit ICT und sozialen Netzwerken gewohnt sind. Zudem können hier Erfahrungen gesammelt werden, inwieweit sich der Einsatz von Open-Source-Software und eigenen Services im Gegensatz zu den proprietären ICT Systemen und Services von Drittanbietern realisieren und integrieren lässt.

5.3.1 Aktueller Stand des Projektes

In der Umsetzung wird ein pädagogisches Konzept auf der Basis von ICT-Plattformen angewendet, das die Zusammenarbeit (collaborative engineering) unterstützt (Molkenthin 2004). Das System basiert vornehmlich auf dem „Jahia“ (jahia.com 2017) Content-Management-System (CMS) für die Website des Studiengangs, Videokonferenzsystemen („Adobe Connected Pro“, „Skype“) und „Google“ Services („Google Cloud“, „Google Classroom“, „Google Sites“) als gemeinsamer Datenpool und für die Erstellung von Dokumenten im Team. Die Kommunikation findet auf individuellen Wegen z.B. per E-Mail, „Facebook“, „Twitter“, Videokonferenz mit Screen-Sharing statt. Bis auf das CMS wird dabei vornehmlich proprietäre Software und unterstützenden Ser-

vices von Drittanbietern (insbesondere Google) genutzt. Tabelle 5.13 zeigt die zusammengefassten Rahmenbedingungen für das Projekt.

Tabelle 5.13: Rahmenbedingungen des Projektes „HydroEurope“ für eine L-TER VRE

Klasse	Beschreibung
Benutzerkreis	Jährlich über 100 verschiedene Personen (10 Studierende in 10 Gruppen, Lehrende)
Zielgruppen	Universitäre Ausbildung, Studierende, Studiengang, Lehrende
Institutionen	University of Nice Sophia-Antipolis (Frankreich), BTU Cottbus-Senftenberg (Deutschland), Technical University of Catalonia (Spanien), Newcastle University (Großbritannien), Warsaw University of Technology (Polen), Vrije Universiteit Brussel (Belgien)
Daten <ul style="list-style-type: none"> ▪ Datenmanagement ▪ Datenvolumen ▪ Datenstrukturen ▪ Datenaufbereitung 	Datenvolumen: ca. 10 GB pro Gruppe und Jahrgang. Berichte, Fotos / Videos, Messdaten (Abfluss), Vektor- und Rasterdaten evtl. als ZIP-File zusammengefasst und mit Metadaten versehen. 15 Tutorials mit Daten bei „Google Classroom“ Sonstige kleine Datenmengen in unstrukturierter Form.
Aufgaben	Lehre und Forschung: Flood Risk Analysis, praxisorientierte Gruppenarbeit, online-Zusammenarbeit, Datenaustausch, Onlinemeetings, Umfragen, Videokonferenz, Screen-Sharing, Versionsverwaltung (für Berichte), gemeinsame Dokumentbearbeitung, Kommunikation. Jede Gruppe erzeugt eine Website für die Öffentlichkeit als Zusammenfassung unter HydroEurope.org
VRE-Schwerpunkt	Community, weniger Forschungsdaten
Zeitlicher Bezug	Die Phase der Gruppenarbeit von Studierenden dauert pro Jahrgang ca. ein Semester. Das Archiv der Gruppenarbeit soll sich über die gesamte Lebensdauer des Studiengangs erstrecken. Voraussichtlich mehrere Jahrzehnte.
Geographischer Bezug	Weltweit

Die alte ICT-Struktur, derzeit basierend auf Kooperationstools von „Google“ (z.B. Google Classroom), soll erweitert und unabhängig von Drittanbietern als gesamtheitliche L-TER VRE für das „HydroEurope“-Projekt aufgebaut werden. Insbesondere soll die L-TER VRE zur Verbesserung einer koordinierten Teamarbeit mit modernen Kommunikationswerkzeugen und Social Media im „Eigenbau“ beitragen.

5.3.2 Zielsetzung der L-TER VRE

Das übergeordnete Ziel des Projektes liegt darin, verschiedene auf dem Markt befindliche Softwareprodukte im Bereich des Wasserwesens in Teamarbeit anzusehen und zu untersuchen, um damit Kompetenzen aufzubauen.

Basis L-TER VRE Funktionalitäten für „HydroEurope“:

- Website mit allgemeinen CMS-Funktionalitäten,
- Gruppen und abstimmbare Zugriffsrechte,
- eigene Social-Media-Community,
- zentrale SQL-Datenbank und Speicherung von GIS-Daten in einer GIS-Datenbank,
- Datenmanagement von verschiedenen Softwareprojekten, Modellen und deren Parametern mit Metadaten,
- Management zusätzlicher Hintergrundinformationen (Bilder, Presseartikel, Koordinaten),
- Wiki für die wissenschaftliche Berichtserstellung zu den Softwareprodukten,
- Wiki für Hilfe, Dokumentation und Teamarbeit mit speziellen Erweiterungen für Inhalte,
- zentrale Benutzerverwaltung, Authentifizierung und SSO,
- allgemeine Informationen für die Öffentlichkeit zum Projekt, Betreiber, beteiligte Institutionen,
- Screen-Sharing, Videokonferenzen,
- Remote-Desktop-Services für den Zugriff auf Windows Anwendungen.

Erweiterte L-TER VRE Funktionalitäten für „HydroEurope“:

- Automatisierung des Forschungsprozesses und der Team Workflows allgemein,
- Langfristige Datenspeicherung für nachfolgende Studienjahrgänge,
- Web-GIS Erweiterungen,
- Durchführung und Verwaltung von Simulationen mit Metadaten
- Lern-Management-System für Studienmodul,
- Quellcode-Repository für die Entwicklung von eigener Software im Team,
- Projektplanungswerkzeuge für die Team-Organisation,
- Wissensmanagement,
- Evaluierung mit Fragebögen.

5.3.3 Konzept und Prototyp

Die Anwendung des „Rapid L-TER VRE“ Konzeptes erfolgte, wie in Kapitel 5.1.3 beschrieben. Eine Kurzform des Konzeptes stellt Tabelle B.3 im Anhang B dar. Sie zeigt die einzelnen Bausteine des „Rapid L-TER VRE“ Konzeptes (vgl. Kapitel 4) mit den jeweiligen Parametern. In Tabelle 5.15 sind die ausgewählten speziellen Erweiterungen für das „Joomla“ CMS gelistet. Als Ergebnis der Konzeptanwendung war der Prototyp für die L-TER VRE nach ca. einer Woche einsatzbereit für Betatests. Abbildung 5.12 zeigt die Homepage der L-TER VRE mit dem Workspace und dem zugehörigem Menü.

Folgende VRE-Services (Tabelle 5.14) sollen nach den Anforderungen (Tabelle 5.13) zum Einsatz kommen. Die erweiterten VRE-Services sind durch „*“ gekennzeichnet:

Tabelle 5.14: Basis- und erweiterte VRE-Services für „HydroEurope“

Kategorie	Software („Schlüselfertig“: T)
CMS	„Joomla“ (T)
Metadatenmanagement und Datenverbünde Metadatenmanagement für Simulationen*	„GeoNetwork“ (T)
Wiki	„MediaWiki“ (T)
Web-Server / Betriebssystem	„nginx“oder „Apache“ Webserver unter „Linux“ OS (T)
Datenbankserver	„Linux“ OS mit: „mysql“ für VRE-Services, „PostgreSQL“ für Messdaten (T)
Simulationen*	Spezifische Software (z.B.: „Q-GIS“, „Netlogo“)
Authentifizierung	„OpenLDAP“ (T)
Community	„Joomla“ mit Erweiterung „JomSocial“ (kostenpflichtige Lizenz) (T)
Code-Repository*	„Mantis“ (T)
Lern-Management-System*	„Moodle“ (T)
Projektmanagement	„Colabtive“ (T)
Cloud-Speicher	„OwnCloud“ (T)
Statistik	„R-Server“

Für das „Joomla“ CMS kamen die folgenden, zusätzlich gewählten Erweiterungen zum Einsatz (Tabelle 5.15):

Tabelle 5.15: Verwendete Erweiterungen (VRE-Tools) zum „Joomla“ CMS für „HydroEurope“

Name	Beschreibung
AdCustomCSS	Layout und Design
ArtData	Visualisierung und Datenaufbereitung
Community Polls	Umfragen
FAQ Book	Hilfesystem
iSEO (kostenpflichtige Lizenz)	Suchmaschinenoptimierung
JChatSocial (kostenpflichtige Lizenz)	Videokonferenz, Chat mit Übersetzung
JomSocial (kostenpflichtige Lizenz)	Community, Gruppen, Benutzerhomepage, Terminkalender, Social Media, Web 2.0
JS Jobs (kostenpflichtige Lizenz)	Jobbörse
K2	Artikel- und Kategorienmanagement
NoNumber ReReplacer	Text, Übersetzungen
OS PDF Indexer	PDF Verschlagwortung
OSE Anti Hacker (kostenpflichtige Lizenz)	Website-Sicherheit
Shmanic LDAP	Zentrale Authentifizierung

Name	Beschreibung
XCloner	Backup

Tabelle 5.16 zeigt die ausgewählten Erweiterungen für „MediaWiki“:

Tabelle 5.16: Verwendete Erweiterungen (VRE-Tools) zu „MediaWiki“ für „HydroEurope“

Name	Beschreibung
Visual Editor	WYSIWYG Editor mit zahlreichen Optionen
Google Charts	Erzeugung von Diagrammen
R / Shiny	Einbindung von „R“ Scripten

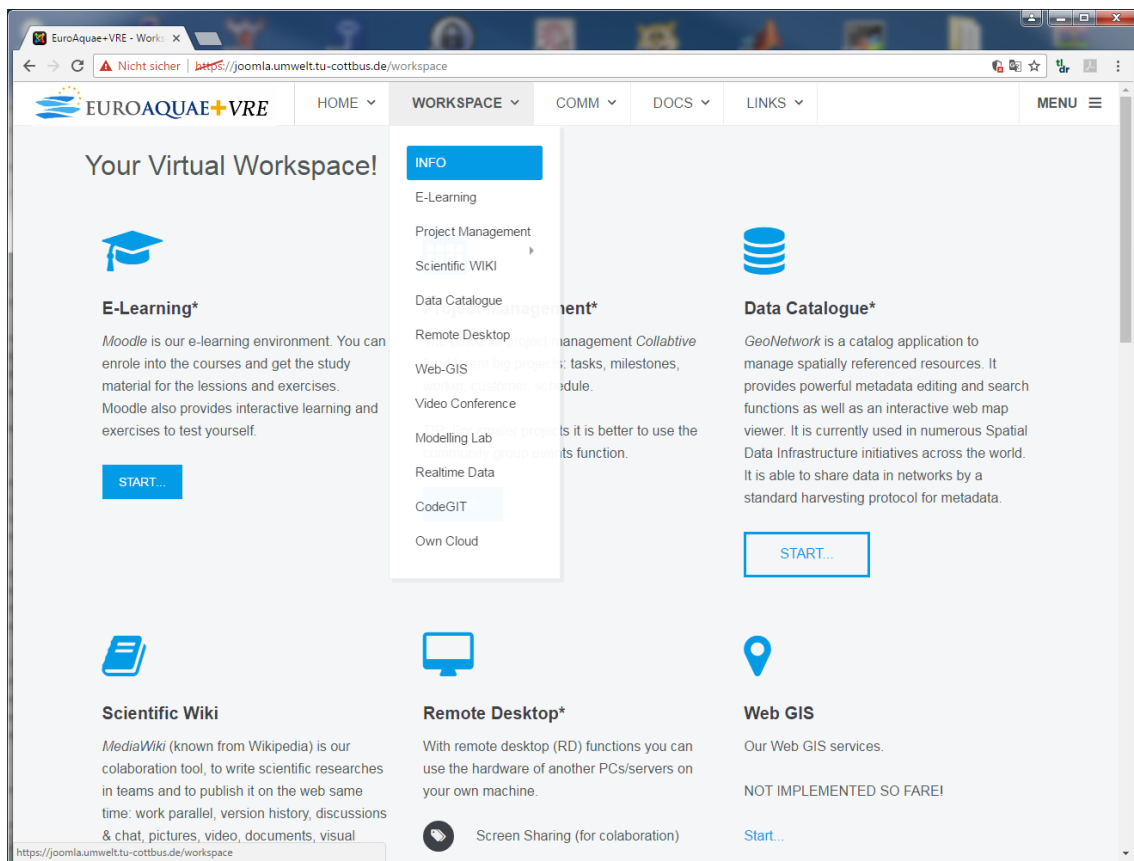


Abbildung 5.12: Homepage des VRE-Prototyps für „HydroEurope“ – Workspace Menü

5.3.4 Bewertung der Konzeptanwendbarkeit

Das „Rapid L-TER VRE“ Konzept bewährte sich für die gegebenen spezifischen Rahmenbedingungen einer artverwandten L-TER VRE für das Projekt „HydroEurope“ aus dem Bereich der kollaborativen und internationalen Lehre im Wasserwesen und der Umweltinformatik. Hier lag der Langzeitbezug weniger bei einzelnen Projekten, sondern in der Kontinuität der Lehre. Das Projekt umfasste eine relativ konstante, „hohe“ Benutzeranzahl.

Für die L-TER VRE „HydroEurope“ fasst das *komprimierte* „Rapid L-TER VRE“ Konzept die wichtigsten Parameter für den Aufbau zusammen (vgl. Tabelle B.3 im Anhang B).

Die projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes wurde anhand der klassifizierten Kriterien (vgl. Kapitel 3.1) überprüft. Hinzu kamen die übrigen Kriterien zur Nachhaltigkeit (Persistenz, ICT, Mensch), schnelle Realisierung (Entwicklungszeit) und zur Virtualisierung (Größe, Struktur, Schwerpunkte, Dynamik) (vgl. Kapitel 3.1). Die Zusammenfassung stellt Tabelle 5.17 dar.

Tabelle 5.17: Projektspezifische Anwendbarkeit des „Rapid L-TER VRE“ Konzeptes für „HydroEurope“

Klasse	Kriterium	Bemerkungen	Konzeptanwendbarkeit
Größe	Nutzerzahl	Authentifizierung mit LDAP Verzeichnisservern und SSO erforderlich, sehr hohe Dynamik, höherer Administrationsaufwand, Faktor „Mensch“ hat größeren Einfluss	✓
	Anzahl Untersuchungsobj.	Strukturierungsmöglichkeiten erforderlich, Erweiterte Suche- und Filterfunktionen	✓
	Datenvolumen	„Small Data“ lassen sich gut digital verarbeiten	✓
	Projektanzahl (Teilprojekte)	Die Komponenten sind allgemein gut für verschiedene Teilprojekte geeignet, jedoch ist der Aufwand zur Konfiguration bei Teilprojekten größer. Teilprojekte erfordern ggf. eine Anpassung weniger eine Weiterentwicklung der L-TER VRE	✓
	Umfang der ICT-Infrastruktur	Viele VRE-Services und Tools erhöhen die Komplexität des Gesamtsystems, mittlere Systemperformance erforderlich	✓
	Geographische Ausdehnung (Untersuchungsgebiet, Benutzer)	PostgreSQL (Geo-)DBMS (Geo-)DBMS, Einhaltung internationaler Gesetze und Standards, kulturelle Unterschiede, zeitliche Unterschiede, Multilinguale Übersetzung	✓
Struktur	Institutionen	Institutionenübergreifende Authentifizierung erforderlich, Hürden bei institutionellen Einrichtungen (z.B.: Verwaltungen und Rechenzentren)	✓
	Datenstrukturen	Individuelle Strukturen erforderlich, „PostgreSQL“ (Geo-Daten), Kopplung von Modellen	✓
	Datenverbünde	Nicht relevant	✓
Schwerpunkte	Fachliche Schwerpunkte	Wenig Programmieraufwand erforderlich	✓
	VRE-Schwerpunkte	Sehr gut geeignet für Communities	✓
Entwicklungszeit	Basis-VRE (Prototyp)	Schnelle Realisierung des Prototyps in ca. einer Woche	✓
Virtualisierung	Größe, Anzahl, Schwerpunkte, Dynamik	Größe: VM Dateigröße relativ „mittel“ (ca.4 GB Basissystem) + 1 TB Datenspeicher Anzahl: Mittlere Anzahl an VMs (ca. 10 Stk.)	✓

Klasse	Kriterium	Bemerkungen	Konzeptanwendbarkeit
		Schwerpunkte: Community Service („Joomla“ CMS mit „JomSoical“ Erweiterung) und VMs für Forschungsdatenmanagement vorhanden Dynamik: VMs benötigen mittlere dynamische Eigenschaften (Load balancing, Hardwareperformance, Speicherplatz, Backup, Archivierung)	

Somit konnte eine *nachhaltige* L-TER VRE, mit den gegebenen Rahmenbedingungen und Anforderungen, als Prototyp *schnell* und gut umgesetzt sowie im Alpha-Stadium (ohne Benutzer) getestet werden.

Die Evaluation des Projektes „HydroEurope“ kommt prinzipiell zu ähnlichen Ergebnissen, wie bei den vorangegangenen Projekten (vgl. Kapitel 5.1.4, 5.2.4). Insbesondere sind diese Ergebnisse für die Authentifizierung mit SSO und für „MediaWiki“ hier ebenso zutreffend. Die folgenden Unterschiede und Bemerkungen sind zu ergänzen:

- Report-Layer und Datenverbünde wurden nicht benötigt.
- Das Lern-Management-System ist eine mächtige Erweiterung im Bereich der Lehre und des Studiums.
- Eine Social-Media-Community scheint für die Zielgruppe ein besonders geeignetes Medium zu sein, um die Kommunikation zu fördern und mehr Zusammengehörigkeitsgefühl zu vermitteln.
- Jobbörsen mit Jobangeboten, Praktika und Abschlussarbeiten beziehen potentielle Arbeitgeber mit ein und geben den Studierenden zusätzliche Motivation.
- Remote-Desktop-Services eignen sich besonders für die verteilte Nutzung von Software auf einem anderen Server über das Internet. Damit kann Installationsaufwand seitens der Anwender minimiert, sowie Kosten für Lizenzen gespart werden.
- Screen-Sharing ist hilfreich für die Unterstützung der Studierenden durch Tutoren, für Vorträge und für die Teamarbeit.
- Videokonferenzen gestalten die Kommunikation reeller und eignen sich für Vorträge, Gruppenmeetings und Einzelgesprächen.
- Cloud Computing bietet pro Nutzer einen eigenen Bereich im Netz zum Speichern und synchronisieren von Daten.

Die Einarbeitung für die Studierenden in die L-TER VRE benötigt einen extra zeitlichen Overhead, wobei die L-TER VRE im Anschluss „nur“ ein bis zwei Semester genutzt werden soll. Dieser Faktor ist „erfolgskritisch“ für die L-TER VRE, da der Vorteil einer L-TER VRE ggf. durch die Zeit der Einarbeitung aufgehoben wird. Allerdings ist der langfristige Aspekt für die Betreiber (Wissenschaftler in der Lehre) zu berücksichtigen. Eine L-TER VRE muss daher umso intuitiver gestaltet sein und auf der

Logik von weit verbreiteten, anderweitigen Softwareprodukten aufbauen. Bei jüngerem Zielpublikum ist das intuitive Verständnis für ICT oftmals größer. Die L-TER VRE sollte dennoch aufgebaut werden, da hier ein generationenübergreifendes Wissen zusammengetragen wird und sich der Studiengang besonders durch sein innovatives ICT-Konzept in der Lehre auszeichnet, wie beschrieben.

6. Gesamtbewertung des „Rapid L-TER VRE“ Baukastenkonzeptes

Das folgende Kapitel fasst die Bewertungen der Konzeptanwendbarkeit der verschiedenen L-TER Projekte hinsichtlich der projektspezifischen Rahmenbedingungen (vgl. Kapitel 5.1.4, 5.2.4, 5.3.4), der theoretischen Bewertung der Nachhaltigkeit, der Entwicklungszeit und der einzelnen Konzeptbestandteile zusammen:

Das „Rapid L-TER VRE“ Baukastenkonzept ist flexibel einsetzbar für die schnelle Realisierung von nachhaltigen L-TER VREs für L-TER Projekte mit einem breiten Spektrum an projektspezifischen Rahmenbedingungen unter Einbezug des Menschen, der als maßgeblicher Faktor wirken kann.

Im Einzelnen:

Das „Rapid L-TER VRE Konzept“ fasste in 10 Punkten die „wichtigen“ Aspekte für L-TER VREs zusammen, die für den erfolgreichen Aufbau von Prototypen zu den drei Anwendungsbeispielen erforderlich waren. Darüber hinaus ließ es sich flexibel für andere L-TER VREs in einem breiten Spektrum an projektspezifischen Eigenschaften nutzen. Die Anwendbarkeit des Konzeptes ist auch auf „größere“ L-TER Projekte mit individuellen Rahmenbedingungen und fachlichen Schwerpunkten übertragbar (theoretisch bewertet). Alle Punkte des Konzeptes tragen erheblich zur Nachhaltigkeit und schnellen Entwicklung von Prototypen bei, was auch für die Weiterentwicklung einer L-TER VRE theoretisch bewertet wurde. Der Aufwand für Systemadministratoren zur Systempflege wurde in vielen Bereichen reduziert gegenüber zu Systemen ohne z.B. zentrales SSO, Verzeichnisserver, Virtualisierung und Metadatenmanagement, sowie den spezifischen Erweiterungen für die einzelnen VRE-Services z.B. für das CMS.

Die Nachhaltigkeit konnte im Rahmen dieser Arbeit nicht praktisch überprüft werden (lange Projektlaufzeit erforderlich). Sie wurde jedoch theoretisch auf Basis der Anwendungsbeispiele bewertet. Entscheidend für die Nachhaltigkeit ist insbesondere: Der Faktor „Mensch“ mit seinem sozialen Umfeld:

Das „Rapid L-TER VRE“ Baukastenkonzept ist geeignet unter dem Aspekt der Nachhaltigkeit.

Die Entwicklungszeit der Prototypen für die Anwendungsbeispiele wurde erheblich reduziert:

Das „Rapid L-TER VRE“ Baukastenkonzept ist geeignet zur schnellen Entwicklung von Prototypen und für „Basis-“ sowie „erweiterte“ L-TER VREs im Produktivbetrieb.

Die Gesamtbewertung mit den Kriterien: Flexibilität, Nachhaltigkeit, Entwicklungszeit, Konzeptbestandteile, fasst Tabelle 6.1 zusammen:

Tabelle 6.1: Gesamtbewertung des „Rapid L-TER VRE“ Baukastenkonzeptes

Kriterium	Bewertung	Eignung
Flexibilität	Projektgröße: kleine bis große Projekte (relativ), Projektstruktur: wenig bis komplexe Strukturen (relativ) bei: Institutionen, Datenstrukturen, Datenverbänden, Projektschwerpunkte: fachspezifisch, Forschungsdaten, Community, Projektzeitraum: minimal ca. 3 Jahre (Aufwand/Nutzen), kurz bis lang, Virtualisierung: dynamische Anpassung durch Virtualisierung.	✓
Nachhaltigkeit	Persistenz: Langzeitdatenspeicherung inkl.: Formate, Standards, Infrastruktur, VMs, Backups, Archivierung ICT: Virtualisierung, Modularisierung, SOA, offene internationale und weltweite Standards, Schnittstellen, Open-Source-Software (kostenlos) mit Entwicklergemeinde, Datenverbände. Die ICT im Konzept ist hinsichtlich der Persistenz, Konstanz, Resistenz, Elastizität, Resilienz, insbesondere z.B. durch Virtualisierungstechnologien dazu geeignet, nach großen Störungen schnell wieder in den Zustand vor der Störung zu gelangen. Innovative Technologien sind in Zukunft ggf. möglich. Faktor „Mensch“: Standards (gesellschaftlich), benutzerorientiert, Transparenz, Dokumentation, Hilfe, Feedback, zielgruppenorientiert.	✓
Entwicklungszeit	durch Virtualisierung erheblich verkürzt, durch „Schlüsselfertige“ Systeme: erheblich verkürzt, durch Programmieraufwand: erheblich verkürzt, durch Baukasten: erheblich verkürzt, Weiterentwicklung: erheblich verkürzt.	✓
Konzeptbestandteile	Das „Rapid L-TER VRE Konzept“ fasste in 10 Punkten alle notwendigen Aspekte zusammen, die für den erfolgreichen Aufbau von Prototypen zu den Anwendungsbeispielen erforderlich waren. Darüber hinaus ließ es sich flexibel für andere L-TER VREs in einem breiten Spektrum an projektspezifischen Eigenschaften Nutzen. Die Anwendbarkeit des Konzeptes ist auch auf „größere“ L-TER Projekte übertragbar. Alle Punkte des Konzeptes haben erheblich in Nachhaltigkeit und Entwicklungsgeschwindigkeit zur Realisierung beigetragen haben.	✓

Eine differenzierte Bewertung erfolgt in den jeweiligen Unterkapiteln für die einzelnen Parameter.

6.1 Bewertung projektspezifischer Rahmenbedingungen

Das „Rapid L-TER VRE“ Konzept bewährte sich flexibel für die drei Anwendungsbeispiele mit den jeweiligen projektspezifischen Rahmenbedingungen (Tabelle 6.2). Die ausgewählten L-TER Projekte decken ein „kleines“ bis „mittleres“ Spektrum an Rahmenbedingungen weitgehend ab (vgl. Abbildung 5.1).

Tabelle 6.2: Bewertung der Flexibilität des „Rapid L-TER VRE“ Konzept

Kriterium	Flexibilität (Anwendungsbeispiele)	Eignung	Flexibilität (theoretisch)	Anwendbarkeit
Projektgröße	klein-mittel (bis ca. 100 Benutzer, dynamischer Benutzerwechsel, bis ca. 20 Teilprojekte, Internationale Projektgruppen, Datenvolumen bis zu wenigen Terrabytes)	✓	groß	✓
Projektstruktur	klein-mittel (bis ca. 20 beteiligte Institutionen, komplexe Datenstrukturen, einzelne Datenverbünde)	✓	groß	✓
Projektschwerpunkte	Fachspezifisch, Forschungsdaten, Community	✓	Fachspezifisch, Forschungsdaten, Community	✓
Zeiteinfluss im Projektzeitraum	Entwicklung eines Basis- Prototypen in wenigen Tagen und Wochen Insgesamt eignet sich das Konzept flexibel unter zeitlichem Bezug zu den einzelnen Lebensphasen einer L-TER VRE für kurze bis lange (relativ) L-TER Projekte	✓	Insgesamt eignet sich das Konzept flexibel unter zeitlichem Bezug zu den einzelnen Lebensphasen: Planung, Entwicklung, Betrieb, Administration, Schulung und Support, Weiterentwicklung	✓
Virtualisierung	4-10 VMs, Ca. 4 GB Basis- system + 100 GB - 1 TB Datenspeicher, für alle Schwerpunkte, dynamisch	✓	Limitierender Faktor sind eher die Hard- und Softwareressourcen.	✓

Durch die Verwendung von VMs konnte für die Projekte erheblich am Umfang von eingesetzten Hardwarekomponenten gespart werden. Jede einzelne „schlüsselfertige“ VM benötigt ansonsten einen eigenen physikalischen Hardware-Server. Auch selbst aufgesetzte Server können und sollten aus den genannten Gründen virtualisiert werden.

So konnten, je nach Leistung eines durchschnittlichen Servers (INTEL i7 Prozessor, 16-32 GB RAM, ECC RAM), ca. fünf bis zehn VMs auf einem Hardware-Server betrieben werden. Dabei ist besonders auch der RAM-Speicher von mindestens ca. 4 GB pro VM, mit integrierter ECC-Fehlerkorrektur zu beachten. Dies gilt jedoch nicht unbedingt für den Produktivbetrieb, in der etwas weniger VMs auf einer gemeinsamen physikalischen Hardware laufen sollten. Eine Optimierung kann dadurch erfolgen, dass diejenigen VMs kombiniert werden, die einen ausgeglichenen „Serverload“ produzieren. Ändern sich im Produktivbetrieb die Anforderung an die Performance der VRE-

Hardware (CPU-Anzahl, Speicherkapazität, RAM-Speicher, etc.), ist die L-TER VRE durch Virtualisierung dynamisch skalierbar. So kann z.B. bei stark anwachsender Zahl von Benutzern oder rechenintensiver Prozesse sofort reagiert werden, was ohne Virtualisierung erheblich komplexer wäre (vgl. Gertsch und Walker 2015). Dies wurde im Ansatz empirisch getestet.

6.1.1 Projektgröße

Die Bewertung der Projektgröße im „Rapid L-TER VRE“ Konzept erfolgte auf Grundlage der jeweiligen Anwendungsbeispiele (vgl. Kapitel 5). Für die Bewertung „größerer“ Projekte wird eine theoretische Bewertung vorgenommen. Die Ergebnisse sind in Tabelle 6.3 zusammengestellt.

Insgesamt eignet sich das Konzept in der flexiblen Anwendung für kleine bis große L-TER Projekte.

Tabelle 6.3: Bewertung der Projektgröße im „Rapid L-TER VRE“ Konzept

Kriterium	Kleine Projekte	Mittlere Projekte	Große Projekte (theoretisch)
Benutzerzahl	Gegeben: Zentrale Authentifizierung durch Verzeichnisserver und SSO nützlich	Gegeben: Zentrale Authentifizierung durch Verzeichnisserver und SSO erforderlich	Gegeben: Zentrale Authentifizierung durch Verzeichnisserver und SSO erforderlich
Anzahl Untersuchungsobjekte	Wenig Relevanz	Gegeben: Strukturierungsmöglichkeiten erforderlich, Erweiterte Suche und Filterfunktionen	Gegeben: Prüfen der maximal zu verarbeitenden Objekte, Spezifische Hard- / Software für die Auswertung von „BigData“
Datenvolumen	Gegeben: „Small Data“ lassen sich sehr gut digital verarbeiten	Gegeben: „Medium-Data“ lassen sich gut digital verarbeiten, dynamische Skalierung ggf. nützlich	Gegeben: „Big Data“ benötigen spezifische Software und Hardware, dynamische Skalierung erforderlich
Projektanzahl (Teilprojekte)	Gegeben: Ein Hauptprojekt: Die VRE-Komponenten sind allgemein für ein Hauptprojekt ausgelegt	Gegeben: Die VRE Komponenten sind allgemein gut für verschiedene Teilprojekte geeignet, jedoch ist der Aufwand zur Anpassung bei Teilprojekten größer.	Überprüfen: maximal mögliche Teilprojekte
Umfang der ICT-Infrastruktur	Gegeben: Wenige VRE-Services und Tools reduzieren die Komplexität des Gesamtsystems, geringe Systemperformance erforderlich	Gegeben: Viele VRE-Services und Tools erhöhen die Komplexität des Gesamtsystems, mittlere Systemperformance erforderlich	Viele VRE-Services und Tools erschweren die Komplexität des Gesamtsystems, hohe Systemperformance erforderlich Daten- und Betriebssicherheit: sehr hohes Niveau

Kriterium	Kleine Projekte	Mittlere Projekte	Große Projekte (theoretisch)
	Daten- und Betriebssicherheit: normales Niveau	Daten- und Betriebssicherheit: hohes Niveau	
Geographische Ausdehnung (Untersuchungsgebiet, Benutzer)	Gegeben: (Geo)-DBMS nützlich, Einhaltung regionaler Gesetze und Standards, keine Fremdsprachenübersetzung	Gegeben: (Geo)-DBMS, Einhaltung internationaler Gesetze und Standards, kulturelle Unterschiede, zeitliche Unterschiede, mindestens Englische Übersetzung	Gegeben: Einhaltung internationaler Gesetze und Standards, kulturelle Unterschiede, zeitliche Unterschiede, Jahreszeitliche Unterschiede, multilinguale Übersetzung

6.1.2 Projektstruktur

Die Bewertung der Projektstruktur im „Rapid L-TER VRE“ Konzept erfolgte auf Grundlage der jeweiligen Anwendungsbeispiele (vgl. Kapitel 5). Die Ergebnisse sind in Tabelle 6.4 zusammengestellt.

Insgesamt eignet sich das Konzept in der flexiblen Anwendung für wenig komplexe bis hin zu komplexen Strukturen in L-TER Projekten.

Tabelle 6.4: Bewertung der Projektstruktur im „Rapid L-TER VRE“ Konzept

Kriterium	Einfache Strukturen	Mittlere Strukturen	Komplexe Strukturen
Institutionen	Gegeben: Institutionenübergreifende Authentifizierung nicht unbedingt erforderlich	Gegeben: Institutionenübergreifende Authentifizierung erforderlich, Hürden bei institutionellen Einrichtungen (z.B.: Verwaltungen und Rechenzentren)	Gegeben: Institutionenübergreifende Authentifizierung erforderlich (ggf. international), Größere Hürden bei institutionellen Einrichtungen (z.B.: Länder, Verwaltungen und Rechenzentren)
Datenstrukturen	Gegeben: Individuelle Strukturen erforderlich, Geo-DBMS	Gegeben: Individuelle Strukturen in mittlerem Umfang, Geo-DBMS, NoSQL DBMS nützlich, Kopplung von Modellen, Semantic Web	Gegeben: Individuelle Strukturen in großem Umfang, Geo-DBMS, NoSQL DBMS nützlich, Kopplung von Modellen, Semantic Web
Datenverbände	Gegeben: bedingt relevant, flache Hierarchien	Gegeben: INSPIRE, GDI-DE, landes-, bundesweite und internationale Netzwerke, mittlere Hierarchien	Gegeben: Weltweite Datenverbände in tiefen Hierarchien

6.1.3 Projektschwerpunkte

Die Bewertung der Projektschwerpunkte im „Rapid L-TER VRE“ Konzept erfolgte auf Grundlage der jeweiligen Anwendungsbeispiele (vgl. Kapitel 5). Die Ergebnisse sind in Tabelle 6.5 zusammengestellt.

Insgesamt eignet sich das Konzept in der flexiblen Anwendung für fachliche und VRE-Schwerpunkte (Forschungsdaten, Community) in L-TER Projekten.

Tabelle 6.5: Bewertung der Projektschwerpunkte im „Rapid L-TER VRE“ Konzept

Kriterium	Kleine Anzahl	Mittlere Anzahl	Hohe Anzahl
Fachliche Schwerpunkte	Gegeben: Wenig fachliche Schwerpunkte, wenig individuelle Programmierung erforderlich	Gegeben: Interdisziplinär, mittlerer Programmieraufwand erforderlich: Datenaufbereitung, -Auswertung, Workflows, Statistik, Datenbanken für Zeitreihen, VRE-Tools	Gegeben: Multidisziplinär, hoher Programmieraufwand erforderlich
VRE-Schwerpunkte	Gegeben: Gut geeignet für Communities Forschungsdaten: „GeoNetwork“, wenig individuelle Programmierung erforderlich	Gegeben: Sehr gut geeignet für Communities (Potentialentfaltung) Forschungsdaten: „GeoNetwork“, mittlerer Programmieraufwand erforderlich	Gegeben: Gut geeignet für Communities (erfordert viel Systemperformance, Informationsflut für Benutzer) Forschungsdaten: „GeoNetwork“, hoher Programmieraufwand erforderlich

6.1.4 Die Zeit in Bezug auf Projektzeitraum und Lebensphasen

Die Bewertung der zeitabhängigen Faktoren im Projektzeitraum mit der Berücksichtigung der einzelnen Lebensphasen im „Rapid L-TER VRE“ Konzept erfolgt auf Grundlage der jeweiligen Anwendungsbeispiele (vgl. Kapitel 5) für die Prototypen. Für die Bewertung der gesamten Lebensphase der L-TER VRE wurde eine theoretische Bewertung vorgenommen. Die Ergebnisse sind in Tabelle 6.6 zusammengestellt.

Insgesamt eignet sich das Konzept flexibel unter zeitlichem Bezug zu den einzelnen Lebensphasen einer L-TER VRE für kurze bis lange (relativ) L-TER Projekte.

Tabelle 6.6: Bewertung der Zeit in Bezug auf Projektzeitraum und Lebensphasen im „Rapid L-TER VRE“ Konzept

Kriterium	Kurze Laufzeit (ca. 1-2 Jahre)	Mittlere Laufzeit (ca. 3-10 Jahre)	Lange Laufzeit (theoretisch über 10 J.)
Planung	Gegeben: Basis- und ggf. erweiterte L-TER VRE	Gegeben: Basis- und erweiterte L-TER VRE	Gegeben: Basis- / erweiterte-, innovative L-TER VRE
Entwicklung	Gegeben: Wenige Tage zur Realisierung der Basis-Prototypen (geschätzt)	Gegeben: Wenige Wochen zur Realisierung der Basis-Prototypen	Gegeben: Wenige Monate zur Realisierung der Basis-Prototypen (geschätzt)

Kriterium	Kurze Laufzeit (ca. 1-2 Jahre)	Mittlere Laufzeit (ca. 3-10 Jahre)	Lange Laufzeit (theoretisch über 10 J.)
	Wenig Risiko im Einsatz von Technologien, Software und Standards	(geschätzt) Erhöhtes Risiko im Einsatz von Technologien, Software und Standards	Hohes Risiko im Einsatz von Technologien, Software und Standards
Betrieb	Gegeben: Persistente Datenspeicherung und Betriebssicherheit auf normalem Niveau	Gegeben: Persistente Langzeitdatenspeicherung und Betriebssicherheit auf hohem Niveau	Vermutlich gegeben: Persistente Langzeitdatenspeicherung und Betriebssicherheit auf höchstem Niveau
Schulung und Support	Gegeben: Nutzungszeit/Schulungsaufwand kann kritisch sein	Gegeben: Nutzungszeit „gerechtfertigt“ Schulungsaufwand	Gegeben: Nutzungszeit übersteigt deutlich den Schulungsaufwand
Administration	Gegeben: Normaler Pflege- und Supportbedarf	Gegeben: Hoher Pflege- und Supportbedarf	Gegeben: Höchster Pflege- und Supportbedarf
Weiterentwicklung	Gegeben: Normaler Entwicklungs- und Dokumentationsbedarf	Gegeben: Hoher Entwicklungs- und Dokumentationsbedarf	Gegeben: Höchster Entwicklungs- und Dokumentationsbedarf

6.1.5 Virtualisierung

Die Bewertung der Virtualisierung im „Rapid L-TER VRE“ Konzept erfolgte auf Grundlage der jeweiligen Anwendungsbeispiele (vgl. Kapitel 5). Für die Bewertung des Projektzeitraums wird eine theoretische Bewertung vorgenommen. Die Ergebnisse sind in Tabelle 6.7 zusammengestellt.

Insgesamt eignet sich das Konzept in der flexiblen Anwendung durch den Einsatz von Virtualisierungstechnologien zur dynamischen Skalierung der VRE-Infrastruktur in Abhängigkeit der projektspezifischen Rahmenbedingung in L-TER Projekten.

Tabelle 6.7: Bewertung der Virtualisierung im „Rapid L-TER VRE“ Konzept

Kriterium	Ohne Virtualisierung	Einfache Virtualisierung	Komplexe Virtualisierung
Projektgröße	Datenvolumen kleiner als mit Virtualisierung (Overhead)	VM Dateigröße relativ „mittel“ Basissystem im Gigabyte Bereich) + Datenspeicher bis zum Terrabyte Bereich	VM Dateigröße relativ „groß“ (Basissystem im Terrabyte Bereich) + Datenspeicher bis zum Terrabyte Bereich oder höher
Projektstruktur	Strukturierung weniger als mit Virtualisierung	VMs: ca. 4-10 Stk.	Relativ größere Anzahl an VMs, spezifische Technologien z.B. „LXC-Container“ für gemeinsamen VM „Overhead“

Kriterium	Ohne Virtualisierung	Einfache Virtualisierung	Komplexe Virtualisierung
Projektschwerpunkte	Keine Relevanz	Community Service („Joomla“ CMS mit „JomSoical“ Erweiterung) und VMs für Forschungsdatenmanagement vorhanden	(siehe „Einfache Virtualisierung“) Community-VM benötigt viel Systemperformance und Speicherplatz-
Projektzeitraum	Ggf. langsamere Realisierung (Prototypen, Basis-L-TER VREs), Administration (Backup, Updates), Weiterentwicklung (ohne „Snapshots“, Systemfehlerbeseitigung)	VMs benötigen mittlere dynamische Eigenschaften (Load balancing, Hardwareperformance, Speicherplatz, Backup, Archivierung)	VMs benötigen hohe dynamische Eigenschaften (Load balancing, Hardwareperformance, Speicherplatz, Backup, Archivierung)

6.2 Bewertung der Nachhaltigkeit

Die Bewertung der Nachhaltigkeit des „Rapid L-TER VRE“ Konzept erfolgte auf Grundlage der jeweiligen Anwendungsbeispiele (vgl. Kapitel 5). Da die Nachhaltigkeit im Rahmen dieser Arbeit nicht praktisch überprüft werden konnte (lange Projektlaufzeit), wird eine theoretische Bewertung vorgenommen. Die Ergebnisse wurden differenziert bewertet und sind in Tabelle 6.8 zusammengestellt:

Tabelle 6.8: Bewertung der Nachhaltigkeit des „Rapid L-TER VRE“ Konzept

Kriterium	Nachhaltigkeit (theoretisch)	Eignung
Persistenz	Insgesamt eignet sich das Konzept für eine nachhaltige Persistenz von L-TER VREs in L-TER Projekten. Dazu bedarf es zusätzlich zur Langzeitdatenspeicherung auf nachhaltige Formate, Standards und Infrastrukturen zu bauen und diese auch persistent mit den Forschungsdaten zu sichern und zu archivieren.	✓
Flexibilität	Insgesamt eignet sich das Konzept für flexible nachhaltige L-TER VREs in L-TER Projekten. Dazu bedarf es flexibler ICT, die sich auf das jeweilige Projekt spezifisch (Größe, Struktur, Schwerpunkte) anpassen lässt, und den Menschen sowie sein gesellschaftliches Umfeld einbezieht.	✓
ICT	Insgesamt eignet sich das Konzept durch nachhaltige ICT. Für die ICT sind allgemein alle der fünf Nachhaltigkeitskriterien für L-TER VREs (Persistenz, Konstanz, Resistenz, Elastizität, Resilienz) von erheblicher Bedeutung. Die ICT im Konzept ist hinsichtlich dieser Kriterien, insbesondere z.B. durch Virtualisierungstechnologien, dazu geeignet, nach großen „Störungen“ schnell wieder in den Zustand vor einer Störung zu gelangen. Der Einsatz von Open-Source-Software ist durch die Entwicklergemeinschaft nachhaltig und kostenlos in der Anschaffung. Die ICT ist modular aufgebaut, unterstützt SOA und nutzt nachhaltige Standards (techn.). Datenverbünde können nach „INSPIRE“ und individuell hierar-	✓

Kriterium	Nachhaltigkeit (theoretisch)	Eignung
	chisch aufgebaut werden. Innovative Technologien sind vermutlich in Zukunft in der Lage, die Nachhaltigkeit bedeutend zu steigern („DNA-Coding“, „BlockChain“).	
Mensch	Insgesamt trägt das Konzept flexibel für verschiedene Benutzer und Zielgruppen in der Nachhaltigkeit von L-TER Projekten bei. Dazu gehören: Standards (menschl.), Benutzerorientiertheit, Transparenz. Ebenso Dokumentation, Hilfe, Feedback, die auch im Team gepflegt werden können, z.B. durch Wikis.	✓

Das „Rapid L-TER VRE“ Konzept ist insgesamt in seinem Aufbau auf einen nachhaltigen Betrieb von LTER-VREs ausgelegt, wie das in dieser Arbeit vorgestellte Konzept für L-TER VREs (vgl. Kapitel 4) beschreibt.

Ein längerfristiger Betrieb der L-TER VREs im Sinne der Nachhaltigkeit ist durch die gewählten Open-Source-Produkte gegeben, da die Produkte eine große Entwicklergemeinschaft mit regelmäßigen Entwicklungszyklen und eine lange „Tradition“ haben. Sie deckten die Basisanforderungen der L-TER VRE zusammen mit den verfügbaren Erweiterungen ab.

Zur Nachhaltigkeit gehört auch die persistente Langzeitspeicherung von Daten. Allgemein kann davon ausgegangen werden, dass bei der Langzeitspeicherung von Daten viele Unsicherheitsfaktoren zu berücksichtigen sind. Auch die verwendeten Datenformate und zugehörige Software müssen persistent archiviert und dokumentiert werden. Ein Transfer von alten zu „zukünftigen“ Technologien auf Hard- und Softwareebene sollte rechtzeitig und kontinuierlich erfolgen, um die Persistenz der Daten zu gewährleisten. Das Konzept ist dazu geeignet und erläutert verschiedene Möglichkeiten hauptsächlich auf Server- (Betriebssystem, Software) und Hardware-Ebene, um eine Langzeitdatenspeicherung zu realisieren. Dazu tragen insbesondere auch VMs mit VM-Management-Software sowie allgemeine Backup Software bei, die in Verbindung mit geeigneter Hardware stehen. Generell scheint es eine geeignete Strategie zu sein, möglichst viele „Snap-Shots“ von unterschiedlichen Projektzeitpunkten zu erstellen und aktuelle Backups von gesamten VMs (Container-Datei), Dateisystemen und Datenbanken, an verteilten Orten, redundant aufzubewahren. Dies war für die Entwicklung der Prototypen in den Anwendungsbeispielen sehr Hilfreich bei der Wiederherstellung von früheren Entwicklungsstadien im Falle eines Systemfehlers durch unbeabsichtigte Änderungen oder inkompatible Updates.

6.2.1 Persistenz

Der Einfluss der Persistenz in Bezug zur Datennachhaltigkeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.9 differenziert bewertet und zusammengefasst. Zur Bewertung wird der derzeitige Stand der Technik (SdT) herangezogen. Mit „SdT“ gekennzeichnete Parameter erfüllen die geforderten persistenten Eigenschaften des L-TER VRE Konzeptes. Des Weiteren wird der Begriff „Nachhaltiger Stand der Technik“

(nSdT) eingeführt, der einen (besonders „+“) nachhaltigen SdT durch dessen spezifische Eigenschaften beschreibt.

Insgesamt eignet sich das Konzept für persistente und nachhaltige Speicherung von Daten in L-TER VREs bzw. deren Langzeitarchiven, die auch unabhängig von der restlichen L-TER VRE persistent bleiben müssen. Dazu bedarf es, zusätzlich zur Langzeitdatenspeicherung, auf nachhaltige Formate, Standards und Infrastrukturen zu bauen und diese auch persistent mit den Forschungsdaten zu sichern und zu archivieren. Das Langzeitarchiv der Daten ist letztendlich das, was über Generationen hinweg (auch nach Einstellung der aktiven Forschung in der übrigen L-TER VRE) erhalten werden muss.

Tabelle 6.9: Bewertung der Persistenz in Bezug zur Nachhaltigkeit im „Rapid L-TER VRE“ Konzept

Kriterium	Einfache Persistenz (SdT)	Mittlere Persistenz (nSdT)	Hohe Persistenz (nSdT+)
Daten	Auf dem „Stand der Technik“ (SdT): Backups, Datenkonsistenz, Strukturenkonsistenz, Datenbankkonsistenz	Auf einem nachhaltigen „Stand der Technik“ (nSdT): Archive, Fehlerkorrekturverfahren für Datenkonsistenz, Strukturenkonsistenz, Datenbankkonsistenz; Persistente Identifikatoren	(mehrfach) Redundanter nSdT
Langzeitdatenspeicherung	(SdT): Archivsicherheit; Archiv-Backups	(nSdT): Archivsicherheit: VM, Fehlerkorrekturverfahren	(nSdT+): (mehrfach) Redundanter nSdT z.B. auch Archivsicherheit bei: „Blockchain“-Technologien
Formate	(SdT): Beschreibung der Formate mit Backup	(nSdT): Beschreibung der Formate im Archiv	(nSdT+): (mehrfach) redundante Beschreibung der Formate mit z.B. mit „BlockChain“
Standards	(SdT): Beschreibung der Standards mit Backup	(nSdT): Beschreibung der Standards im Archiv	(nSdT+): (mehrfach) redundant z.B. mit „BlockChain“
Infrastruktur	(SdT): VM Backup der Infrastruktur	(nSdT): VM Backup der Infrastruktur („Snap-Shots“)	(nSdT+): (mehrfach) redundante VM Archivierung der Infrastruktur („SnapShots“)
Virtualisierung	(SdT): Backup der VM	(nSdT): Backup und Archivierung der VM	(nSdT+): Backup und Archivierung der VM

6.2.2 Flexibilität

Der Einfluss der projektspezifischen Flexibilität in Bezug zur Nachhaltigkeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.10 differenziert bewertet und zusammengefasst.

Insgesamt eignet sich das Konzept für flexible, nachhaltige L-TER VREs in L-TER Projekten. Dazu bedarf es flexibler ICT, die sich auf das jeweilige Projekt spezifisch (Größe, Struktur, Schwerpunkte) anpassen lässt.

Tabelle 6.10: Bewertung der Flexibilität in Bezug auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept

Kriterium	Einfache Flexibilität (SdT)	Mittlere Flexibilität (nSdT)	Hohe Flexibilität (nSdT+)
Größe	(SdT): ICT bietet persistente, konstante Eigenschaften. Konstante Nutzerzahl und Projekte.	(nSdT): ICT bietet persistente, dynamische Eigenschaften, die sich flexibel skalieren lassen. Dynamische Nutzerzahl.	(nSdT+): ICT bietet persistente, hochdynamische Eigenschaften, die sich hoch flexibel skalieren lassen. Hochdynamische Nutzerzahl.
Struktur	(SdT): ICT bietet persistente, konstant strukturierende Eigenschaften. Konstante Nutzer- und institutionelle Struktur.	(nSdT): ICT bietet persistente, dynamisch strukturierende Eigenschaften, die sich flexibel skalieren lassen. Dynamische Nutzerstruktur und institutionelle Struktur.	(nSdT+): ICT bietet persistente, hochdynamisch strukturierende Eigenschaften, die sich hoch flexibel skalieren lassen. Hochdynamische Nutzerstruktur und institutionelle Struktur.
Schwerpunkte	(SdT): ICT bietet persistente, konstante Schwerpunkte. Konstante Nutzer-, institutionelle- und gesellschaftliche Schwerpunkte.	(nSdT): ICT bietet persistente, dynamische Schwerpunkte, die sich flexibel skalieren lassen. Dynamische Nutzer-, institutionelle und gesellschaftliche Schwerpunkte.	(nSdT+): ICT bietet persistente, hochdynamische Schwerpunkte, die sich hoch flexibel skalieren lassen. Hochdynamische Nutzer-, institutionelle und gesellschaftliche Schwerpunkte.
Virtualisierung	(SdT): ICT bietet konstante VMs.	(nSdT): ICT bietet dynamische VMs, die sich flexibel skalieren lassen.	(nSdT+): ICT bietet „hochdynamische“ VMs, die sich flexibel skalieren lassen.

6.2.3 ICT

Der Einfluss der ICT auf die technische Nachhaltigkeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.11 differenziert bewertet und zusammengefasst.

Insgesamt eignet sich das Konzept flexibel mit „einfacher“ bis „hoher“ Nachhaltigkeit in L-TER Projekten:

Tabelle 6.11: Bewertung der ICT in Bezug auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept

Kriterium	Einfache ICT	Nachhaltige ICT	Hoch nachhaltige ICT
Open Source	(SdT): kostenlos in Anschaffung	(nSdT): Weiterentwicklung durch Open-Source-Community	(nSdT+): L-TER VRE selbst als Open Source Framework
Infrastruktur	(SdT): ohne VM mehr Hardwarebedarf	(nSdT): Mit VM weniger Hardwarebedarf, skalierbar	(nSdT+): Mit VM, skalierbar, „LXC Container“
Systemarchitektur	(SdT): modularer Aufbau	(nSdT): modularer Aufbau, SOA, ECC RAM-Speicher (Fehlerkorrektur)	(nSdT+)
Standards (techn.)	(SdT): offene Standards, Spezifische Formate, z.B. PDF-A	(nSdT): Internationale Standards, Web 2.0, Web 2.5, Web 3.0, Spezifische Formate: z.B.: ISO 9660 für ISO/VM-Images mit Fehlerkorrektur, Spezifische Standards z.B. Fehlerkorrekturverfahren	(nSdT+): Weltweite Standards, Web X? „Blockchain“-Technologien? „DNA-Codierung“?
Schnittstellen	(SdT):	(nSdT): Tensoren	(nSdT+)
Datenverbünde	(SdT): „Geo-Network“: projektinterne Netzwerke, flache Hierarchien	(nSdT) „Geo-Network“: INSPIRE, projektübergreifende, fachliche-Netzwerke, mittlere Hierarchien	(nSdT+) „Geo-Network“: INSPIRE, Individuell, öffentliche-Netzwerke, tiefe Hierarchien
Virtualisierung	(SdT): Einfluss auf alle 5 Kriterien der Nachhaltigkeit (vgl. Kap 3.1): Backup, Flexibilität	(nSdT): Deutlicher Einfluss auf alle 5 Kriterien der Nachhaltigkeit: Backup, Flexibilität, „SnapShots“, Entwicklung, schnelle Wiederherstellung nach Störungen, Systemskalierung	(nSdT+): Erheblicher Einfluss auf alle 5 Kriterien der Nachhaltigkeit: Backup, Archivierung, Flexibilität, „SnapShots“, Weiterentwicklung, schnelle Wiederherstellung nach großen Störungen, dynamische Systemskalierung, LXC Container

6.2.4 Faktor „Mensch“

Der Einfluss des Faktor „Mensch“ mit seinem sozialen und gesellschaftlichem Umfeld auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.12 differenziert bewertet und zusammengefasst. Dabei handelt es sich um eine nicht konkret untersuchte Einschätzung.

Insgesamt trägt das Konzept flexibel für verschiedene Benutzer und Zielgruppen in der Nachhaltigkeit von L-TER Projekten bei.

Tabelle 6.12: Bewertung des Menschen (und Gesellschaft) in Bezug auf die Nachhaltigkeit im „Rapid L-TER VRE“ Konzept

Kriterium	Mensch (isoliert)	Community	Öffentlichkeit
Standards (gesellschaftl.)	geeignet: regional, national: Sprache, Zeit, Gesetze, etc.	geeignet: national, international: Sprache, Zeit, Jahreszeit, Gesetze, etc.	geeignet: weltweit: Sprache, Zeit, Jahreszeit, Gesetze, etc.
Benutzerorientiert	geeignet	geeignet: Community-ICT	geeignet: Homepage
Transparenz	geeignet	geeignet: Nutzerakzeptanz durch Daten, Datenschutz, Dokumentation, Hilfe, Data Policies, Hintergrundinformationen, Rechtsformen, Institutionen, Struktur	geeignet: Akzeptanz über die Projektteilnehmer hinaus „gesicherte“ Transparenz und Vertrauen, durch „BlockChain“?
Dokumentation	geeignet: Dokumentation kann nur von Teilgruppe gepflegt werden	geeignet: Dokumentation kann von allen Projektteilnehmern gepflegt werden	geeignet: Dokumentation kann von allen Projektteilnehmern und darüber hinaus gepflegt werden
Hilfe	geeignet: Hilfe kann nur von Teilgruppe gepflegt werden	geeignet: Hilfe kann von allen Projektteilnehmern gepflegt werden	geeignet: Hilfe kann von allen Projektteilnehmern und darüber hinaus gepflegt werden
Feedback	geeignet: eingeschränkte Feedbacks mit geringerer Akzeptanz und zur Evaluation sowie Weiterentwicklung	geeignet: projektinterne Feedbacks für mittlere Akzeptanz und zur Evaluation sowie Weiterentwicklung	geeignet: öffentliche Feedbacks für hohe Akzeptanz und zur Evaluation sowie Weiterentwicklung
Zielgruppen	geeignet: L-TER VRE ist von besonderem Interesse für eine spezielle Teilgruppe des Projektes, ggf. „vertraulich“ oder „geheim“.	geeignet: L-TER VRE ist von Interesse für das gesamte Projekt und ggf. in den Fachdisziplinen	geeignet: L-TER VRE ist von öffentlichem Interesse

6.3 Bewertung der Entwicklungszeit

Per Definition ist das „Rapid L-TER VRE Konzept“ für eine „rapide“ und nachhaltige Entwicklung von Prototypen mit einer „schnellen“ Übergangsphase in den Produktivbetrieb ausgelegt.

Um das „Rapid L-TER VRE“ Baukastenkonzept hinsichtlich einer „schnellen“ Realisierung zu überprüfen, wurden die Entwicklungszeit für Basis-VRE (Prototyp) und die geschätzte Entwicklungsdauer für eine erweiterte L-TER VRE als zusätzliche Kriterien verwendet.

Insgesamt eignet sich das Konzept zur „schnellen“ (Weiter-)Entwicklung von L-TER VRE Prototypen, sowie Basis- und erweiterten L-TER VREs im Produktivbetrieb für flexible L-TER Projekte.

Basis-Prototypen für L-TER VREs können innerhalb weniger Tage und Wochen realisiert werden.

Im Vergleich dazu wird für z.B. für komplette Eigenentwicklungen ein Zeitbedarf von Monaten bis Jahren geschätzt.

Tabelle 6.13: Bewertung der „schnellen“ Entwicklung im „Rapid L-TER VRE“ Konzept

Kriterium	Entwicklungszeit (theoretisch)	Eignung
Virtualisierung	Erhebliche Zeitersparnis durch die Eigenschaften von VMs, wie z.B. „Snapshots von lauffähigen Systemumgebungen und in Verbindung mit „schlüselfertigen“ Systemen	✓
„Schlüselfertige“ Systeme	Erhebliche Zeitersparnis durch vorinstallierte und vorkonfigurierte Systeme als Image	✓
Programmieraufwand	Programmieraufwand ist erheblich geringer durch den Einsatz von Open-Source-Software, „schlüselfertige“ Systeme, Modularisierung, SOA. Individueller Programmieraufwand für: fachspezifische Schwerpunkte, Automatisierung von Berichten, Systemübergreifende Tools, SSO	✓
Weiterentwicklung	Basis-VRE: relativ schnell, erweiterte VRE: individuell	✓
Baukasten	Umfangreicher Baukasten mit Service-Kategorien, Tools und ausgewählten Softwarebeispielen verkürzt die Entwicklungszeit erheblich.	✓

Die Anwendung des Konzeptes kann nicht nur die Implementation der VRE, sondern auch die Planung erheblich beschleunigt. Den größten Anteil tragen dazu die im Konzept vorgestellten Methoden für den Einsatz von „schlüselfertigen“ Servern und deren Open-Source-Software für einzelne VRE-Services und -Tools in Verbindung mit VMs und VM-Management-Systemen (vgl. Kapitel 4.3) bei. Für die Administration sind diese Technologien ebenfalls sehr hilfreich. Durch die relativ einfache Handhabung der VMs kann allgemein viel Zeit im Servermanagement gespart werden. Zum Beispiel lassen sich verschiedene Entwicklungsschritte bzw. Backups der VRE auf Knopfdruck sichern und wiederherstellen.

Weiterhin können durch die Anwendung des Konzeptes zeitintensivere Implementierungen von speziellen Funktionalitäten identifiziert und für spätere Entwicklungsschritte zurückgestellt werden, damit möglichst schnell zumindest eine Basis-VRE in den Produktivbetrieb übergehen kann. Dies ist überall dort sinnvoll, wo eigene Programmierung und Speziallösungen erforderlich sind (vermehrt bei den VRE-Tools und im Report-Layer) oder die Konfiguration sehr komplex sein kann, wie im Falle einer zentralen Authentifizierung mit SSO.

6.3.1 Virtualisierung

Der Einfluss der Virtualisierung auf die Entwicklungszeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.14 differenziert bewertet und zusammengefasst.

Insgesamt eignet sich die das Konzept durch Virtualisierung für eine „schnelle“ Entwicklung flexibler L-TER Projekte. Die Zeitersparnis in der Entwicklung ist erheblich gegenüber kompletten „Eigenentwicklungen“.

Tabelle 6.14: Bewertung der Zeitersparnis durch „Virtualisierung“

L-TER VRE	Durchschnittlicher Zeitbedarf / Server und Technologie (geschätzt)	
	Mit Virtualisierung	Ohne Virtualisierung
Installation	Im Bereich von ca. 10 Minuten pro VM	mehrere Stunden pro Server, falls ein Image geschrieben wird ggf. ähnlich schnell
Administration	z.B. komplettes Server-Backup mit Snapshot im Bereich von 10 Minuten bei Basis-System pro VM	im Bereich von Stunden pro Server
Dynamische Skalierung	der Hardware-Performance (CPUs, RAM, Speicher) wenige Minuten im Idealfall	Tage
Entwicklung	Die Zeitersparnis ist sehr individuell: VMs unterstützen zur Entwicklung notwendige: Installation, Administration und dynamische Skalierung. „SnapShots“ von VMs unterstützen die Entwicklungsgeschwindigkeit erheblich durch die Speicherung der Entwicklungsschritte. Zudem sind VMs aufgrund ihrer Eigenschaften mit einer lauffähigen Umgebung der gesamten Infrastruktur eine gute Basis für die Weiterentwicklung in einem Entwicklerteam oder einer offenen Community.	Tage bis Jahre

6.3.2 „Schlüselfertige“ Systeme

Der Einfluss von „schlüselfertigen“ Systeme auf die Entwicklungszeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.15 differenziert bewertet und zusammengefasst.

Insgesamt eignet sich die das Konzept durch „schlüselfertige“ Systeme für eine „schnelle“ Entwicklung flexibler L-TER Projekte. Die Zeitersparnis in der Entwicklung ist erheblich gegenüber kompletten „Eigenentwicklungen“.

Tabelle 6.15: Bewertung der Zeitersparnis durch „schlüsselfertige“ Systeme

Phase der VRE	Durchschnittlicher Zeitbedarf / Server und Technologie (geschätzt)	
	„Schlüsselfertige“ VMs	Ohne „schlüsselfertige“ VMs
Planung	Individuelle Zeitverkürzung durch Testinstallationen und durch die Beispielsoftware im „Baukasten“ des Konzeptes	Tage bis Monate
Installation	Im Bereich von 10 Minuten pro VM	Stunden bis Wochen
Konfiguration	Individuelle Zeitverkürzung durch Vor-Konfiguration und -Installation	Minuten bis Wochen

6.3.3 Programmieraufwand

Im Vergleich zu kompletten Eigenentwicklungen (Programmieraufwand) von VREs, die durchaus Monate bis mehrere Jahre dauern können, wird der Programmieraufwand durch die Anwendung des „Rapid L-TER VRE“ Baukastenkonzeptes erheblich reduziert und dadurch der Entwicklungszeitraum für L-TER VREs verkürzt (Abschätzung). Der Einfluss des Programmieraufwandes auf die Entwicklungszeit im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.16 differenziert bewertet und zusammengefasst.

Insgesamt eignet sich das Konzept durch die Reduzierung des „eigenen“ Programmieraufwandes für eine „schnelle“ Entwicklung flexibler L-TER Projekte. Die Zeiterparnis in der Programmierung ist erheblich gegenüber kompletten „Eigenentwicklungen“.

Tabelle 6.16: Bewertung des Programmieraufwandes für das „Rapid L-TER VRE“ Konzept

Konzeptbestandteil	Einfluss auf Programmieraufwand	Erforderlicher Aufwand (geschätzt)
Open Source	Der wesentliche Teil an Programmierung entfällt durch Open Source Community. Offener Quellcode als Basis für L-TER VRE, individuelle Änderungen und Weiterentwicklungen.	gering
Offene Standards	Standards müssen nicht selbst entwickelt werden	gering
Modularität	Modularer Aufbau der eingesetzten Komponenten ermöglicht als Basis die Einbindung von Erweiterungen	gering
SOA	Serviceorientierte Architekturen vereinfachen Einbindung, Austausch und Wiederverwendung von VRE-Services	gering
VRE-Services	Basis VRE-Services werden abgedeckt durch Baukasten. Die Services basieren in der Regel auf Open-Source-Software.	gering
VRE-Tools	Basis VRE-Tools werden großteils abgedeckt durch Baukasten. Die Services basieren in der Regel auf Open-Source-Software.	gering-mittel

Konzeptbestandteil	Einfluss auf Programmieraufwand	Erforderlicher Aufwand (geschätzt)
Authentifizierung und SSO	Authentifizierung mittels Verzeichnisserver (LDAP) für Basis-VREs ist schnell zu realisieren. Ein VRE-übergreifendes SSO wird nicht von allen Komponenten gleichermaßen unterstützt. Programmierung und Konfiguration ist mit spezifischem Know-how verbunden.	gering-mittel
Report-Layer	Für viele Report-Funktionen können die VRE-Tools der spezifischen VRE-Services auf Basis von Open-Source-Software zum Einsatz kommen. Für die Realisierung eines übergreifenden Report-Layers kann das CMS konfiguriert und programmiert werden. Eine individuelle Generierung von Berichten, für verschiedene Zielgruppen aufbereitet, erfordert meist eigenen Programmieraufwand. Für interoperable Datenverbünde ist die Software „Geo-Network Open Source“ nach relativ wenig Konfigurationsaufwand einsatzbereit.	gering-mittel
Fachspezifische Schwerpunkte	Hier ist meist eigener Programmieraufwand erforderlich, um VRE-Services und –Tools individuell anzupassen, und zu entwickeln.	mittel-hoch

6.3.4 Weiterentwicklung

Der Zeitbedarf der Weiterentwicklung im „Rapid L-TER VRE“ Konzept wird in Tabelle 6.17 differenziert bewertet und zusammengefasst.

Insgesamt eignet sich das Konzept durch die nachhaltige Planung der Entwicklung und Weiterentwicklung für eine „schnelle“ Realisierung flexibler L-TER Projekte. Die Zeitersparnis in der (Weiter-)Entwicklung ist erheblich gegenüber einer nicht nachhaltigen Planung.

Tabelle 6.17: Bewertung der Weiterentwicklung im Rahmen des „Rapid L-TER VRE“ Konzeptes

Konzeptbestandteil	Einfluss auf Weiterentwicklung	Erforderlicher Aufwand (geschätzt)
Aufteilung	Aufteilung in Basis- und erweiterte L-TER VRE, Abschätzung des Aufwandes bei der Auswahl der Basisfunktionalitäten und erweiterten Funktionalitäten, L-TER VRE als Basis für Erweiterungen, innovative Technologien, fortwährender Entwicklungsprozess	Basis: gering Erweitert: gering-hoch
Nachhaltigkeit	Kontinuierliche Weiterentwicklung nach Konzept	mittel-hoch
Open Source	Große Open-Source-Community, L-TER VRE selbst als Open-Source-Framework (hier ist individueller Aufwand/Nutzen schwer abzuschätzen)	gering

Konzeptbestandteil	Einfluss auf Weiterentwicklung	Erforderlicher Aufwand (geschätzt)
Offene Standards	Sind generell vorhanden und müssen eingebunden und ggf. individuell erweitert werden	gering-mittel
Modularität	Viele Erweiterungen verfügbar, die eingebunden und ggf. individuell erweitert werden müssen	gering
SOA	Viele Architekturen vorhanden, die ggf. für eigene Services zu zusätzlichem Programmieraufwand führen. Erweiterung, Austausch und Wiederverwendung von VRE-Services.	gering-mittel
VRE-Services	Viele „erweiterte“ VRE-Services im Baukasten.	gering
VRE-Tools	Viele „erweiterte“ VRE-Tools im Baukasten. Meist müssen individuelle fachspezifische Tools selbst weiterentwickelt werden.	gering-mittel
SSO	Es sollte darauf geachtet werden, dass neue Komponenten über eine zentrale Authentifizierung mit Verzeichnisservern (LDAP) und SSO verfügen.	gering-mittel
Report-Layer	Weiterentwicklung zu zielgruppenorientierten, automatisch generierten Berichten.	mittel
Fachspezifische Schwerpunkte	Hier ist meist eigener Programmieraufwand erforderlich, um VRE-Services und -Tools individuell und innovativ weiterzuentwickeln.	mittel-hoch

Folgende Aspekte ergeben sich aus den Projektbeispielen:

- Updates: Es ist hilfreich, parallel eine zweite Testumgebung zu betreiben, um den laufenden Betrieb nicht zu gefährden. Dies wird durch die Virtualisierungstechnologien stark vereinfacht.
- Häufige Updates der verschiedenen Module, Services und Tools erschweren die Arbeit der Administratoren. Aber eine einzige Software, die alles kann, existiert ohnehin nicht.
- Die Kompatibilität von einzelnen Softwareprodukten und deren Erweiterungen (mit verschiedenen Versionen) ist problematisch. Insbesondere bei CMS mit den unterschiedlichen Erweiterungen, die oft nicht mit der Entwicklung des CMS Schritt halten.

6.3.5 Baukasten für VRE-Services und -Tools mit Softwareempfehlungen

Der Baukasten diene als Basis für individuelle L-TER VREs mit der Einteilung von VRE-Services nach ihrer Funktion und mit relevanten Softwarebeispielen zu VRE-Services und -Tools. Er kann erheblich Zeit bei der Suche nach geeigneter Software einsparen (Tabelle 6.18).

Insgesamt eignet sich Konzept durch den integrierten „Baukasten“ mit Kategorien, VRE-Services und -Tools, sowie Software-Quellen (Anhang, Quellenverzeichnis) mit zusätzliche Erweiterungen für eine „schnelle“ Entwicklung flexibler L-TER Projekte auf der Basis von Open-Source-Software. Die Zeitersparnis in der Planung und Entwicklung ist erheblich gegenüber der „uninformierten“ Suche von VRE-relevanter, einzelner und kompatibler Softwareprodukte aus einem Pool von tausenden Produkten und deren oft erheblichen Anzahl an Erweiterungen.

Tabelle 6.18: Bewertung des „VRE-Baukastens“ in Bezug zur Entwicklungszeit im „Rapid L-TER VRE“ Baukastenkonzept

Baukasten	Einfluss auf Entwicklungsgeschwindigkeit	Erforderlicher Aufwand (geschätzt)
Einteilung von VRE-Services	Der wesentliche Teil an Programmierung entfällt durch Open Source Community. Offener Quellcode als Basis für L-TER VRE, individuelle Änderungen und Weiterentwicklungen.	gering
Softwarebeispiele: VRE-Service	Standards müssen nicht selbst entwickelt werden	gering
Softwarebeispiele: VRE-Tools	Modularer Aufbau der eingesetzten Komponenten ermöglicht als Basis die Einbindung von Erweiterungen	gering

6.4 Bewertung der Konzeptbestandteile

Die differenzierten Bewertungen der 10 Konzeptbestandteile im „Rapid L-TER VRE“ Konzept in Bezug auf Nachhaltigkeit und Entwicklungszeit sind in Tabelle 6.19 zusammengefasst.

Insgesamt eignet sich die das Konzept durch die einzelnen Bestandteile (10 Punkte) zur „schnellen“ Entwicklung flexibler L-TER Projekte. Die Zeitersparnis durch die Anwendung des Konzeptes ist erheblich gegenüber anderen Konzepten.

Tabelle 6.19: Bewertung der Konzeptbestandteile im „Rapid L-TER VRE“ Konzept

Bestandteil	Nachhaltigkeit	Eignung	Entwicklungszeit	Eignung
Open-Source	Entwickler-Community, Faktor „Mensch“, Benutzerorientierung durch Nutzer-Community impliziert	✓	Offener Quellcode	✓
ICT	Virtualisierung, Langzeitdatenspeicherung, Standards (techn.)	✓	Übersicht zu Technologien, Technologien in Software verankert	✓
„Schlüsselfertige“ Systeme	Einheitliche Betriebssystem-Basis, Virtualisierung	✓	Installation und Konfiguration	✓
Homepage, CMS	Modularität, Funktionalitäten des CMS	✓	Basis der VRE und Content-Management	✓

Bestandteil	Nachhaltigkeit	Eignung	Entwicklungszeit	Eignung
VRE-Services	SOA	✓	Baukasten, Basis-Services (vgl. Open Source)	✓
VRE-Tools	Modularität	✓	Baukasten, Basis-Tools (vgl. Open Source), ggf. individuelle Entwicklung von Tools	✓
Auth., SSO	Zentrale Authentifizierung, Akzeptanz durch SSO	✓	(vgl. ICT), Konfigurationsaufwand, Know-how erforderlich	✓
Report-Layer	zielgruppenorientiert, Datenverbünde	✓	Basis Report-Tools, Programmierung für automatische Berichterzeugung und Layer, Datenverbünde schnell INSPIRE-konform realisierbar mit „GeoNetwork“	✓
Hilfe, Feedback, Doku	zielgruppenorientiert	✓	(vgl. Services, Tools)	✓
Weiterentwicklung	Einteilung in Basis- und erweiterte VRE, Kontinuität, innovative ICT, nachhaltiges Konzept	✓	schnelle Realisierung von Prototypen durch Aufteilung in Basis- und erweiterte VRE, Aufwandsabschätzung	✓

6.4.1 Open Source

Zum Aufbau der VRE ließen sich verschiedene geeignete Open-Source-Software-Produkte finden und integrieren. Die Beispiele aus dem „VRE-Baukasten“ (vgl. Anhang) waren dabei eine hilfreiche Unterstützung auf der Suche nach geeigneter Software. Ein längerfristiger Betrieb der L-TER VREs im Sinne der Nachhaltigkeit scheint durch die gewählten Open-Source-Produkte gegeben zu sein, da die Produkte eine große Entwicklergemeinschaft mit regelmäßigen Entwicklungszyklen und eine lange Tradition haben. Sie decken weitgehend die Basisanforderungen von L-TER VREs zusammen mit den verfügbaren Erweiterungen ab.

6.4.2 ICT-Basistechnologien

Die gewählte webbasierte Informationsstruktur, die auf Webtechnologien und offenen Standards aufbaut, scheint nachhaltig geeignet für L-TER VREs zu sein, wie aus den Anwendungsbeispielen (vgl. Kapitel 5) und der Einschätzung der Nachhaltigkeit eingesetzter ICT (vgl. Kapitel 6.2.3) hervorgeht. Mit einer Übersicht zu den Technologien war dieser Punkt hilfreich, um die benötigten VRE-Services und -Tools umzusetzen und

dazu die geeignete Software und die erforderlichen Standards auszuwählen. Von besonderer Bedeutung waren die folgenden Technologien und Standards:

- Webtechnologien: Web 2.0, 2.5, 3.0, HTML,
- Content-Management-Systeme,
- Technologien zur Virtualisierung,
- Datenbankmanagementsysteme,
- Technologien zur Authentifizierung und SSO,
- Datenmodelle und Semantic Web (Web 3.0),
- Serviceorientierte Architekturen (SOA) (Web 2.5),
- Datenverbünde für INSPIRE-konformen Datenaustausch in Netzwerken,
- Langzeitdatenspeicherung (z.B. Scriptgesteuerte Archivierung, PDF-A),
- Remote Desktop, Screen- und Application Sharing,
- Hochleistungsrechnen (HPC) und verteiltes Rechnen (GRID)
(optional, nicht in den Anwendungsbeispielen).

6.4.3 Schlüsselfertige Server und Virtualisierung

Schlüsselfertige Server erwiesen sich in Verbindung mit Technologien zur Virtualisierung als eine erhebliche Unterstützung, um Zeit und Hardwareressourcen einzusparen. Sie waren sehr hilfreich zum Test von Software, für Prototypen und sind geeignet für den Produktivbetrieb und die Weiterentwicklung – somit in jeder Lebensphase der L-TER VRE. Für viele der benötigten VRE-Services waren „schlüsselfertige“ Server in der Liste aus dem Anhang F zu finden. Die durchschnittlich benötigte Zeit für den Download und die Installation der Server, sowie die Erstellung der VMs lag bei ca. 15 Minuten pro Server. Im Vergleich dazu benötigt die Installation auf normalem Wege mehrere Stunden oder Tage und erfordert tiefgründiges Wissen z.B. zu Linux-Serversystemen. Bei der Konfiguration der eigentlichen Software, die speziell für den gedachten VRE-Service anfällt, ist der Geschwindigkeitsvorteil meist nicht ganz so groß. Zwar ist hier schon viel Arbeit im Vorfeld durch eine Grundkonfiguration oder bereits installierte Erweiterungen abgenommen – jedoch müssen die Einstellungen überprüft werden und die Konfiguration ist immer mit einer individuellen (manuellen) Anpassung der L-TER VRE verbunden.

6.4.4 Homepage und CMS

Mit Hilfe dieses Konzeptbestandteils können solide CMS als Basis-Systeme für die L-TER Projekte implementiert werden. Es hat sich gezeigt, dass ein CMS mit Homepage eine gute Basis für den Aufbau einer VRE darstellt, da viele erforderliche Eigenschaften und Funktionen bereits integriert sind und es dazu flexibel in der Anwendung ist. Die Bedeutung des CMS mit Homepage für L-TER VREs ist durch diesen Punkt im Konzept relevant hervorgehoben. Das „Rapid L-TER VRE“ Konzept ist auch dabei behilf-

lich, die Auswahl für ein geeignetes CMS zum Projekt zu treffen. Durch die Beschreibung der einzelnen CMS (vgl. Anhang C, Tabelle C.5) und die Literatur (Breitenstrom et al. 2013) ist die Auswahl eines geeigneten CMS für das L-TER Projekt gut möglich. Die Liste der Erweiterungen (vgl. Anhang D) ist hilfreich, um nicht den Überblick bei den verfügbaren Erweiterungen auf der Homepage des CMS-Anbieters zu verlieren. Manche Erweiterungen sind kostenpflichtig. Jedoch rentiert sich oft ihr Einsatz, da die Preise verhältnismäßig niedrig sind und es sich um Software-Module handelt, die nur schwer oder zeitaufwendig selbst zu programmieren und ggf. relativ leicht austauschbar sind. Da es sich bei diesen kleinen Erweiterungen um keine wesentlichen Hauptbestandteile der L-TER VRE handelt (wie das CMS selbst), kann aus den vorangegangenen Gründen vom Open-Source-Konzept abgewichen werden. Die Einbindung der VRE-Services und -Tools in der L-TER VRE erfolgt durch das Setzen der jeweiligen Links und die Verwendung von „iFrames“ über das CMS.

6.4.5 VRE-Services

Dieser Bestandteil des „Rapid-L-TER VRE“ Konzeptes repräsentiert die Basisfunktionalitäten in Form von VRE-Services in einer L-TER VRE. Durch die Anwendung des Konzeptes können relevante VRE-Services leicht aus dem „VRE-Baukasten“ ausgewählt und durch konkrete Produktempfehlungen schnell implementiert werden. „Schlüsselfertige“ Systeme und VMs für die einzelnen VRE-Services haben zu einer schnellen Realisierung des Prototyps beigetragen. Die im Konzept vorgeschlagenen VRE-Services erfüllen die Basisanforderungen und Rahmenbedingungen der L-TER Projekte. Für die zusätzlichen Funktionalitäten einer erweiterten VRE können durch eine modularisierte Architektur und SOA zusätzliche VRE-Services und spezifische VRE-Tools individuell mit verringertem Aufwand ergänzt werden.

Als „Basis-Service“ der L-TER VREs sind besonders folgende Services hervorzuheben: „CMS-Service“, „Datenbank-Service“, „Metadatenmanagement-Service“, „Harvesting-Service“, „Wiki-Service“ und „Community-Service“. Auf diese wird im Folgenden eingegangen:

Die Bewertung des „CMS-Services“ erfolgte bereits als zentraler Bestandteil des Konzeptes.

Der Service für das *Metadatenmanagement* und dem maschinellen Austausch von Metadaten (*Harvesting-Service*) kann mit der Software „GeoNetwork Open Source“ realisiert werden. Durch den integrierten Map-Server lassen sich auch GIS-Daten visualisieren. „GeoNetwork“ spielt in L-TER VREs in Bezug auf die Metadaten der Forschungsdaten eine zentrale Rolle. So können für L-TER VREs beispielsweise leicht und schnell INSPIRE-konforme Datenverbünde für den Umweltbereich realisiert werden. Metadaten, die aus anderen Services oder Systemen stammen, lassen sich von „GeoNetwork“ als Metadatenmanagementsystem importieren. Im Rahmen einer erweiterten L-TER VRE muss jedoch darauf geachtet werden, dass die Verknüpfung von Metadaten

aus anderen VRE-Services mit „GeoNetwork“ automatisiert durch VRE-Tools verbunden sind.

Als Software für den *Wiki-Service* kann „MediaWiki“ eingesetzt werden, für das ebenso zahlreiche Erweiterungen mit zusätzlichen Funktionalitäten zur Verfügung stehen. Darunter ist insbesondere die Erweiterung „Visual-Editor“ hervorzuheben, welche die Möglichkeit bietet, einen WYSIWYG-Editor für eine vereinfachte Anwendung bereitzustellen. Dadurch sind z.B. Formatierungen übersichtlicher vorzunehmen sowie Formeln, Zitate, geografische Karten, Bilder und Videos direkt in Texte einzubinden. Das gilt ebenso bei „Goolge Charts“ für die Erzeugung und Einbettung von Diagrammen, sowie bei „R“ für die Ausführung interaktiver „R-Skripte“ mit „R-Server“. Insgesamt ist das Wiki sehr flexibel als „Allround-Tool“ einsetzbar.

Für den *Community-Service* (Social Media, Social Networks) hat sich die CMS-Erweiterung „JomSocial“ bewährt (kostenpflichtige Lizenz), deren Funktionalitäten sich grundlegend mit denen von „Facebook“ vergleichen lassen. Sie erweitert den CMS-Service zum Community-Service. Für „Joom Social“ existieren zahlreiche Erweiterungen, mit denen individuelle Funktionalitäten realisiert werden können. Dazu gehörten Funktionalitäten des Community-Services auf der Basis von Web 2.0 z.B.: dynamische Gruppen mit Gruppenhomepage, Benutzerprofile mit Benutzerhomepage, Terminkalender, Timelines, Social Media (Einbetten und Teilen von multimedialen Inhalten) und Bewertungen.

6.4.6 VRE-Tools

Die meisten Funktionalitäten einer Basis L-TER VRE können durch die bereits vorhandenen oder zusätzlich installierten VRE-Tools der einzelnen VRE-Services realisiert werden. Bei individuellen Aufgaben müssen in einer erweiterten L-TER VRE selbstprogrammierte Tools in die VRE eingebunden sein. Dies gilt z.B. für den Report-Layer, der in der Basis-VRE für das Projekt aus zeitlichen Gründen nur eingeschränkt ohne Funktionen zur automatisierten Generierung von Berichten implementiert wurde. Insbesondere betrifft dies auch die individuellen Methoden zur Datenanalyse mit Pre- und Postprocessing. Sie sind sehr zeitaufwendig in der Programmierung oder bei der Integration von Statistik-Services („R-Statistics“) mit den Daten der VRE.

Zu den Aufgaben der VRE-Tools gehört auch die Langzeitdatenspeicherung, die für L-TER VREs per Definition von besonderer Bedeutung ist. Im Rahmen dieser Arbeit konnte aus zeitlichen Gründen keine Evaluation über einen längeren Zeitraum erfolgen. Zur Langzeitspeicherung von Daten können unterschiedliche Hard- und Softwaretechniken individuell genutzt werden (vgl. Kapitel 2.5.4). Verschiedene eigene Skripte zur Automatisierung von serverseitiger Backup- und Virtualisierungssoftware wurden dazu erfolgreich getestet (Anhang E). Darunter Skripte zum Kopieren von Daten mit Funktionen des Betriebssystems, Sichern der Datenbanken mit z.B. „phpmyadmin“ und CMS-Tools, sowie Skripte zum automatischen Start/Stop der VMs und der Erstellung von

VM-Snapshots mit der Übertragung der Snapshots und Containerdateien auf einen Netzwerkspeicher. So kann eine persistente Langzeitdatenhaltung realisiert werden. Die Maßnahmen richten sich dabei nach dem jeweiligen Sicherheitsanspruch und dem Zeitrahmen.

6.4.7 Authentifizierung

Für die L-TER Projekt kann ein „OpenLDAP“ Service (Verzeichnisserver) zur zentralen Authentifizierung zum Einsatz kommen. Die eingesetzten Produkte für die relevanten VRE-Services und deren Erweiterungen unterstützen diese Technologie und den Standard weitgehend. Das CMS und die verschiedenen VRE-Services können dadurch erfolgreich in der L-TER VRE mit einer zentralen Authentifizierung eingebunden werden. Auch bei einem kleinen Nutzerkreis sollte zur Verbindung der Nutzerdatenbanken ein LDAP Verzeichnisserver eingesetzt werden, da ansonsten jeder Service eine eigene Nutzerverwaltung verwendet (mehrfacher administrativer Aufwand) und daher eine zentrale Authentifizierung für die einzelnen Services kaum möglich ist.

Aufwendiger ist in einer erweiterten L-TER VRE die Umsetzung von SSO für das gesamte System übergreifend zu realisieren, da es nicht von allen potentiell relevanten VRE-Services gleichermaßen unterstützt wird. Für die Basis-VREs kann daher ein „pseudo SSO“ mit Hilfe von Cookies und der Funktion „angemeldet bleiben“ eingerichtet werden. Dies wurde von allen Basis-Services unterstützt. Gleichzeitig stellt SSO jedoch eine Grundeigenschaft von VREs dar, sodass dies einen kritischer Punkt für den Gesamterfolg einer VRE sein kann.

6.4.8 Report-Layer

Für den Report-Layer werden die Report-Tools des CMS und der jeweiligen VRE-Services genutzt, die standardmäßig oder als Erweiterung zur Verfügung stehen. Sie liefern die gewünschten service-spezifischen Kennzahlen z.B. über Forschungsdaten, Benutzer und deren Beiträgen. Die Kennzahlen können passend zum Kontext für Administratoren und Benutzer angeordnet sein. Für erweiterte L-TER VRE lassen sich die einzelnen Report-Tools in einem übergeordneten Report-Layer des CMS zusammenfassen. Diese Kennzahlen dienen der allgemeinen Information und sind z.B. entscheidende Bewertungskriterien für den Erfolg von VREs.

Ebenso ist die automatische Generierung von individuellen Berichten mit Programmieraufwand möglich. Dazu eignen sich Tools zur Erzeugung von Templates, die mit einer programmierten Logik die Daten der Datenbanken und vorgegebenen Textbausteine zusammensetzt. Dies kann erhebliche Zeit bei der Erstellung von wissenschaftlichen Berichten einsparen, wobei die Berichte trotzdem noch einmal von Menschenhand kontrolliert und ggf. überarbeitet werden sollten.

Für optionale Datenverbünde mit anderen Forschungsnetzwerken kann „Geonet-work“ nach dem INSPIRE-Standard zum Einsatz kommen. Der Report-Layer ist somit auch geeignet für interoperable Datenverbünde der L-TER Projekte.

6.4.9 Benutzerhilfe, Feedback und technische Dokumentation

Das „Rapid L-TER VRE“ Konzept zeigt verschiedene Wege für Benutzerhilfe, Feedback und Dokumentation und gibt praxistaugliche Empfehlungen für die Auswahl geeigneter Software. In den L-TER Projekten ließen sich dazu verschiedene Varianten im CMS und durch andere spezifische VRE-Services umsetzen. Für das CMS kann z.B. „FAQ-Book“ als Hilfesystem mit Feedback, sowie „K2“ und die native Artikelfunktion von „Joomla“ für Hilfetexte zum Einsatz kommen (vgl. Anhang D). Wikis sind ein adäquates Instrument, um Hilfen, Dokumentationen und andere Dokumente schnell und nachvollziehbar im Team zu erstellen. Als Wiki-Service kann die Software „MediaWiki“ genutzt werden. Durch die zahlreichen verfügbaren Erweiterungen ist das Wiki ein „Allround-Tool“ für eine teamorientierte Erstellung von Dokumenten mit multimediale Inhalten.

6.4.10 Weiterentwicklung

Das Konzept hilft durch die integrierte Abschätzung des Aufwandes bei der Auswahl, welche Funktionalitäten in der Basis-VRE und welche in der erweiterten VRE vorhanden sein sollten. Für die Projekte ist es nützlich, die L-TER VREs in Basis- und erweiterte VREs aufzuteilen. Damit können in wenigen Tagen und Wochen die jeweiligen Basis-Prototypen erstellt werden, die auch leicht in den Produktivbetrieb übergehen. Sind die Basisfunktionen vorhanden und auf zukünftige Erweiterungen ausgelegt, ist die Erweiterung einer VRE bedeutend einfacher. Diese Vorgehensweise der Aufteilung erscheint angebracht für eine schnelle Umsetzung der L-TER VREs. Innovative Technologien können in einem fortwährenden Entwicklungsprozess für den nachhaltigen Erfolg der L-TER VREs nützlich sein (vgl. Kapitel 4.3).

7. Schlussfolgerungen

Im Folgenden wird zunächst der wissenschaftliche Beitrag der Arbeit zusammengefasst und danach die relevanten Schlussfolgerungen daraus gezogen. Des Weiteren werden offene Fragen und zukünftiger Forschungsbedarf in diesem Bereich erörtert.

7.1 Wissenschaftlicher Beitrag der Arbeit

Das übergeordnete Ziel der Arbeit lag darin, ein Konzept zum Aufbau von nachhaltigen virtuellen Forschungsumgebungen für Projekte in der ökologischen Langzeitforschung auf der Basis von Open-Source-Software zu erstellen. VREs sollen die Kreativität, Effektivität und Produktivität in der Forschung und Lehre steigern, um somit die Forschungsergebnisse zu verbessern und den gesamten Forschungsprozess nachvollziehbar und nachhaltig verfügbar zu machen. Die Arbeit richtet sich vorwiegend an Projektplaner und Entscheidungsträger, die eine VRE aufbauen oder ihre vorhandene ICT-Infrastruktur erweitern möchten. Sie kann konzeptionell als Leitfaden für die Planung, den Aufbau und den Betrieb von (L-TER) VREs verwendet werden.

Da VREs oder ähnliche ICT-Infrastrukturen im Rahmen von „e-Science“ als „Rückgrat“ für die Organisation der Projekte prädestiniert sind, ist der Aufbau solcher Infrastrukturen oft schon zur Voraussetzung bei der Antragstellung von Projekten geworden. Ein Prototyp sollte daher bereits vor oder bei der Antragstellung existieren - spätestens zum Projektbeginn muss eine Produktivumgebung zur Verfügung stehen. Vor diesem Hintergrund ist das in der vorliegenden Arbeit erstellte Konzept „Rapid L-TER VRE“ speziell für eine schnelle Entwicklung von VRE Prototypen mit einer kurzen Übergangsphase in den Produktivbetrieb ausgelegt.

Es muss jedoch nicht immer gleich eine komplette oder neue L-TER VRE sein. Auf einer nachhaltigen Infrastruktur basierende ICT-Services können später noch zusammenwachsen. VREs sollten daher aufgebaut werden, wenn:

- prinzipiell der Mehrwert einer L-TER VRE den dazu notwendigen hohen Aufwand und Ressourcenbedarf übersteigt (vgl. Kapitel 3.3). Andernfalls kann es besser sein, bereits verfügbare Ressourcen zu prüfen und zu erweitern. Dies kann unter Umständen nur sehr schwer abgeschätzt werden,
- ein hoher Bedarf an Interaktivität zwischen den Benutzern, erweiterter Kommunikation mit Communities, und eine zentrale Plattform für digitale Forschungswerkzeuge benötigt werden,
- wenn es sich wegen des Aspektes der Langzeitnutzung insbesondere um L-TER VREs handelt.

Projekte in der ökologischen Langzeitforschung (L-TER) sind durch den Langzeitaspekt und Raumbezug in der Regel umfangreiche Projekte mit mittleren und großen Datenmengen (Medium- und Big Data) und vielen unterschiedlichen Untersuchungsobjekten auf verschiedenen Skalenebenen. Besonders der Langzeitaspekt unterscheidet sie von anderen Projekten. Im Sinne der ökologischen Langzeitforschung ist ggf. der Wissenstransfer über Generationen von Wissenschaftlern hinweg sicherzustellen. Dadurch wird ein vergleichsweise sehr langer Betriebszeitraum für VREs erforderlich.

Für die Entwicklung eines nachhaltigen Konzeptes wurden daher nach einer Literaturrecherche die speziellen Aspekte und Anforderungen an virtuelle Forschungsumgebungen von Projekten in der ökologischen Langzeitforschung genauer untersucht. Grundlage hierfür war der Stand des Wissens zu allgemeinen VREs, Datenmanagement und ICT, die interdisziplinär in Verbindung gebracht wurden mit den Zielen, Methoden und Daten der ökologischen Langzeitforschung.

Auf dieser Basis entstand ein „Baukastenkonzept“ zur Planung, dem Aufbau und Betrieb von nachhaltigen L-TER VREs, das auch für allgemeine VREs verwendet werden kann (vgl. Kapitel 4). Es ist für eine schnelle (rapide) Entwicklung von Prototypen mit einer schnellen Übergangsphase in den Produktivbetrieb ausgelegt und trägt daher den Namen „Rapid L-TER VRE“. Die schnelle Entwicklung wird vor allem durch „schlüsselfertige“ Server auf Open-Source-Basis und den Einsatz von virtuellen Maschinen (VMs) ermöglicht.

Das vorgestellte „Rapid L-TER VRE“ Baukastenkonzept basiert auf den folgenden Bestandteilen:

1. Einsatz von Open-Source-Software (mit Beispielen im Anhang)
2. ICT-Basistechnologien
3. „Schlüsselfertige“ Server und Virtualisierung
4. Homepage und CMS,
5. VRE-Services,
6. VRE-Tools,
7. Authentifizierung und SSO
8. Report Layer,
9. Benutzerhilfe, Feedback und Dokumentation
10. Weiterentwicklung

Das *Baukastenkonzept* bezieht sich hier insbesondere auf die VRE-Services und -Tools, die je nach Anforderungen der Projekte und Forschungsgemeinschaften die „Bausteine“ eines „Baukastens“ der L-TER VRE bilden (vgl. Anhang C, D, F).

Der Langzeitaspekt von L-TER VREs ist dabei in allen Bausteinen integriert. Die Umsetzung der gesamten VRE erfolgt schrittweise nach dem vorgestellten Life-Cycle-Modell für VREs. Dazu sollten parallel in der Aufbau- und Entwicklungsphase die

einzelnen Bausteine des Konzeptes nach dem Life-Cycle-Modell für einzelne Module durchlaufen werden.

Das erstellte Konzept wurde anhand der projektspezifischen Rahmenbedingungen, der einzelnen Punkte des „Rapid L-TER VRE“ Baukastenkonzeptes, der Nachhaltigkeit und der Entwicklungszeit mit folgendem Ergebnis bewertet (vgl. Kapitel 6):

Das „Rapid L-TER VRE“ Baukastenkonzept ist flexibel einsetzbar für die schnelle Realisierung von nachhaltigen L-TER VREs für L-TER Projekte mit einem breiten Spektrum an projektspezifischen Rahmenbedingungen unter Einbezug des Menschen, der als maßgeblicher Faktor wirken kann.

Im Einzelnen:

Das „Rapid L-TER VRE Konzept“ fasste in 10 Punkten alle notwendigen Aspekte zusammen, die für den erfolgreichen Aufbau von Prototypen zu den drei Anwendungsbeispielen erforderlich waren. Darüber hinaus ließ es sich flexibel für andere L-TER VREs in einem breiten Spektrum an projektspezifischen Eigenschaften nutzen. Die Anwendbarkeit des Konzeptes ist auch auf „größere“ L-TER Projekte mit individuellen Rahmenbedingungen und fachlichen Schwerpunkten übertragbar (theoretisch bewertet). Alle Punkte des Konzeptes tragen erheblich zur Nachhaltigkeit und schnellen Entwicklung von Prototypen bei, was auch für die Weiterentwicklung einer L-TER VRE theoretisch bewertet wurde. Auch der administrative Aufwand wird in vielen Bereichen deutlich reduziert. Tabelle 7.1 (Wiederholung von Tabelle 6.1) fasst die Gesamtbewertung des „Rapid L-TER VRE“ Konzeptes zusammen.

Die Nachhaltigkeit konnte im Rahmen dieser Arbeit nicht praktisch überprüft werden. Sie wurde jedoch konzeptionell auf Basis der Anwendungsbeispiele bewertet:

Das „Rapid L-TER VRE“ Baukastenkonzept ist geeignet unter dem Aspekt der Nachhaltigkeit.

Die Entwicklungszeit der Prototypen für die Anwendungsbeispiele wurde erheblich reduziert:

Basis-Prototypen für L-TER VREs können innerhalb weniger Tage und Wochen realisiert werden.

Im Vergleich dazu wird für z.B. für komplette Eigenentwicklung ein Zeitbedarf von Monaten bis Jahren geschätzt.

Das „Rapid L-TER VRE“ Baukastenkonzept ist geeignet zur schnellen Entwicklung von Prototypen und für Basis- sowie erweiterte L-TER VREs im Produktivbetrieb.

Tabelle 7.1: Gesamtbewertung des „Rapid L-TER VRE“ Konzeptes

Kriterium	Bewertung	Eignung
Flexibilität	<p>Projektgröße: kleine bis große Projekte (relativ),</p> <p>Projektstruktur: wenig bis komplexe Strukturen (relativ) bei: Institutionen, Datenstrukturen, Datenverbänden,</p> <p>Projektschwerpunkte: fachspezifisch, Forschungsdaten, Community,</p> <p>Projektzeitraum: minimal ca. 3 Jahre (Aufwand/Nutzen), kurz bis lang,</p> <p>Virtualisierung: dynamische Anpassung durch Virtualisierung.</p>	✓
Nachhaltigkeit	<p>Persistenz: Langzeitdatenspeicherung inkl.: Formate, Standards, Infrastruktur, VMs, Backups, Archivierung</p> <p>ICT: Virtualisierung, Modularisierung, SOA, offene internationale und weltweite Standards, Schnittstellen, Open-Source-Software (kostenlos) mit Entwicklergemeinde, Datenverbände. Die ICT im Konzept ist hinsichtlich der Persistenz, Konstanz, Resistenz, Elastizität, Resilienz, insbesondere z.B. durch Virtualisierungstechnologien dazu geeignet, nach großen Störungen schnell wieder in den Zustand vor der Störung zu gelangen. Innovative Technologien sind in Zukunft evtl. möglich.</p> <p>Faktor „Mensch“: Standards (gesellschaftlich), benutzerorientiert, Transparenz, Dokumentation, Hilfe, Feedback, zielgruppenorientiert.</p>	✓
Entwicklungszeit	<p>durch Virtualisierung erheblich verkürzt,</p> <p>durch „Schlüselfertige“ Systeme: erheblich verkürzt,</p> <p>durch Programmieraufwand: erheblich verkürzt,</p> <p>durch Baukasten: erheblich verkürzt,</p> <p>Weiterentwicklung: erheblich verkürzt.</p>	✓
Konzeptbestandteile	<p>Das „Rapid L-TER VRE Konzept“ fasste in 10 Punkten alle notwendigen Aspekte zusammen, die für den erfolgreichen Aufbau von Prototypen zu den Anwendungsbeispielen erforderlich waren.</p> <p>Darüber hinaus ließ es sich flexibel für andere L-TER VREs in einem breiten Spektrum an projektspezifischen Eigenschaften Nutzen.</p> <p>Die Anwendbarkeit des Konzeptes ist auch auf „größere“ L-TER Projekte übertragbar.</p> <p>Alle Punkte des Konzeptes haben erheblich in Nachhaltigkeit und Entwicklungsgeschwindigkeit zur Realisierung beigetragen haben.</p>	✓

Die folgende Übersicht (Tabelle 7.2) fasst zusammen, aus welchen Software-Komponenten sich nachhaltig Basis L-TER VREs mit einem soliden Funktionsumfang auf dem Stand der Technik aufbauen lassen. Die schlüsselfertigen VMs sind spezielle Versionen der entsprechenden Software, die sofort einsatzbereit als virtueller Server verfügbar sind.

Tabelle 7.2: Open-Source-Software für eine Basis L-TER VRE nach dem „Rapid L-TER Konzept“.

Name	Beschreibung	„Schlüsselfertige“ VM
„Linux“ Webserver, „mysql“	OS mit Web-Server und DBMS	Ja,
„Joomla“ oder anderes CMS	CMS und deren Erweiterungen	Ja
„GeoNetwork Open Source“	Metadatenkatalog, WebGIS, Datenaustausch	bedingt
„MediaWiki“	Wiki (Software für Teamwork bei Dokumenten)	Ja
„Moodle“	E-Learning-Plattform	Ja
„Virtual-Box“ / „VM-Ware“	Virtualisierung und Virtuelle Maschinen	Plattform
„OpenLDAP“	Verzeichnisserver zur zentralen Authentifizierung	Ja
„OwnCloud“	Cloud Server	Ja
„R-Statistics“-Server	Statistik-Software „Shiny“-Server zur Einbindung in Websites (Basisversion von „Shiny“ ist kostenlos)	Nein
„Netlogo“	Modellierung und Simulation, PC-Software erzeugt Java Applets, die in Websites eingebunden werden können	-

Im Anhang finden sich zusätzliche Software Empfehlungen (Open Source) für VRE- und CMS-Komponenten, die aus praktischer Erfahrung im Rahmen dieser Arbeit stammen und zum Teil im längeren Produktivbetrieb getestet wurden.

Allgemeine Schlussfolgerungen:

Bei Projekten im Umweltbereich und in der ökologischen Langzeitforschung ist neuerdings in den Projektanträgen als „quasi Voraussetzung“ für die Förderung ein Konzept für das elektronische Datenmanagement bzw. für eine VRE ähnliches Konstrukt im Sinne der guten wissenschaftlichen Praxis durch den Antragsteller vorzulegen. Das Thema „VRE“ ist daher besonders aktuell, weil das dazu erforderliche Know-how und die technischen Lösungen meistens nicht in den Forschungsbereich der Antragsteller fallen und sich bisher keine Standards oder „best practice“ Lösungen durchgesetzt haben. Bereits bei der Antragsstellung sollte ein solides Konzept mit zumindest einem L-TER VRE Prototypen bereitstehen, damit die Projektbeschreibung für den Antrag detailliert und erkennbar fundiert auf das System eingehen kann. Eine schnelle Implementierung von Prototypen ist daher besonders wichtig. Bei Projektstart sollte zuerst eine L-TER VRE als Basis vorhanden sein, damit die weiteren Forschungsaktivitäten geordnet stattfinden können. Auch aus diesem Grund sollte eine L-TER VRE so schnell wie möglich vom Prototypen und der Testphase in den Produktivbetrieb übergehen können. Vor diesem Hintergrund legt das „Rapid L-TER VRE“ Konzept einen besonderen Schwerpunkt auf die schnelle Entwicklung von VREs.

Nach der INSPIRE Richtlinie (INSPIRE 2007) der EU müssen die Mitgliedsländer für die Bereitstellung und Pflege von Metadaten und Suchdiensten für Metadaten im Umweltbereich seit 2007 mit einer zweijährigen Übergangsfrist Sorge tragen. Sie haben den Aufbau von Organisationsstrukturen mit geeigneten Netzwerkdiensten und Technologien zu gewährleisten, damit eine Koordinierung und die Berichterstattung an die EU möglich werden. Dazu gehört auch die Integration in die Geodateninfrastruktur Deutschland (GDI-DE). Um dieser Pflicht nachzukommen, werden ICT Werkzeuge benötigt, die für das Metadatenmanagement, den automatischen Austausch von Metadaten (Harvesting) und eine Berichterstattung geeignet sind. Diese Werkzeuge können von VREs bereitgestellt werden, wie in dieser Arbeit gezeigt wurde. Das „Rapid L-TER VRE“ Konzept erfüllt die Vorgaben der INSPIRE Richtlinie und daher prädestiniert für den Einsatz für INSPIRE-konforme VREs (vgl. Kapitel 2.6 und 5.1.3). Dazu enthält das Konzept die Basiskomponente für das Metadatenharvesting mit dem OAI-PMH Protokoll und für die Berichterstattung. Mit dem vorgeschlagenen Metadaten-Management-System „GeoNetwork Open Source“ lässt sich sowohl der automatische Metadatenaustausch realisieren, als auch die konforme Beschreibung mit Metadaten nach den Vorgaben aus den Anhängen I-III der INSPIRE Richtlinie.

Viele VREs (bzw. ähnliche Infrastrukturen) waren in der Vergangenheit nicht so erfolgreich, was sich zum einen aus der Literaturrecherche und zum anderen aus Gesprächen mit Projektbeteiligten von auf Fachkonferenzen und Workshops ergab. Das hatte sowohl technisch bedingte, als auch menschliche Ursachen, wie z.B. eine mangelnde Kooperation bei der Herausgabe von eigenen Forschungsdaten, die zu „leeren Archiven“ führten (engl. empty archives) (Nelson 2009). Für den Erfolg einer VRE kommt es nicht nur auf die verwendete Technologie bzw. Software an, sondern auch im besonderen Maße darauf, dass sie nachhaltig und maßgeschneidert auf die Projekte angewendet und konfiguriert wird – alles so einfach und benutzerfreundlich, wie möglich. Dabei müssen kulturelle, sprachliche, zeitliche und örtliche Unterschiede sowie die Vorbildung der Nutzer bei der Planung berücksichtigt werden. Die Zeit für den Aufbau der VRE und die Einarbeitungszeit kann für kleinere und kurze Projekte erheblich sein, sodass sich VREs eher bei größeren oder länger andauernden Projekten lohnen. Dies prädestiniert insbesondere L-TER Projekte für die Unterstützung durch VREs, wobei der Erfolg jedoch sehr individuell und hinsichtlich der einzelnen Akteure (Benutzer, Betreiber, Förderer) zu bewerten ist.

Die Verwendung von proprietärer Software und Cloud-Diensten von Drittanbietern als Teilkomponenten für den Aufbau von VREs bringt die Gefahr der Abhängigkeit vom Anbieter mit sich. Damit einhergehen datenschutzrechtliche Probleme und Sicherheitsrisiken. Aus diesem Grund ist es zweckdienlich, VREs weitgehend nur mit Open-Source-Software und auf eigenen Servern aufzubauen und zu betreiben. Obwohl in den letzten Jahren gerade die Systeme auf der Basis von Open Source einen technischen Stand erreicht haben, der eine ausgereifte Entwicklung komplexer VREs und vergleichbarer Software zulässt, gibt es weder unter den proprietären Produkten, noch unter den Open-Source-Projekten ein System, das sofort einsatzfähig wäre und allen Projektan-

forderungen gerecht und beliebig skaliert werden könnte. Vermutlich wird es dies auch in Zukunft aufgrund der Vielfältigkeit der Anforderungen und Funktionen sowie der rasanten Geschwindigkeit im Fortschritt der ICT-Entwicklung nicht geben. Einzelne Teilbereiche entwickeln sich mitunter so rasant, dass spezielles Know-how erforderlich wird, Module nicht mehr kompatibel sind und die Struktur immer komplexer wird. Nur eine Modularisierung in Verbindung mit Open-Source-Software und unter Berücksichtigung internationaler Standards kann nach dem Baukastenprinzip zu nachhaltigen, globalen, effizienten und ökonomischen Lösungen beitragen. So kann nach dem Open-Source-Prinzip eine weltweite Entwicklergemeinde an einem speziellen Modul bzw. Problem viel effizienter und mit mehr „Personal“ arbeiten, da auch der Benutzerkreis von kleinen spezifischen Modulen viel größer ist und die besondere fachliche Ausrichtung meist bessere Wirkung entfaltet.

Die wissenschaftliche Erkenntnis, die zu der Entwicklung von VREs im allgemeinen beigetragen hat besagt, dass VREs per Definition dazu gedacht sind, Forschung und Wissenschaft durch geeignete Methoden der ICT zu forcieren. In der Literatur finden sich jedoch wenig konkrete Hinweise dazu, wie der Nutzen durch VREs *direkt* zu messen ist, da hier vermutlich viele Einflüsse zusammentreffen und eine quantitative Bewertung daher sehr unscharf wäre. Durch VREs soll allgemein die Produktivität, Kreativität, Kommunikation und die Ergebnisse gesteigert und die Kosten gesenkt werden. Doch in der Praxis lässt sich dies nicht realistisch vergleichen. Es würde dazu das exakt gleiche Projekt einmal mit und einmal ohne VRE benötigt, jeweils mit den gleichen Anfangsvoraussetzungen, was jedoch nicht möglich ist, da z.B. die Personen andere wären oder bei den gleichen Personen schon das Vorwissen des vorangegangenen Projektes mit einfließt.

Die Gründung einer eigenständigen Organisation (z.B. eine GmbH) als Träger der VRE kann ggf. einige Vorteile haben. Zum einen ermöglichen unabhängige Rechtsformen das Schließen von Verträgen zum Beispiel mit Dienstleistern zur Bereitstellung der Infrastruktur. Oftmals ist es problematisch, die bestehenden Infrastrukturen der beteiligten Institutionen selber zu nutzen, da hier Kapazitäten, Sicherheitsaspekte und rechtliche Rahmenbedingungen eine Rolle spielen. Zum anderen kann die Verwaltung und Organisation der VRE an zentraler Stelle erfolgen. Dies vereinfacht den Betrieb und die Kosten.

Unter den beschriebenen VRE-Frameworks zum individuellen Aufbau von VREs, wie z.B. „gCube“ und „eSciDoc“ scheint es derzeit wenig aktuelle Weiterentwicklungen der Systeme zu geben. Dies wird ersichtlich an eingestellten Projekten, nicht vorhandenen Software-Updates, wenigen Erweiterungen und einer mäßigen Verbreitung. Es handelt sich zwar um Open-Source-Frameworks – diese bauen aber zu einem bedeutenden Teil auf eigenem Code auf. Sie fokussieren nicht den für L-TER Projekte notwendigen Ansatz der Nachhaltigkeit in Form von Persistenz, Dynamik und Flexibilität, die schnelle Realisierung von Prototypen und VREs im Produktivbetrieb sowie einen Baukasten für Open-Source-Services und -Tools, wie vergleichsweise das in dieser Arbeit vorgestellte „Rapid L-TER VRE“ Baukastenkonzept. Das Konzept könnte zu einem

Framework weiterentwickelt werden, wie das folgende Kapitel 7.2 zur zukünftigen Forschung zeigt.

Faktor „Mensch“:

Für die Bewertung des Erfolges einer nachhaltigen L-TER VRE, sowie als Teil (Resource) einer L-TER VRE, ist der Mensch als Individuum mit seinem sozialen sowie gesellschaftlichen Umfeld ein buchstäblich „maßgeblicher“ Faktor.

Der „Erfolg“ wird von einzelnen Personen oder Benutzergruppen (z.B.: Benutzer, Betreiber, Förderer, Forschungsgemeinschaften, Entscheidungsträger und Öffentlichkeit) ganz individuell beurteilt und teilweise sehr konträr (vgl. Kapitel 2.5.7).

Dies kann (nach eigener Einschätzung) z.B. dazu führen, dass Neid entsteht, weil andere Personen durch die eigenen offenen Forschungsdaten und Prozesse profitieren. Oder, das Problembewusstsein der Benutzer ist nicht vorhanden, warum es sich beispielsweise lohnt, eine L-TER VRE aufzubauen und dafür erhöhten eigenen und fremden Arbeitsaufwand zu investieren.

In diesem Zusammenhang ist auch die Transparenz und Offenheit eines Systems oder Projektes für die Akzeptanz und das Vertrauen der Menschen von besonderer Bedeutung, da es sich z.B. um persönliches, geistiges Eigentum handelt, das vor unberechtigter Nutzung geschützt werden soll, oder den Daten und Forschungsergebnissen anderer vertraut werden muss.

Diese Aspekte können zu „leeren Archiven“ (engl. empty archives, vgl. Nelson, und Bryn, 2009) oder im vorliegenden Fall zu ungenutzten L-TER VREs führen, die zwar auf dem technisch neusten Stand im Produktivbetrieb zur Verfügung stehen, aber nicht (mehr) genutzt werden und keine aktuellen Inhalte haben.

Auch müssen gesellschaftliche, regionale und kulturelle Standards und Unterschiede berücksichtigt werden. Dazu gehört die Konformität mit Gesetzen, Richtlinien, Data Policies, Fremdsprachen, Nutzungsgewohnheiten (z.B. Vorkenntnisse oder Vorlieben bei für Software), Zeitzonen, spezifische Feiertage, etc.).

Bislang wurde der Faktor „Mensch“ in vielen Projekten mit digitalen Infrastrukturen (e-Science) vernachlässigt. Die gesamte L-TER VRE muss daher auf den Menschen im Mittelpunkt ausgerichtet sein. Dazu gehören eigentlich selbstverständlicher Weise auch eine am Menschen orientierte nachhaltige Benutzung, Hilfe mit Support und Schulung, sowie Dokumentationen zu den L-TER VREs.

Die schrittweise Entwicklung von einer Basis- zu einer erweiterten L-TER VRE gibt die Möglichkeit, ein Feedback durch den Faktor „Mensch“ in der Weiterentwicklung zu berücksichtigen. Dies kann zu einer erheblichen Zeitersparnis in der Gesamtentwicklungszeit führen.

Der Mensch als Fehlerquelle in technischen Systemen kann erheblichen Einfluss auf die gesamte Persistenz von VREs haben. Dies betrifft besonders sensible Bereiche,

wie z.B. die Entwicklung und Administration von L-TER VREs, die ggf. die (irreversible) Zerstörung des gesamten Systems bewirken können.

Letztendlich lässt sich alles auf die drei Hauptfaktoren für den Erfolg einer nachhaltigen L-TER VRE reduzieren:

„Technik“, „Nachhaltigkeit“ und „Mensch“.

Mit der vorliegenden Dissertation soll letztendlich ein wissenschaftlicher Beitrag dazu geleistet werden, dass sich in Zukunft „digitale Forschungsgemeinschaften“ auf der Basis von L-TER VREs entwickeln, die nachhaltig Mensch und Technik im Sinne von „e-Science“ im Umweltbereich, insbesondere in der ökologischen Langzeitforschung, zusammenführen.

7.2 Offene Fragen und zukünftige Forschung

Das „Rapid L-TER VRE Konzept“ hat sich bei der praktischen Umsetzung von L-TER VRE Prototypen für Projekte mit verschiedenen Rahmenbedingungen in dieser Arbeit bewährt. Die Forschungsfragen aus der Einleitung in Kapitel 1.2 wurden in der Arbeit implizit in den spezifischen Kapiteln und in der Schlussfolgerung beantwortet. Für die Validierung eines nachhaltigen Konzeptes ist jedoch noch der Produktivbetrieb der VREs mit Benutzern und Administratoren und Entwicklern über einen längeren Zeitraum erforderlich. Es bleibt daher die Frage offen, ob sich das Konzept wirklich im langfristigen, praktischen Einsatz bewährt und wie es sich verbessern lässt.

Bewährt sich das Konzept in der Praxis, so könnte aus dem „Rapid L-TER VRE“ Konzept ein *Open-Source „Rapid L-TER VRE“ Framework* entstehen. Die Installations- und Konfigurationsskripte des Frameworks sollten mit verschiedenen Szenarien und der Vorauswahl von Funktionalitäten sowie Basis-Parametern automatisiert alle erforderlichen VMs vorkonfiguriert und schlüsselfertig erstellen. Dazu gehört auch Verbindung der einzelnen VMs und Komponenten als gesamtheitliche VRE. Dies sollte über eine unkomplizierte und beschleunigte Form der Installation erfolgen, wie z.B. durch mehrstufige „Installationsassistenten“ (engl. Software Wizzards). Ein derartiges Framework würde erheblich Zeit, Personal und Ressourcen bei der Erstellung einer nachhaltigen VRE einsparen. Lediglich technischen und inhaltlichen Einstellungen wären manuell für das jeweilige Projekt anzupassen.

In der Zukunft müssen VRE-Frameworks die Installation und Konfiguration von Standard Open-Source-Software und Virtuellen Maschinen (VMs) übernehmen. Auf der Basis dieser Standard-Software werden dort neue Erweiterungen speziell für VREs benötigt, wo allgemeine Erweiterungen ihre Grenzen finden. Andernfalls werden VRE-Frameworks vermutlich keine weite Verbreitung finden und kaum dem jeweils aktuellen Stand der Technik entsprechen. Das „Rapid L-TER VRE“ Konzept ist konzeptionell

dazu geeignet, zu einem nachhaltigen VRE-Framework durch automatisierte Installations- und Konfigurationsroutinen auf der Basis von vorgefertigten VMs weiterentwickelt zu werden.

Auf der Basis von „Blockchain-Technologien“ ließen sich eigene (ggf. Open-Source) VRE-Services für die „Blockchain-Forschungsdatenspeicherung“ inklusive Forschungsprozess entwickeln. Dies ermöglicht digital überprüfbare und nachvollziehbare „Vertragsketten“, die kryptografisch durch „Signaturen“ abgesichert und persistent in verteilten DBMS gespeichert sind. Dadurch können z.B. Urheberrechte und Verfahrensschritte des Forschungsprozesses der Autoren und Wissenschaftler unveränderlich nachvollzogen werden. Sie könnten ebenso als Grundlage z.B. im Rahmen des Umwelt Monitorings bei folgendem Anwendungsszenario dienen: Der Sensor als technisches Gerät (für Umweltparameter) und zugehörige Verarbeitungsprozesse in der VRE werden von einer Behörde digital signiert. Die VRE verarbeitet in signierten Verarbeitungsprozessen die signierten Sensordaten und erzeugt daraus signierte Forschungsergebnisse. Eine Behörde erhält die signierten Forschungsergebnisse z.B. in Form von Grenzwertüberschreitungen und überprüft die Signatur, um Manipulationen „auszuschließen“.

Mit dem Einsatz von „Blockchain-Technologien“ könnte für den Faktor „Mensch“ die *Transparenz* der Forschungsergebnisse und ihrer Prozesse sowie das *Vertrauen* erheblich gesteigert werden. Dies kündigt vielleicht der Anfang einer neuen „Ära der gesicherten Transparenz von Umweltdatendaten“ an.

Literatur- und Quellenverzeichnis

- 9starinc (2016): 9STAR: Secure Access IT Software Solutions. Online verfügbar unter <https://9starinc.com/>, zuletzt geprüft am 22.09.2017.
- Ahlf, Henning (2016): NoSQL-Datenbanken. In: *Das Wirtschaftsstudium: wisu : Zeitschrift für Ausbildung, Prüfung, Berufseinstieg und Fortbildung* 45 (10), S. 1133–1139.
- Alegue, Jean Duclos; Gnauck, Albrecht (2006): Using Fourier Analysis in Investigating Water Quality Indicators. In: Albrecht Gnauck (Hg.): *Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2005; [9. Workshop "Modellierung und Simulation von Ökosystemen"; 26.10.2005 - 28.10.2005]*. Aachen: Shaker (Berichte aus der Umweltinformatik, 98), 42-54.
- Allan, R.; Alden, A.; Boyd, D. et al. (2004): Roadmap for a UK Virtual Research Environment. Report of the JCSR VRE Working Group. Online verfügbar unter https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/JISC_UK/V040622J.pdf, zuletzt geprüft am 16.08.2017.
- Anderson, Christa; Griffin, Kristin L. (2010): *Windows Server 2008 R2 Remote Desktop Services Resource Kit*. Redmond, Wash.: Microsoft Press.
- apache.org (2017): Homepage: Apache License. Online verfügbar unter <https://www.apache.org/licenses/LICENSE-2.0>, zuletzt aktualisiert am 19.09.2017, zuletzt geprüft am 07.12.2017.
- Archer, Margaret Scotford (Hg.) (1998): *Critical realism. Essential readings*. London: Routledge (Critical realism).
- Arnaud-Haond, Sophie; Duarte, Carlos M.; Diaz-Almela, Elena; Marbà, Núria; Sintes, Tomas; Serrão, Ester A. (2012): Implications of extreme life span in clonal organisms. Millenary clones in meadows of the threatened seagrass *Posidonia oceanica*. In: *PloS one* 7 (2), e30454. DOI: 10.1371/journal.pone.0030454.
- Assante, Massimiliano; Candela, Leonardo; Castelli, Donatella; Frosini, Luca; Lelii, Lucio; Manghi, Paolo et al. (2008): An Extensible Virtual Digital Libraries Generator. In: Birte Christensen-Dalsgaard, Donatella Castelli, Bolette Ammitzbøll Jurik und Joan Lippincott (Hg.): *Research and advanced technology for digital libraries. 12th European conference, ECDL 2008, Aarhus, Denmark, September 14 - 19, 2008 ; proceedings, Bd. 5173*. Berlin, Heidelberg: Springer (Lecture notes in computer science, 5173), S. 122–134.
- Atkins, D. E.; Cyberinfrastructure, National Science Foundation Blue-Ribbon Advisory Panel on (2003): *Revolutionizing Science and Engineering Through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure: National Science Foundation*. Online verfügbar

- unter <https://books.google.de/books?id=TWNQAAAAMAAJ>, zuletzt geprüft am 30.11.2017.
- Baumann, Silke (2014): Langzeitarchivierung innerhalb Virtueller Forschungsumgebungen im Bereich Digital Humanities. Hg. v. Konrad Umlauf. Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, 353).
- Bengel, Günther (2014): Grundkurs Verteilte Systeme. Grundlagen und Praxis des Client-Server und Distributed Computing. 4. Aufl. Wiesbaden: Springer Vieweg. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-8348-2150-8>, zuletzt geprüft am 30.11.2017.
- Berman, Francine; Brady, Henry; Kastrinos, Nikolaos; Blumenthal, Marjory (2005): Final Report. NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. In: . Online verfügbar unter https://immagic.com/eLibrary/ARCHIVES/GENERAL/US_NSF/N050512B.pdf, zuletzt geprüft am 16.08.2017.
- Birken, Philipp (2011): Ein viertes Paradigma der Wissenschaft. Online verfügbar unter <http://birchlog.blogspot.de/2011/04/ein-viertes-paradigma-der-wissenschaft.html>, zuletzt geprüft am 15.08.2017.
- Blokland, Kees; Mengerink, Jeroen; Pol, Martin; Rubruck, Doris (2016): Cloud-Services testen. Von der Risikobetrachtung zu wirksamen Testmaßnahmen. Heidelberg: dpunkt.Verlag.
- Borgida, Alexander T. (Hg.) (2009): Conceptual modeling: foundations and applications. Essays in honor of John Mylopoulos. Unter Mitarbeit von John Mylopoulos. Berlin, Heidelberg, New York, NY: Springer (Lecture notes in computer science, 5600 : Festschrift).
- Brase, Jan; Klump, Jens (2007): Zitierfähige Datensätze: PrimärdatenManagement durch DOIs. In: Rafael Ball (Hg.): Wissenschaftskommunikation der Zukunft. 6. - 8. November 2007; Beiträge und Poster. Jülich: Forschungszentrum, Zentralbibliothek (Schriften des Forschungszentrums Jülich, Bd. 18), S. 159–167.
- Breitenstrom, Christian; Micklisch, Clemens; Mochól, Małgorzata (2013): Sicherheitsstudie Content Management Systeme (CMS). Bundesamt für Sicherheit in der Informationstechnik. Online verfügbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Studien/CMS/Studie_CMS.pdf?__blob=publicationFile, zuletzt geprüft am 30.11.2017.
- BTU Cottbus-Senftenberg (2017): Forschungsplattform Hühnerwasser/ Research Platform Chicken Creek - BTU Cottbus-Senftenberg. BTU Cottbus-Senftenberg. Online verfügbar unter <https://www.b-tu.de/chicken-creek/>, zuletzt aktualisiert am 01.03.2017, zuletzt geprüft am 14.09.2017.

- Buddenbohm, Stefan; Hofmann, Matthias; Neuroth, Heike (2014): Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen. Unter Mitarbeit von Stefan Buddenbohm, Harry Enke, Matthias Hofmann, Jochen Klar, Heike Neuroth und Uwe Schwiegelshohn. Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen (DARIAH-DE working papers, 7).
- Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hg.) (2011): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen.
- CAHDE (2009): Ad hoc Committee on e-democracy: Electronic democracy ("e-democracy"). Recommendation CM/Rec(2009)1 adopted by the Committee of Ministers of the Council of Europe on 18 February 2009 and explanatory memorandum. Strasbourg: Council of Europe Publ (Legal instruments).
- Candela, L.; Castelli, D.; Pagano, P. (2011): History, Evolution and Impact of Digital Libraries. In: I. Iglezakis, T.-E. Synodinou und S. Kapidakis (Hg.): E-publishing and digital libraries: Legal and organizational issues. PA, USA: Hershey (Information Science Reference).
- Candela, L.; Castelli, Donatella; Pagano, Pasquale (2010): Making Virtual Research Environments in the Cloud a Reality: the gCube Approach. In: ERCIM News (83), S. 32–33.
- Candela, Leonardo; Akal, Fuat; Avancini, Henri; Castelli, Donatella; Fusco, Luigi; Guidetti, Veronica et al. (2007): DILIGENT: integrating digital library and Grid technologies for a new Earth observation research infrastructure. In: International Journal on Digital Libraries 2007 (7), S. 59–80.
- Candela, Leonardo; Castelli, Donatella; Pagano, Pasquale (2013): Virtual research environments: An overview and a research agenda. In: Data Science Journal 2013 (12), S. 75–81. Online verfügbar unter <https://datascience.codata.org/articles/10.2481/dsj.GRDI-013/galley/241/download/>, zuletzt geprüft am 16.08.2017.
- Candela, Leonardo; Castelli, Donatella; Pagano, Pasquale; Simi, Manuele (2005): From Heterogeneous Information Spaces to Virtual Documents. In: Edward A. Fox, Erich Neuhold, Pimrumpai Premssmit und Vilas Wuwongse (Hg.): Digital Libraries: Implementing Strategies and Sharing Experiences. 8th International Conference on Asian Digital Libraries, ICADL 2005, Bangkok, Thailand, December 12-15, 2005, Proceedings. Berlin Heidelberg: Springer-Verlag (Lecture notes in computer science, 3815), S. 11–22.
- Carusi, Annamaria; Reimer, Torsten (2010): Virtual Research Environment Collaborative Landscape Study. Hg. v. JISC. Online verfügbar unter <https://www.webarchive.org.uk/wayback/archive/20140615234259/http://www.jisc.ac.uk/media/documents/publications/vrelandscape.pdf>, zuletzt geprüft am 16.08.2017.

- Conference on Environment and Development; United Nations Conference on Environment and Development (1992): Agenda 21, Rio Declaration, Forest Principles. Drafts. New York: United Nations. Online verfügbar unter http://www.un.org/depts/german/conf/agenda21/agenda_21.pdf, zuletzt geprüft am 15.08.2017.
- Convention on Biological Diversity (2017): Homepage. Convention on Biological Diversity (CBD). Online verfügbar unter <https://www.cbd.int/>, zuletzt geprüft am 30.11.2017.
- corba.org (2017): Common Object Request Broker Architecture™ (COBRA) - Homepage. Online verfügbar unter <http://www.corba.org/>, zuletzt geprüft am 30.08.2017.
- creativecommons.org (2017): Homepage. Online verfügbar unter <https://creativecommons.org>, zuletzt geprüft am 30.11.2017.
- Cyberinfrastructure Council (2007): Cyberinfrastructure Vision for the 21st Century Discovery. National Science Foundation.
- D4Science (2017): Homepage. Online verfügbar unter <https://www.d4science.org/>, zuletzt geprüft am 16.08.2017.
- debian.org (2017): Homepage: Debian - Debian-Gesellschaftsvertrag. Online verfügbar unter http://www.debian.org/social_contract#guidelines, zuletzt aktualisiert am 23.07.2017, zuletzt geprüft am 19.09.2017.
- Dehler, Elmar; Hardy, Dirk; Troßmann, Hubert (2015): Datenbanken. Entwickeln, programmieren, anwenden. 2. Aufl., 1. Dr. Haan-Gruiten: Verl. Europa-Lehrmittel Nourney, Vollmer (Europa-Fachbuchreihe Programmierung für die IT-Ausbildung).
- Deka, Ganesh Chandra (Hg.) (2016): NoSQL. Database for storage and retrieval of data in cloud. Boca Raton, FL, London, New York: CRC Press Taylor & Francis Group.
- Deutsche Forschungsgemeinschaft (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft". Weinheim: Wiley-VCH.
- Deutsche Forschungsgemeinschaft (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. Hg. v. Deutsche Forschungsgemeinschaft - Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme - Unterausschuss für Informationsmanagement. Online verfügbar unter http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf, zuletzt geprüft am 17.08.2017.

- Deutsche Forschungsgemeinschaft (2013): Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft". Weinheim: Wiley-VCH.
- Deutsche Initiative für Netzwerkinformation e.V (2008): DINI-Thesen zur Informations- und Kommunikationsstruktur der Zukunft. Zehn Thesen zur Entwicklung von Service und Servicestrukturen für Information und Kommunikation in Forschung, Lehre und Studium. Göttingen. Online verfügbar unter https://dini.de/fileadmin/docs/DINI_thesen.pdf, zuletzt geprüft am 20.09.2017.
- Deutsche Initiative für Netzwerkinformation e.V. (2011): Deutsche Initiative für Netzwerkinformation e.V. - Homepage: AG Virtuelle Forschungsumgebungen. Online verfügbar unter <http://dini.de/ag/vforum/>, zuletzt geprüft am 17.08.2017.
- Devedžić, Vladan; Gašević, Dragan (Hg.) (2009): Web 2.0 & Semantic Web. 1. Aufl. Boston, MA: Springer US.
- Diedrich, Oliver (2016): Was ist Open Source? Heise Medien. Online verfügbar unter <https://www.heise.de/ct/artikel/Die-Woche-Was-ist-Open-Source-222121.html>, zuletzt aktualisiert am 28.06.2007, zuletzt geprüft am 19.09.2017.
- Dietzel, Michael; Petsch, Mathias; Gold, Stefan: Cloud Computing vs. SOA - eine Analyse moderner Architekturparadigmen für betriebliche Informationssysteme. @Ilmenau, Techn. Univ., Masterarbeit, 2011.
- DILIGENT (2017): Home - diligentproject.org. Online verfügbar unter <http://diligent.ercim.eu/>, zuletzt aktualisiert am 06.01.2011, zuletzt geprüft am 16.08.2017.
- Dimoulis, Georgios (2015): Die Auswahl des richtigen Open Source CMS. Marktüberblick und Nutzwertanalyse von WordPress, TYPO3, Joomla!, Drupal und eZ Publish. 1. Auflage. Hamburg: Diplomica Verlag.
- Docker.com (2017): Homepage - Enterprise Container Platform for High-Velocity Innovation. Online verfügbar unter <https://www.docker.com/>, zuletzt geprüft am 19.09.2017.
- Dröschmeister, Rainer (Hg.) (1998): Die Bedeutung ökologischer Langzeitforschung für Naturschutz. Referate und Ergebnisse des gleichnamigen Symposiums auf der Insel Vilm vom 26. - 30. Mai 1997. Deutschland. Münster: BfN-Schr.-Vertrieb im Landwirtschaftsverl. (Schriftenreihe für Landschaftspflege und Naturschutz, H. 58).
- Eberhard, Kurt (1999): Einführung in die Erkenntnis- und Wissenschaftstheorie. Geschichte und Praxis der konkurrierenden Erkenntniswege. 2., durchges. und erw. Aufl. Stuttgart: Kohlhammer (Urban-Taschenbücher, 386).
- EGEE (2017): Homepage: EGEE Portal: Enabling Grids for E-science. Online verfügbar unter <https://eu-egee-org.web.cern.ch/eu-egee-org/index.html>, zuletzt aktualisiert am 15.01.2014, zuletzt geprüft am 27.09.2017.

- egi.eu (2017): Homepage: EGI | EGI Advanced Computing Services for Research. Online verfügbar unter <https://www.egi.eu/>, zuletzt geprüft am 27.09.2017.
- e-Infrastructure Reflection Group (2010): Blue Paper. E-IRG. Online verfügbar unter https://www.lrz.de/forschung/projekte/projekte_completed/e-irgsp3/, zuletzt geprüft am 16.08.2017.
- Elsayed, Ibrahim (2012): Dataspace Support Platform for e-Science. Dataspace-based Preservation of Scientific Studies. neue Ausg. Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften.
- Erbs, Heinz-Erich; Karczewski, Stephan; Schestag, Inge (2003): Datenbanken. Datenmodelle, Objekte, WWW, XML. 1. Aufl. Berlin, Offenbach: VDE-Verl.
- escidoc.org (2016): Homepage: escidoc.org. Online verfügbar unter <https://www.escidoc.org/>, zuletzt geprüft am 05.06.2016.
- Etzkorn, Letha Hughes (2017): Introduction to middleware. Web services, object components, and cloud computing. Boca Raton, London, New York: CRC Press Taylor & Francis Group (A Chapman & Hall book).
- Europäische Kommission (2006): Study on the Economic impact of open source software on innovation and the competitiveness of the Information and Communication Technologies (ICT) sector in the EU. Unter Mitarbeit von Rishab Aiyer Ghosh: MERIT (ENTR/04/112).
- Europäische Union (2017): Homepage: FP7. Online verfügbar unter https://ec.europa.eu/research/fp7/index_en.cfm, zuletzt geprüft am 20.09.2017.
- Europäisches Parlament (2000): Richtlinie 2000/60/EG des Europäischen Parlaments und des Rates vom 23. Oktober 2000 zur Schaffung eines Ordnungsrahmens für Maßnahmen der Gemeinschaft im Bereich der Wasserpolitik. Europäisches Parlament. Online verfügbar unter <http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32000L0060>, zuletzt aktualisiert am 15.08.2017.
- Europäisches Parlament (2017): Natura 2000 Homepage. Europäisches Parlament. Online verfügbar unter http://ec.europa.eu/environment/nature/natura2000/index_en.htm, zuletzt geprüft am 15.08.2017.
- European Environment Agency (2016): Homepage: Shared Environmental Information System (SEIS). Online verfügbar unter <https://www.eea.europa.eu/about-us/what/shared-environmental-information-system-1/shared-environmental-information-system>, zuletzt aktualisiert am 09.08.2016, zuletzt geprüft am 20.09.2017.
- Federal Geographic Data Committee (2017): Content Standard for Digital Geospatial Metadata (CSDGM). Online verfügbar unter <https://www.fgdc.gov/metadata/csdgm-standard>, zuletzt geprüft am 19.08.2017.

- Feldmann, Florian (2015): Binding credentials. Securing (SSO) authentication. Bochum, Ruhr-Universität Bochum, Dissertation. Online verfügbar unter <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:hbz:294-45159>, zuletzt geprüft am 30.11.2017.
- Filetti, Mirko (2014): Augmented Reality - Points of Interest in Environmental Mobile Applications. In: Nguyen Xuan Thinh (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2013 [vom 30.10. bis zum 1.11.2013]. Berlin: Rhombos-Verl., S. 139–150.
- Filetti, Mirko; Gnauck, Albrecht (2011): Mirko Filetti, Albrecht Gnauck. A Concept of a Virtual Research Environment for LongTerm Ecological Projects with Free and Open Source Software. In: Jiří Hřebíček, Gerald Schimak und Ralf Denzer (Hg.): Environmental Software Systems. Frameworks of eEnvironment. 9th IFIP WG 5.11 International Symposium, ISESS 2011, Brno, Czech Republic, June 27-29, 2011. Proceedings. Berlin, Heidelberg: IFIP International Federation for Information Processing (SpringerLink Bücher, 359).
- Foster, I.; Zhao, Y.; Raicu, L.; Lu, S. (2009): Cloud Computing and Grid Computing 360-Degree Compared. In: Proceedings of the 5th Grid Computing Environments Workshop. New York, NY: ACM.
- Foster, Ian; Kesselman, Carl (Hg.) (1999): The grid. Blueprint for a new computing infrastructure. San Francisco, Calif.: Kaufmann. Online verfügbar unter <http://www.loc.gov/catdir/description/els033/98030570.html>, zuletzt geprüft am 30.11.2017.
- Franklin, Michael; Halevy, Alon; Maier, David (2005): From databases to dataspace. In: SIGMOD Rec. 34 (4), S. 27–33. DOI: 10.1145/1107499.1107502.
- Fraser, Michael (2005): Virtual Research Environment: Overview and Activity. Online verfügbar unter <http://www.ariadne.ac.uk/issue44/fraser/>, zuletzt geprüft am 15.08.2017.
- Free Software Foundation (1996): Freie Software. Was ist das? Hg. v. Free Software Foundation. Online verfügbar unter <https://www.gnu.org/philosophy/free-sw>, zuletzt aktualisiert am 15.08.2017.
- freebsd.org (2017): Homepage: The FreeBSD Copyright. Online verfügbar unter <https://www.freebsd.org/copyright/freebsd-license.html>, zuletzt geprüft am 07.12.2017.
- Fritzlar, Heinrich; Huber, Andreas; Rudl, Alexandra (Hg.) (2012): Open Source im öffentlichen Sektor: flexibler, sicherer, günstiger. Was der öffentliche Sektor von dem Zukunftstrend lernen kann. Boizenburg: Hülsbusch (Schriftenreihe des Innovators Club - Deutschlandforum Verwaltungsmodernisierung, Bd. 3).
- fsf.org (2017): Homepage: Free Software Foundation. Online verfügbar unter <https://www.fsf.org/>, zuletzt geprüft am 19.09.2017.

- Fuchs-Kittowski, Klaus; Wohlgemuth, Volker (2011): Umweltinformatik und Umweltforschung - in ihrer Institutionalisierung und Interdisziplinarität. In: Michael Böcher: Interdisziplinarität und Institutionalisierung der Wissenschaft. Hg. v. Klaus Fischer. Berlin: wvb Wiss. Verl. (Wissenschaftsforschung, 2010), S. 99–151. Online verfügbar unter http://www.wissenschaftsforschung.de/ JB10_Fuchs.pdf, zuletzt geprüft am 07.12.2017.
- Furht, Borko (Hg.) (2011): Handbook of Augmented Reality. 1., st Edition. New York, NY: Springer New York.
- García Márquez, Fausto Pedro; Lev, Benjamin (Hg.) (2017): Big Data Management. Springer International Publishing AG. 1st ed. 2017. Cham: Springer International Publishing.
- gcube-system.org (2017a): Catalogue of Applications - gCube Website. Online verfügbar unter <https://www.gcube-system.org/catalogue-of-applications>, zuletzt geprüft am 10.12.2017.
- gcube-system.org (2017b): Homepage: gCube. Online verfügbar unter <https://www.gcube-system.org/>, zuletzt geprüft am 27.09.2017.
- gcube-system.org (2017c): Publications - gCube Website. Online verfügbar unter <https://www.gcube-system.org/papers>, zuletzt geprüft am 10.12.2017.
- geonetwork-opensource.org (2017): Homepage: GeoNetwork opensource. Online verfügbar unter <https://geonetwork-opensource.org/>, zuletzt geprüft am 18.12.2017.
- geoportal.de: Geoportal.de - Geodaten aus Deutschland - GDI-DE. Geoportal. Online verfügbar unter <http://www.geoportal.de/DE/GDI-DE/gdi-de.html?lang=de>, zuletzt geprüft am 28.09.2017.
- Gerken, Wolfgang (2016): Datenbanksysteme für Dummies. 1. Auflage. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA (Lernen leichter gemacht).
- Gertsch, Daniel Th; Walker, Janina (2015): Virtual Book. Grundlagenwerk über die Virtualisierung ; [Grundlagen, Techniken, Architekturen der Virtualisierung]. 1. Aufl. Frick: swiboo.ch (Lehrmittel Virtualisierung).
- Gnauck, A. (2006): Funktoren und Signale - Zur Signalanalyse ökologischer Prozesse. In: Albrecht Gnauck (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2004. Aachen: Shaker (Berichte aus der Umweltinformatik, 93), S. 192–224.
- Gnauck, A. H.; Straškraba, M. (1985): Freshwater Ecosystems. Modelling and Simulation. Burlington: Elsevier Science (Developments in Environmental Modelling). Online verfügbar unter <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10771921>, zuletzt geprüft am 30.11.2017.

- Gnauck, Albrecht (Hg.) (2001): Systemtheorie und Modellierung von Ökosystemen. Mit 22 Tabellen. Heidelberg: Physica-Verl. (Umweltwissenschaften).
- Gnauck, Albrecht (2004): Signalmodelle von Zeitreihen der Wassergüte. In: Albrecht Gnauck (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2002. Aachen: Shaker (ASIM-Mitteilungen, 82), S. 62–79.
- Gnauck, Albrecht (Hg.) (2011): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2010 ; [14. Workshop "Modellierung und Simulation von Ökosystemen" ; 27.10.2010 - 29.10.2010]. Workshop Kölpinsee; Gesellschaft für Informatik; Workshop "Modellierung und Simulation von Ökosystemen". Aachen: Shaker (Berichte aus der Umweltinformatik, AM 132).
- Gnauck, Albrecht; Luther, Bernhard (2006): Simulation and Parameter Optimisation of a Shallow Lake Eutrophication Model. In: Albrecht Gnauck (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2004. Aachen: Shaker (Berichte aus der Umweltinformatik, 93), S. 14–29.
- Gnauck, Albrecht; Pillmann, Werner (2011): Environmental Meta-Modles: The ICT-ENSURE Project. In: Albrecht Gnauck (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2010 ; [14. Workshop "Modellierung und Simulation von Ökosystemen" ; 27.10.2010 - 29.10.2010]. Aachen: Shaker (Berichte aus der Umweltinformatik, AM 132), S. 1–16.
- Gnauck, Albrecht.; Li, Bai-Lian Larry; Feugo, Jean Duclos Alegeu; Luther, Bernhard (2010): The Role of Statistics for Long-Term Ecological Research. In: Felix Müller, Cornelia Baessler, Hendrik Schubert und Stefan Klotz (Hg.): Long-Term Ecological Research. Between Theory and Application. Dordrecht: Springer Science+Business Media B.V, S. 107–128.
- gnu.org (2017): Homepage: gnu.org. Online verfügbar unter <https://www.gnu.org/home.de.html>, zuletzt geprüft am 07.12.2017.
- Goldfarb, Charles F.; Prescod, Paul (2000): Das XML-Handbuch. [Anwendungen, Produkte, Technologien]. 2., aktualisierte und erw. Aufl. München: Addison-Wesley (Charles F. Goldfarb series on open information management).
- Google (2017): Charts : Google Developers. Online verfügbar unter <https://developers.google.com/chart/>, zuletzt aktualisiert am 02.05.2017, zuletzt geprüft am 22.09.2017.
- Grass, Robert N.; Heckel, Reinhard; Puddu, Michela; Paunescu, Daniela; Stark, Wendelin J. (2015): Robust chemical preservation of digital information on DNA in silica with error-correcting codes. In: Angewandte Chemie (International ed. in English) 54 (8), S. 2552–2555. DOI: 10.1002/anie.201411378.
- Gregersen, J. B.; Gijsbers, P. J. A.; Westen, S. J. P. (2007): Openmi: Open modelling interface. Journal of Hydroinformatics, 9(3):175–191, 2007. In: Journal of Hydroinformatics (9(3)), S. 175–191. Online verfügbar unter

- <http://www.iwaponline.com/jh/009/jh0090175.htm>, zuletzt geprüft am 02.11.2017.
- Gries, Thomas (1986): Fehlerkorrekturverfahren mittels Reed-Solomon-Codes für adaptive Teilband-Sprachcodierer. Berlin: Inst. für Fernmeldetechnik. Online verfügbar unter http://opus.kobv.de/tuberlin/volltexte/2008/1781/pdf/gries_thomas.pdf, zuletzt geprüft am 30.11.2017.
- Grune, Dick; Jacobs, Cerial J.H (2004): Parsing Techniques. 2. ed. 2/2005. New York, NY: Springer Verlag New York (Monographs in Computer Science).
- Hamburger, Valentin (2016): Building VMware Software-Defined Data Centers. S. 254, Pact Publishing Birmingham (UK), 2016
- Haynes, David (2004): Metadata for information management and retrieval. London: Facet Publ (Become an expert).
- Hey, Tony; Tansley, S.; Tolle, K. (2009): Jim Gray on eScience. A transformed scientific method. In: Tony Hey, S. Tansley und K. Tolle (Hg.): The Fourth Paradigm: Data-Intensive Scientific Discovery. 2. printing, Version 1.1., October 2009. Redmond, Wash.: Microsoft Research, S. XVII–XXXI.
- Higgins, Sarah (2008): The DCC Curation Lifecycle Model. In: The International Journal of Digital Curation (3-2008), S. 134–140. Online verfügbar unter <http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48>, zuletzt geprüft am 18.08.2017.
- Hitzler, Pascal (2008): Semantic Web. Grundlagen. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (eXamen.press). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-540-33994-6>, zuletzt geprüft am 30.11.2017.
- Hobbie, John E.; Carpenter, Stephen R.; Grimm, Nancy B.; Gosz, James R.; Seasted, Timothy R. (2003): The US Long Term Ecological Research Program. In: BioScience 53 (1), S. 21–32. DOI: 10.1641/0006-3568(2003)053[0021:TULTER]2.0.CO;2.
- Holl, Sonja (2014): Automated optimization methods for scientific workflows in e-science infrastructures. Jülich: Forschungszentrum Jülich (Schriften des Forschungszentrums Jülich IAS Series, 24).
- Hoogerwerf, M.; Lösch, M.; Schirrwagen, J.; Callaghan, S.; Manghi, P.; Iatropoulou, K. (2013): Linking Datan and Publications: Towards a Cross-Disciplinary Approach. In: International Journal of Digital Curations (8), S. 244–254.
- Hřebíček, Jiří; Schimak, Gerald; Denzer, Ralf (Hg.) (2000): Environmental Software Systems. Environmental Information and Decision Support. International Federation for Information Processing; International Symposium on Environmental Software Systems. Boston, MA: Springer US.

-
- hydroeurope.org (2017): Homepage: HydroEurope. Online verfügbar unter <http://www.hydroeurope.org/>, zuletzt geprüft am 21.09.2017.
- ICT-ENSURE (2017): Homepage: ICT-ENSURE Literature Information System. Online verfügbar unter <https://www.iai.kit.edu/ictensure/site?mod=litdb>, zuletzt geprüft am 20.09.2017.
- IEEE (2017): Homepage. Online verfügbar unter <https://www.ieee.org/index.html>, zuletzt geprüft am 19.08.2017.
- Imboden, Dieter M.; Koch, Sabine; Stürchler, Nikolas (2008): Systemanalyse. Einführung in die mathematische Modellierung natürlicher Systeme ; mit 8 Tabellen. 3., korrigierter Nachdr. Berlin: Springer.
- INSPIRE (2007): Infrastructure for spartial information in Europe - Homepage. Online verfügbar unter <https://inspire.ec.europa.eu/>, zuletzt aktualisiert am 30.08.2017, zuletzt geprüft am 30.08.2017.
- iso.org (2017): ISO - International Organization for Standardization. Online verfügbar unter <https://www.iso.org/home.html>, zuletzt geprüft am 11.12.2017.
- jahia.com (2017): Homepage. Online verfügbar unter <https://www.jahia.com/de/home>, zuletzt geprüft am 14.12.2017.
- Jansen, S.; Brinkkemper, S.; Cusumano, M. A. (2013): Software ecosystems. Analyzing and managing business networks in the software industry. Cheltenham: Edward Elgar Publishing. Online verfügbar unter <http://ebookcentral.proquest.com/lib/subhh/detail.action?docID=1190645>, zuletzt geprüft am Hobbie, John E.; Carpenter, Stephan R.; Grim, Nancy B.; Gosz, James R.; Seastedt, Timothy R.
- Jansen, Slinger; Brinkkemper, Sjaak (Hg.) (2013): Software ecosystems. Analyzing and managing business networks in the software industry. Cheltenham: Edward Elgar Publishing Ltd.
- Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. Mannheim: GESIS (GESIS technical reports, 2012,07).
- Jensen, Uwe; Katsanidou, Alexia; Zenk-Möltgen, Wolfgang (2011): Metadaten und Standards. Potsdam: Fachhochschule Potsdam.
- joomla.org (2017): Homepage: Joomla! The CMS Trusted By Millions for their Websites. Online verfügbar unter <https://www.joomla.org/>, zuletzt geprüft am 22.11.2017.
- joomlart (2017): Homepage: Joomla Templates and Joomla Extensions - JoomlaArt | Joomla Templates and Extensions Provider. Online verfügbar unter <https://www.joomlart.com/>, zuletzt geprüft am 22.09.2017.

- Joomlashine (2017): Homepage: JoomlaShine | Hi-quality Joomla 3.x templates & extensions. Online verfügbar unter <https://www.joomlashine.com/>, zuletzt geprüft am 22.09.2017.
- Jørgensen, S. E. (1976): A eutrophication model for a lake. In: *Ecological Modelling* (2), S. 147–162.
- Joseph, Joshy; Fellenstein, Craig (2004): *Grid computing*. Upper Saddle River, N.J.: Prentice Hall Professional Technical Reference (On demand series). Online verfügbar unter <http://proquest.tech.safaribooksonline.de/0131456601>, zuletzt geprüft am 30.11.2017.
- Karve, Swagat (2017): *Reed-Solomon Codes. Theory and Techniques*. 1. Auflage. Saarbrücken: LAP LAMBERT Academic Publishing.
- Kim, Ji-Sun (2009): *Tangible User Interface for CAVE based on Augmented Reality Technique. A new 3D user interface system for a CAVE application, based on Virtual Reality (VR), Augmented Reality (AR), and Tangible User Interface (TUI)*. Saarbrücken: VDM Verlag Dr. Müller.
- KlimaScout (2017): Homepage: KlimaScout – Das Wiki zur Anpassung an den Klimawandel. Online verfügbar unter <http://www.klimascout.de/>, zuletzt geprüft am 20.09.2017.
- Klinke, Sigbert (2014): *Extension:R - MediaWiki*. Online verfügbar unter <https://www.mediawiki.org/wiki/Extension:R>, zuletzt aktualisiert am 08.09.2017, zuletzt geprüft am 22.09.2017.
- Klinke, Sigbert (2016): *Extension:Shiny - MediaWiki*. Online verfügbar unter <https://www.mediawiki.org/wiki/Extension:Shiny>, zuletzt aktualisiert am 08.09.2017, zuletzt geprüft am 22.09.2017.
- Klump, J., Bertelmann, R. (2013): *Forschungsdaten* - In: Kühlen, R., Semar, W., Strauch, D. (Eds.), *Grundlagen der praktischen Information und Dokumentation*, De Gruyter, 575-583
- Klünter, Dieter; Laser, Jochen (2008): *LDAP verstehen, OpenLDAP einsetzen. Grundlagen und Praxiseinsatz*. 2., überarb. und erw. Aufl. Heidelberg: dpunkt-Verl.
- Konst, Stefan (2000): *Sichere Log-Dateien auf Grundlage kryptographisch verketteter Einträge*. Braunschweig (2017): TU Braunschweig. Online verfügbar unter <http://publikationsserver.tu-braunschweig.de/get/64933>, zuletzt geprüft am 30.11.2017.
- Köwitsch, Philipp-Fernando (2016): *Validierung einer Deponiegasprognose für die Altablagerung Großziethen mit künstlichen neuronalen Netzen*. Bachelor Thesis.
- ks-webagentur.de (2017): Homepage: Website mit WordPress vorne bei den CMS-Systemen. Online verfügbar unter <http://ks-webagentur.de/website-mit->

- wordpress-vorne-bei-cms-systeme/, zuletzt aktualisiert am 27.11.2017, zuletzt geprüft am 12.12.2017.
- Kumaran, S.; Senthil (2017): Practical LXC and LXD. Linux Containers for Virtualization and Orchestration. Berkeley, CA: Apress. Online verfügbar unter <http://dx.doi.org/10.1007/978-1-4842-3024-4>, zuletzt geprüft am 30.11.2017.
- Larisch, Dirk (2000): Verzeichnisdienste im Netzwerk. NDS, active directory und andere. München, Wien: Hanser.
- Lave, Jean; Wenger, Etienne (1991): Situated learning. Legitimate peripheral participation. Cambridge: Cambridge Univ. Pr (Learning in doing).
- Lazinger, Susan Smernoff; Tibbo, Helen R. (2001): Digital preservation and metadata. History, theory, practice. Englewood, Colo.: Libraries Unlimited.
- LDS Umweltamt (2005): Landkreis Dahme-Spreewald Umweltamt. Online verfügbar unter <https://www.dahme-spreewald.info/sixcms/detail.php/526>, zuletzt geprüft am 30.09.2017.
- Lee, David Kuo Chuen; Deng, Robert (Hg.) (2018]-): Handbook of blockchain, digital finance, and inclusion. London: Academic Press.
- Lek, Sovan; Guégan, Jean-François (Hg.) (2000): Artificial Neuronal Networks. Berlin, Heidelberg: Springer (Environmental Science). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-642-57030-8>, zuletzt geprüft am 30.11.2017.
- Li, Chi-Yu; Molkenthin, Frank (2014): Time series scenario composition framework in Hydroinformatics Systems. Dissertation. BTU Cottbus - Senftenberg, Cottbus. Online verfügbar unter <http://nbn-resolving.de/urn:nbn:de:kobv:co1-opus4-33650>, zuletzt geprüft am 30.11.2017.
- Liebig, Thorsten (2013): Reasoning with OWL - system support and insights. Ulm: Universität Ulm. Fakultät für Ingenieurwissenschaften und Informatik (Ulmer Informatik-Berichte).
- Lin, Yu-Pin; Petway, Joy R.; Anthony, Johnathen; Mukhtar, Hussnain; Liao, Shih-Wei; Chou, Cheng-Fu; Ho, Yi-Fong (2017): Blockchain: The Evolutionary Next Step for ICT E-Agriculture. Basel: MDPI. Online verfügbar unter <http://www.mdpi.com/2076-3298/4/3/50/pdf>, zuletzt geprüft am 31.12.2018.
- Little, Roderick J. A.; Rubin, Donald B. (2014): Statistical Analysis with Missing Data. 2., Auflage. New York, NY: John Wiley & Sons (Wiley Series in Probability and Statistics).
- Llewellyn Smith, C.; Borysiewicz, L.; Casselton, L, Conway, G.; Hassan, M.; Leach, M. et al. (2011): Knowledge, networks and nations. Global scientific collaboration in the 21st century. London: The Royal Society (Policy document, 03/11). Online verfügbar unter http://royalsociety.org/uploadedFiles/Royal_Society_Content/Influencing_Polic

- y/Reports/2011-03-28-Knowledge-networks-nations.pdf, zuletzt geprüft am 30.11.2017.
- Ludwig, Jens; Enke, Harry (Hg.) (2013): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt, Göttingen: Hülsbusch; Niedersächsische Staats- und Universitätsbibliothek. Online verfügbar unter <http://resolver.sub.uni-goettingen.de/purl?isbn-978-3-86488-032-2>, zuletzt geprüft am 30.11.2017.
- Luther, Bernhard; Gnauck, Albrecht (2012): Zur Datenrekonstruktion in Zeitreihen von ökologischen Langzeituntersuchungen. In: Albrecht Gnauck (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2011 ; [15. Workshop "Modellierung und Simulation von Ökosystemen" ; 26.10.2011 - 28.10.2011]. Aachen: Shaker (Berichte aus der Umweltinformatik, 138), 49-59.
- Lutz, Peter (2009): Grundriss des Urheberrechts. Heidelberg: Müller (Jura auf den Punkt gebracht).
- Lynch, Clifford (1998): The Dublin Core Descriptive Metadata Program. Strategic Implications for Libraries and Networked Information Access. In: 0000214924. Online verfügbar unter <http://www.arl.org/newsltr/196/dublin.html>, zuletzt geprüft am 30.11.2017.
- Lynch, Clifford (2008): The Institutional Challenges of Cyberinfrastructure and E-Research. In: EDUCAUSE Review 2008 (Vol. 43, No. 6). Online verfügbar unter <http://er.educause.edu/~media/files/article-downloads/erm0867.pdf>, zuletzt geprüft am 16.08.2017.
- Massé, Mark (2012): REST API design rulebook. [designing consistent RESTful Web Service Interfaces]. Beijing: O'Reilly.
- Molkenthin, Frank (2000): WWW based hydroinformatics systems. Zugl.: Cottbus, Techn. Univ., Habil.-Schr., 2000. Cottbus: Inst. für Bauinformatik (Progress in Bauinformatik).
- Molkenthin, Frank (2004): Web-based Collaborative Engineering based on Information Sharing - HydroWeb. An Education Experiment in the Internet. Weimar: Professur Informatik im Bauwesen.
- mozilla.org (2017): Homepage: Mozilla Public Licence. Online verfügbar unter <https://www.mozilla.org/en-US/MPL/>, zuletzt geprüft am 07.12.2017.
- Müller, F.; Bergmann, R.; Dannowski, R.; Dippner, Joachim W.; Gnauck, A.; Haase, P. et al. (2016): Assessing resilience in long-term ecological data sets. In: Ecological indicators : integrating monitoring, assessment and management 65 (2016), S. 10–43. DOI: 10.1016/j.ecolind.2015.10.066.
- Müller, Felix; Baessler, Cornelia; Schubert, Hendrik; Klotz, Stefan (Hg.) (2010a): Long-Term Ecological Research. Between Theory and Application. Dordrecht: Springer Science+Business Media B.V. Online verfügbar unter

- <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10398879>, zuletzt geprüft am 30.11.2017.
- Müller, Felix; Gnauck, Albrecht; Wenkel, Karl-Otto; Schubert, Hendrik; Bredemeier, Michael (2010b): Theoretical Demands for Long-Term Ecological Research and Management of Long-Term Datasets. In: Felix Müller, Cornelia Baessler, Hendrik Schubert und Stefan Klotz (Hg.): Long-Term Ecological Research. Between Theory and Application. Dordrecht: Springer Science+Business Media B.V., S. 11–25.
- Müller, Felix; Leupelt, Maren (Hg.) (1998): Eco Targets, Goal Functions, and Orientors. ebrary, Inc. Berlin, New York: Springer. Online verfügbar unter <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10644722>, zuletzt geprüft am 30.11.2017.
- Munson, Matthew; Schöch, Christof; Rose, Sebastian; Scholger, Walter; Benardou, Agiati (2016): DARIAH-DE - Aufbau von Forschungsinfrastrukturen für die e-Humanities. Abschlussbericht : Laufzeit des Vorhabens: 01.03.2011 bis 29.02.2016. Version 1.0 vom 30.08.2016. Göttingen: SUB Göttingen. Online verfügbar unter <http://edok01.tib.uni-hannover.de/edoks/e01fb17/890522650.pdf>, zuletzt geprüft am 17.08.2017.
- Nagel, Wolfgang E.; Kröner, Dietmar H.; Resch, Michael M. (Hg.) (2016): High Performance Computing in Science and Engineering '16. Transactions of the High Performance Computing Center, Stuttgart (HLRS) 2016. Cham: Springer. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-319-47066-5>, zuletzt geprüft am 30.11.2017.
- Nakamoto, Satoshi (2008): Bitcoin: A Peer-to-Peer Electronic Cash System. Hg. v. Bitcoin.org. Online verfügbar unter <https://bitcoin.org/bitcoin.pdf>, zuletzt geprüft am 28.12.2017.
- National Science Foundation (2007): Cyberinfrastructure Vision for 21st Century. (NSF 07-28). Online verfügbar unter <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp>, zuletzt geprüft am 16.08.2017.
- Nelson, Bryn (2009): Data sharing. Empty archives. In: Nature 461 (7261), S. 160–163. DOI: 10.1038/461160a.
- Neumann, Michael (2003): Ökologische Informatik. In: Forum Geoökologie 2003 (14-1), S. 4–6. Online verfügbar unter http://dr-michael-neumann.de/pdf/MN_2003_Forum_d_Geoec_14_1.pdf, zuletzt geprüft am 16.08.2017.
- Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, M. (Hg.) (2009): Nestor Handbuch - Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung [im Rahmen des Projektes: Nestor - Kompetenznetzwerk Langzeitarchivierung und

- Langzeitverfügbarkeit digitaler Ressourcen für Deutschland]. Version 2.0, Juni 2009. Boizenburg: Hülsbusch.
- Neuroth, Heike; Strathmann, Stefan; Oßwald, Achim; Scheffel, Regine; Klump, Jens; Ludwig, Jens (Hg.) (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland. Boizenburg, Göttingen: Hülsbusch; Univ.-Verl. Göttingen.
- Notay, Kunwar Vikramjeet Singh (2015): Model coupling in hydroinformatics systems through the use of autonomous tensor objects. Dissertation. BTU Cottbus - Senftenberg, Cottbus. Online verfügbar unter <http://d-nb.info/1114284149/34>, zuletzt geprüft am 30.11.2017.
- Nowakowski, Piotr; Ciepiela, Eryk; Harężlak, Daniel; Kocot, Joanna; Kasztelnik, Marek; Bartysiński, Tomasz et al. (2011): The Collage Authoring Environment. In: *Procedia Computer Science* 4, S. 608–617. DOI: 10.1016/j.procs.2011.04.064.
- Odum, Eugene Pleasants (1967): *Ökologie*. München, Basel, Wien: Bayerischer Landwirtschaftsverl. (Moderne Biologie).
- Oey, Kai Jan (2012): *Softwarearchitektur heute. Nutzen und Kosten serviceorientierter Architekturen*. neue Ausg. Saarbrücken: AV Akademikerverlag.
- Open Archives Initiative (2017): OAI Protocol for Metadata Harvesting - Homepage. Open Archives Initiative. Online verfügbar unter <http://openarchives.org/>, zuletzt aktualisiert am 22.02.2017, zuletzt geprüft am 30.08.2017.
- Open Geospatial Consortium (2017): Homepage. Open Geospatial Consortium (OGC). Online verfügbar unter <http://www.opengeospatial.org/>, zuletzt geprüft am 15.08.2017.
- OpenMI.org (2017): Homepage: OpenMI.org. Online verfügbar unter <http://www.openmi.org/>, zuletzt aktualisiert am 19.10.2017, zuletzt geprüft am 02.11.2017.
- opensource.org (2017): Homepage: Open Source Initiative. Online verfügbar unter <https://opensource.org>, zuletzt aktualisiert am 19.09.2017, zuletzt geprüft am 19.09.2017.
- O'Reilly, Tim (2007): What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. In: *O'Reilly, Communications & Strategies* (1), S. 17. Online verfügbar unter <https://ssrn.com/abstract=1008839>, zuletzt geprüft am 16.08.2017.
- Oxford Digital Library (2009): Metadata in the Oxford Digital Library. Online verfügbar unter <http://www.odl.ox.ac.uk/metadata.htm>, zuletzt aktualisiert am 26.03.2009, zuletzt geprüft am 19.08.2017.

-
- Ozturan, Can (2016): Blockchain Infrastructure for E-Science. digitalinfrastructures.eu. Online verfügbar unter <https://www.digitalinfrastructures.eu/content/blockchain-infrastructure-e-science>, zuletzt geprüft am 31.12.2017.
- Page, Bernd; Hilty, Lorenz M. (Hg.) (1995): Umweltinformatik. Informatikmethoden fuer Umweltschutz und Umweltforschung. 2., aktualisierte und erw. Aufl. München: Oldenbourg (Handbuch der Informatik, 13.3).
- Pape, Dave (2001): The CAVE Virtual Reality System. Online verfügbar unter <https://www.evl.uic.edu/pape/CAVE/>, zuletzt geprüft am 15.08.2017.
- Pausch, Julia (2017): CMS als kostensparende Alternative zur Erstellung von Webseiten. Wie arbeitet Joomla! und welche Vorteile bietet das Content-Management-System für kleinere und mittlere Firmen. 1. Auflage. Saarbrücken: AV Akademikerverlag.
- Pedrycz, Witold; Chen, Shyi-Ming (Hg.) (2017): Data Science and Big Data. An Environment of Computational Intelligence. Springer International Publishing AG. 1st edition 2017. Cham: Springer International Publishing (Studies in Big Data).
- Perske, Rainer (2012): Single Sign-On (SSO). In: Fortschritte des integrierten Informationsmanagements an Hochschulen: integrierte Bereitstellung, einheitlicher Zugang und individuelle Verteilung; Tagungsband des Abschlussworkshops zum Projekt MIRO - Münster Information System for Research and Organization, Münster, 3.11.2010. Münster: Verl.-Haus Monsenstein und Vannerdat, S. 142–151.
- Pillmann, Werner; Gnauck, Albrecht (2012): Einbindung von Ökosystem- und Umweltmodellen in die Nachhaltigkeitsmodellierung. In: Albrecht Gnauck (Hg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2011 ; [15. Workshop "Modellierung und Simulation von Ökosystemen" ; 26.10.2011 - 28.10.2011]. Aachen: Shaker (Berichte aus der Umweltinformatik, 138), S. 1–11.
- Pillmann, Werner; Schade, Sven; Smits, Paul (2011): Environmental informatics. Aachen: Shaker (Innovations in sharing environmental observations and information, EnviroInfo Ispra 2011; proceedings of the 25th International Conference EnviroInfo, October 5-7, 2011, Joint Research Centre Ispra Institute for Environment and Sustainability / W. Pillmann, S. Schade, P. Smits (eds.) ; Part 1).
- Pollock, D. S.G.; Green, Richard C.; Nguyen, Truong (1999): Handbook of Time Series Analysis, Signal Processing, and Dynamics. 1. Aufl. s.l.: Elsevier professional (Signal Processing and its Applications). Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=330175>, zuletzt geprüft am 30.11.2017.
- Popst, Hans (Hg.) (1993): Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken. RAK-WB. Deutsches Bibliotheksinstitut. Berlin: Dt. Bibliotheksinst.

- Portnoy, Matthew (2016): *Virtualization Essentials*. 2nd ed. Newark: Wiley. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=4644086>, zuletzt geprüft am 30.11.2017.
- Preve, Nikolaos P. (Hg.) (2013): *Grid Computing. Towards a Global Interconnected Infrastructure*. Guildford, Surrey: Springer London (Computer Communications and Networks).
- Puppe, Frank (1993): *Systematic Introduction to Expert Systems. Knowledge Representations and Problem-Solving Methods*. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-642-77971-8>, zuletzt geprüft am 30.11.2017.
- Qiu, Meikang (Hg.) (2017): *The Third IEEE International Conference on Big Data Security on Cloud - IEEE BigDataSecurity 2017, the Third IEEE International Conference on High Performance and Smart Computing - IEEE HPSC 2017, the Second IEEE International Conference on Intelligent Data and Security - IEEE IDS 2017. Proceedings : 26-28 May 2017, Beijing, China. IEEE International Conference on High Performance and Smart Computing; IEEE International Conference on Intelligent Data and Security; Institute of Electrical and Electronics Engineers; IEEE International Conference on Big Data Security on Cloud; IEEE BigDataSecurity; IEEE HPSC; IEEE IDS*. Piscataway, NJ: IEEE. Online verfügbar unter <http://ieeexplore.ieee.org/servlet/opac?punumber=7976094>, zuletzt geprüft am 30.11.2017.
- Raj, Pethuru; Raman, Anupama; Nagaraj, Dhivya; Duggirala, Siddhartha (2016): *High-Performance Big-Data Analytics. Computing Systems and Approaches*. Softcover reprint of the original 1st edition 2015. Cham: Springer International Publishing; Springer (Computer Communications and Networks).
- Recknagel, Friedrich (Hg.) (2006): *Ecological informatics. Scope, techniques and applications ; with a CD-ROM*. 2. ed. Berlin, Heidelberg, New York: Springer.
- Reinsel, David; Gantz, John; Rydning, John (2017): *Data Age 2025: The Evolution of Data to Life-Critical*. Hg. v. IDC White Paper. SEAGATE. Online verfügbar unter <http://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>, zuletzt geprüft am 15.08.2017.
- Renn, Ortwin; Deuschle, Jürgen; Jäger, Alexander; Weimer-Jehle, Wolfgang (2007): *Leitbild Nachhaltigkeit. Eine normativ-funktionale Konzeption und ihre Umsetzung*. Wiesbaden: VS Verlag für Sozialwissenschaften | GWV Fachverlage GmbH Wiesbaden. Online verfügbar unter <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10223387>, zuletzt geprüft am 30.11.2017.
- Resch, Michael; Keller, Rainer; Himmler, Valentin (2008): *Tools for High Performance Computing. Proceedings of the 2nd International Workshop on Parallel Tools for High Performance Computing, July 2008, HLRS, Stuttgart*. 1. Aufl. s.l.: Sprin-

- ger-Verlag. Online verfügbar unter <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10284411>, zuletzt geprüft am 30.11.2017.
- Research Councils UK (2005): RCUK Position statement on access to research outputs. Research Councils UK. Online verfügbar unter <http://www.rcuk.ac.uk/documents/documents/2005statement-pdf/>, zuletzt geprüft am 16.08.2017.
- Ritschl, Valentin; Weigl, Roman; Stamm, Tanja (Hg.) (2016): *Wissenschaftliches Arbeiten und Schreiben. Studium Pflege, Therapie, Gesundheit*. Heidelberg: Springer-Verlag Berlin Heidelberg 2016, DOI 10.1007/978-3-662-49908-5_2
- Rogall, Holger (2002): *Neue Umweltökonomie - Ökologische Ökonomie. Ökonomische und ethische Grundlagen der Nachhaltigkeit, Instrumente zu ihrer Durchsetzung*. Wiesbaden, s.l.: VS Verlag für Sozialwissenschaften. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-322-99733-3>, zuletzt geprüft am 30.11.2017.
- Rosina, Peter Christian (2016): *Semantic Web Technologies for Method Management. Model, Methodology and Enterprise Architecture Integration*. Augsburg, Universität Augsburg, Diss., 2016. Augsburg: Universität Augsburg.
- Roure, D. de; Goble, C.; Stevens, R. (2009): The design and realisation of the myExperiment Virtual Research Environment for social sharing of workflows. In: *Future Generation Computer Systems* (25), S. 561–567.
- Ruchter, Markus (2002): *Web-basierte Umweltinformationssysteme für die Öffentlichkeit. Konzepte, Potentiale und Realisierbarkeit*. Als Ms. gedr. Karlsruhe: Forschungszentrum (Wissenschaftliche Berichte / Forschungszentrum Karlsruhe, Technik und Umwelt, FZKA 6760).
- Runkler, Thomas A. (2015): *Data Mining. Modelle und Algorithmen intelligenter Datenanalyse*. 2., aktualisierte Auflage. Wiesbaden: Springer Vieweg (Computational Intelligence).
- Rusch-Feja, Diann (2002): The Open Archives Initiative and the OAI protocol for metadata harvesting. Rapidly forming a new tier in the scholarly communication infrastructure. *Literaturverz.* S. 184-186. In: *Learned Publishing*. Online verfügbar unter <http://isacco.catchword.com/vl=10161948/cl=27/nw=1/fm=docpdf/rpsv/catchword/alpsp/09531513/v15n3/s4/p179>, zuletzt geprüft am 30.11.2017.
- Russell, Stuart; Norvig, Peter; Kirchner, Frank (2012): *Künstliche Intelligenz. Ein moderner Ansatz*. 3., aktualisierte Aufl. München: Pearson Higher Education (Always learning).
- Sakamoto, M. (1966): Primary production by phytoplankton community in some Japanese lakes and its dependence on lake depth. In: *Archiv für Hydrobiologie* (62), S. 1–28.

- Sarbadhikari, Haimanti; Srivastava, Shashi Mohan (2017): A Course on Basic Model Theory. Singapore: Springer Singapore. Online verfügbar unter <http://dx.doi.org/10.1007/978-981-10-5098-5>, zuletzt geprüft am 30.11.2017.
- Schicker, Edwin (2017): Datenbanken und SQL. Eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL. 5., aktualisierte und erweiterte Auflage. Wiesbaden: Springer Vieweg (Lehrbuch).
- Schroeder, Kathrin (2001): Persistent Identifier in "Der Deutschen Bibliothek" im Rahmen des Projektes CARMEN-AP4. Frankfurter Buchmesse 2001. Leipzig, Frankfurt, M., Berlin: Die Dt. Bibliothek.
- Schroeder, Kathrin (2004): "Technische Aspekte der Langzeitarchivierung". Persistent Identifier (PI) im Kontext der Langzeitarchivierung - State of the Art ; Bibliothekskongress Leipzig, 24.03.2004. Leipzig, Frankfurt, M., Berlin: Die Dt. Bibliothek.
- shibboleth.net (2017): Homepage: OpenSAML 2 - Shibboleth Wiki. Online verfügbar unter <https://wiki.shibboleth.net/confluence/display/OpenSAML/Home>, zuletzt geprüft am 19.12.2017.
- Sicilia, Miguel-Angel (Hg.) (2014): Handbook of metadata, semantics and ontologies. A to <A>. New Jersey: World Scientific.
- Snowdon, David N. (Hg.) (2004): Inhabited information spaces. Living with your data. London: Springer (Computer supported cooperative work).
- Spisla, David (2014): Medien zur Langzeitspeicherung elektronischer Daten (Bitstream Preservation). München: GRIN Verlag GmbH.
- Stallmann, Richard (2017): Warum Open Source das Ziel Freie Software verfehlt - GNU.org. Online verfügbar unter <https://www.gnu.org/philosophy/open-source-misses-the-point>, zuletzt geprüft am 19.09.2017.
- Steigerwald, Heiko (2013): XML-basierte Standards für Digitale (Multimediale) Medien und Metadaten. München: GRIN Verlag GmbH.
- Straškraba, Milan; Gnauck, Albrecht (1983): Aquatische Ökosysteme. Modellierung u. Simulation. Stuttgart, New York: Fischer.
- Strecker, Michael (2010): Potenzial von Single Sign-on bei Webapplikationen. Eine Analyse anhand Java-basierter Lösungen. München: GRIN Verlag GmbH.
- Süptitz, Thomas; Weis, Stephan J. J.; Eymann, Torsten (2013): Was müssen Virtual Research Environments leisten? – Ein Literaturreview zu den funktionalen und nichtfunktionalen Anforderungen. Online verfügbar unter <http://www.wi2013.de/proceedings/WI2013%20-%20Track%203%20-%20Sueptitz.pdf>, zuletzt geprüft am 16.08.2017.

-
- Tarr, Peri L.; Wolf, Alexander L. (Hg.) (2011): Engineering of Software. The Continuing Contributions of Leon J. Osterweil. Unter Mitarbeit von Leon J. Osterweil. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Taylor, Ian J. (Hg.) (2007): Workflows for e-Science. Scientific workflows for Grids. London: Springer.
- Terzo, Olivier; Mossucca, Lorenzo (Hg.) (2015): Cloud Computing with e-Science Applications. Boca Raton, Fla.: CRC Press.
- The council of the european communities (1992): Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Online verfügbar unter <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31992L0043>, zuletzt geprüft am 15.08.2017.
- Thibodeau, Kenneth (2002): Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years¹. Online verfügbar unter <https://www.clir.org/pubs/reports/pub107/thibodeau.html>, zuletzt geprüft am 17.08.2017.
- Thurnherr, Daniela (2003): Öffentlichkeit und Geheimhaltung von Umweltinformationen. Weiterentwicklung des Umweltvölkerrechts durch die Aarhus-Konvention und deren Bedeutung für das schweizerische Recht. Zugl.: Zürich, Univ., Diss., 2003. Zürich: Schulthess (Zürcher Studien zum öffentlichen Recht, 159).
- Tietz, Christian; Pelchen, Chris; Meinel, Christoph; Schnjakin, Maxim (2017): Management digitaler Identitäten. Aktueller Status und zukünftige Trends. Potsdam: Universitätsverlag (Technische Berichte des Hasso-Plattner-Instituts für Softwaresystemtechnik an der Universität Potsdam, 114). Online verfügbar unter <http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-103164>, zuletzt geprüft am 30.11.2017.
- toolkit.globus.org (2017): Homepage: Globus: OGSA - The Open Grid Services Architecture. Online verfügbar unter <http://toolkit.globus.org/ogsa/>, zuletzt geprüft am 10.12.2017.
- turnkeylinux.org (2017): Homepage: TurnKey GNU/Linux: 100+ free ready-to-use system images for virtual machines, the cloud, and bare metal. Online verfügbar unter <https://www.turnkeylinux.org/>, zuletzt geprüft am 12.12.2017.
- United Nations University (2006): Study on the: Economic impact of open source software on innovation and the competitiveness of the Information and Communication Technologies (ICT) sector in the EU. Online verfügbar unter <https://www.campussource.de/opensource/docs/FLOSSImpactOnEU.pdf>, zuletzt geprüft am 19.09.2017.
- van Gorp, Pieter; Mazanek, Steffen (2011): SHARE. A web portal for creating and sharing executable research papers. In: Procedia Computer Science 4, S. 589–597. DOI: 10.1016/j.procs.2011.04.062.

- Vincent, Chung-Wei Lin (2017): Content, Management, System Aufbau eines Bewertungsprototyps für CMS aus Kommunikativen Perspektiven. Online verfügbar unter http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000006973/002_CMS02GERMANSUMMARY.pdf?hosts, zuletzt geprüft am 22.09.2017.
- Volk, Markus (2011): Open Linked Data, Open Government Data Sets. München: GRIN Verlag GmbH.
- Vu, Kim-Phuong L.; Proctor, Robert W. (Hg.) (2011): Handbook of human factors in web design. 2. ed. Boca Raton, Fla.: CRC Press (Human factors and ergonomics).
- W3C (2017): World Wide Web Consortium (W3C) - Homepage. Online verfügbar unter <https://www.w3.org/>, zuletzt aktualisiert am 18.08.2017, zuletzt geprüft am 19.08.2017.
- Wang, Fei-Yue; Carley, Kathleen M.; Zeng, Daniel; Mao, Wenji (2007): Social Computing. From Social Informatics to Social Intelligence. In: IEEE Intell. Syst. 22 (2), S. 79–83. DOI: 10.1109/MIS.2007.41.
- Watzke, Christian (2017): DFG Sonderforschungsbereiche - Universität Heidelberg. Hg. v. Universität Heidelberg. Online verfügbar unter <https://www.uni-heidelberg.de/einrichtungen/forschung/sonderforschungsbereiche.html>, zuletzt geprüft am 23.09.2017.
- Weerawarana, Sanjiva (2008): Web services platform architecture. SOAP, WSDL, WS-policy, WS-addressing, WS-BPEL, WS-reliable messaging and more. 5. printing. Upper Saddle River, N.J.: Prentice Hall/PTR.
- Wenger, E. (2000): Communities of Practice. Learning, Meaning, and Identity: Cambridge University Press (Learning in Doing: Social, Cognitive and Computational Perspectives). Online verfügbar unter <https://books.google.de/books?id=Jb8mAAAAQBAJ>, zuletzt geprüft am 30.11.2017.
- Wenz, Christian; Hauser, Tobias (2004): Web Services mit PHP. [inkl. XML-RPC, SOAP, WSDL, UDDI ; mit PHP Web Services implementieren ; technische Grundlagen und Praxisbeispiele]. 1. Aufl. Bonn: Galileo Press (Galileo Computing).
- Wiegand, Gerhard (1996): Konzepte der Hierarchietheorie in der Ökologie. In: Karin Mathes, Broder Breckling und Klemens Ekschmitt (Hg.): Systemtheorie in der Ökologie. Beiträge zu einer Tagung des Arbeitskreises "Theorie" in der Gesellschaft für Ökologie: Zur Entwicklung und aktuellen Bedeutung der Systemtheorie in der Ökologie, Schloß Rauischholzhausen im März 1996. 1. Aufl. Landsberg: ecomed (Angewandte Naturwissenschaften), S. 7–24.

-
- Wilkins-Diehr, Nancy (Hg.) (2007): Science gateways - common community interfaces to grid resources. [presentations and discussions at the Global Grid Forum GGF14 Science Gateways Workshop in Chicago, IL, June 2005]. Global Grid Forum GGF14 Science Gateways Workshop. Chichester: Wiley (Concurrency and computation Special issue, 19.2007, 6).
- Wissenschaftsrat (2011): Übergreifende Empfehlungen zu Informationsinfrastrukturen. Berlin. Online verfügbar unter <https://www.wissenschaftsrat.de/download/archiv/10466-11.pdf>, zuletzt geprüft am 20.09.2017.
- Witt, Andreas; Metzinger, Dieter (Hg.) (2010): Linguistic modeling of information and markup languages. Contributions to language technology. Dordrecht: Springer (Text, speech and language technology, 40).
- Wohlgemuth, Volker (Hg.) (2009): Environmental informatics and industrial environmental protection. Concepts, methods and tools. Aachen: Shaker.
- Woodward, Guy (Hg.) (2014): Big data in ecology. 1. ed. Amsterdam: Elsevier AP (Advances in ecological research, 51).
- Wu, Zhaohui; Chen, Hua-jun (2008): Semantic grid. Model, methodology, and applications. Berlin, Heidelberg, Hangzhou: Springer; Zhejiang Univ. Press (Advanced topics in science and technology in China).
- Wulf, A. (1993): The collaboratory oppurtunity. In: Science (261), S. 854–855.
- Yang, Xiaoyu; Wang, Lizhe; Jie, Wei (Hg.) (2013): Guide to e-Science. Next Generation Scientific Research and Discovery. 1., 2011. Guildford, Surrey: Springer London (Computer Communications and Networks).
- Zächerl, Sebastian (2007): Semantic Web - RDF DAML + OIL. München: GRIN Verlag GmbH.
- ZKI (2017): ZKI - Zentren für Kommunikation und Informationsverarbeitung e.V.: Startseite. Online verfügbar unter <https://www.zki.de/>, zuletzt geprüft am 20.09.2017.
- Zurth, Patrick (2015): Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht. Dissertation.

Anhang

A Vor- und Nachteile von Open-Source-Software

Tabelle A.1: Vorteile von Open-Source-Software (Fritzlar et al. 2012)

Bereich	Vorteile
Offenheit, Freiheit, Transparenz	<ul style="list-style-type: none"> ▪ freie Nutzungsrechte ▪ Unabhängigkeit vom Hersteller ▪ offener Quellcode ▪ Beeinflussung künftiger Produktausrichtung ▪ offene Formate & Schnittstellen ▪ nachhaltige Informationsverwertbarkeit ▪ Plattformunabhängigkeit
Produktivität, Innovation	<ul style="list-style-type: none"> ▪ marktorientierte Innovationszyklen ▪ Erweiterbarkeit und Skalierbarkeit
Qualität	<ul style="list-style-type: none"> ▪ weniger anfällig für Viren oder Schadcode, jedoch attraktiver für Angriffe ▪ schnelle Behebung von Sicherheitslücken und Fehlern ▪ geringere Fehlerrate
Wirtschaftlichkeit	<ul style="list-style-type: none"> ▪ geringe Umsetzungszeit ▪ Aufwandsbündelung und Mehrfachnutzung ▪ keine Lizenzkosten ▪ Wissen ist auf breite Basis verteilt und unabhängig von Einzelunternehmen ▪ kein Upgrade-Zwang
Markt & Wettbewerb	<ul style="list-style-type: none"> ▪ Vermeidung von Monopolen ▪ Einbezug von KMU sowie regionaler Unternehmen ▪ Stärkung der Verhandlungsbasis gegenüber kommerziellen Anbietern ▪ größerer Wettbewerb
Außenwirkung	<ul style="list-style-type: none"> ▪ Akzeptanz der Bürger ▪ Imagegewinn durch Einsatz innovativer Technologien und effizienten Einsatz von Steuergeldern ▪ keine Nötigung der Bürger zur Nutzung und zum Erwerb proprietärer Software

Tabelle A.2: Nachteile von Open-Source-Software (Fritzlar et al. 2012)

Bereich	Nachteile
Kompatibilität	<ul style="list-style-type: none"> ▪ Hardwarekompatibilität bei spezieller Hardware z.B. unter dem Betriebssystem „Linux“ ▪ Abwärtskompatibilität ist nicht immer vorhanden ▪ Fachanwendungen gehören meist zu proprietärer Software
Recht und Haftung	<ul style="list-style-type: none"> ▪ Gewährleistung und Haftung durch Lizenzgeber ist nicht automatisch vorhanden ▪ unbewusste potentielle Verletzung von Softwarepatenten ▪ Anzahl, Unterschiede, Kombination von Lizenzen
Nutzung	<ul style="list-style-type: none"> ▪ Schulungsaufwand kann hoch sein ▪ eingeschränkte FLOSS Lösungen oder „Pseudo-FLOSS“ ▪ Akzeptanz der Anwender
Dienstleistungen	<ul style="list-style-type: none"> ▪ nicht für alle FLOSS steht ein flächendeckendes Supportnetz bereit

B Tabellen für die VRE-Beispielprojekte

Tabelle B.1: Komprimiertes „Rapid L-TER VRE“ Konzept für die Altablagerung Großziethen

„Rapid L-TER VRE“ Bestandteile	Beschreibung, (*) Zusätzliche Erweiterungen
Open Source	VRE wird mit Basis-Open-Source-Software implementiert: <ul style="list-style-type: none"> ▪ (Turnkey) Linux Web-/Datenbank Server mit „Joomla“ (CMS) und diversen Erweiterungen. ▪ „GeoNetwork“ (Metadatenkatalog und WebGIS) ▪ (Turnkey) „MediaWiki“ mit Visual Editor Erweiterung (schlüsselselfertige „Wikipedia“- Software für Teamwork an Dokumenten) ▪ (Turnkey) „OpenLDAP“ (schlüsselselfertiger Verzeichnissever)
ICT-Basistechnologien	Direkte Webtechnologien (HTML, http, XML, Web 2.0, etc.) <ul style="list-style-type: none"> ▪ Servertechnologien (Betriebssysteme, OS Services, etc.) ▪ Technologien zur Authentifizierung / SSO (Verzeichnissever, Kryptographie, Token, etc.) ▪ Datenbanken und Datenbank Management Systeme: MySQL (CMS und andere Services), PostgreSQL (GIS-Daten) ▪ Content Management Systeme (CMS) ▪ Programmiersprachen (z.B. Java, Java-Script, Python, Perl, C, PHP, etc.) ▪ Metadaten und Auszeichnungssprachen ▪ Semantik Web Technologien
Schlüsselselfertige Server und Virtualisierung	<ul style="list-style-type: none"> ▪ Installation von „Virtual-Box“ (Virtualisierung) und Erzeugung Virtueller Maschinen (VM) ▪ Installation und Konfiguration von Schlüsselselfertigen Servern auf den VM ▪ Start-/Stopp-Scripte für VM-Steuerung ▪ Backup-Scripte für automatische VM-Snapshots und -Backups
Homepage und CMS	<ul style="list-style-type: none"> ▪ „Joomla“ CMS mit Erweiterungen ▪ Installation Basissystem ▪ Installation der Erweiterungen ▪ Konfiguration ▪ Design und Layout ▪ Optimierung für Mobilgeräte ▪ Struktur (Menü und andere Gliederungen) ▪ Integration der nicht zum CMS gehörenden Service und Tools ▪ Basis-Inhalte einpflegen ▪ CMS-Systemsicherheit und -Backups ▪ Suchmaschinenoptimierung
Services und Tools	<ul style="list-style-type: none"> ▪ klassische CMS-Funktionalitäten („Joomla“) ▪ Social-Media-Community (CMS-Erweiterung)*

	<ul style="list-style-type: none"> ▪ Wiki („MediaWiki“ mit „VisualEditor“, etc.) ▪ Metadaten-Management, WebGIS, (Meta)Data-Harvesting („GeoNetwork“) ▪ Statistik Server (R-Statistics)*
Tools (Middleware)	<ul style="list-style-type: none"> ▪ klassische CMS-Funktionalitäten ▪ Automatisierung des Forschungsprozesses und der Workflows (allgemein) ▪ Langzeitdatenspeicherung (Server-Tools, individuell)* ▪ Oberfläche für SQL-Datenbank (CMS-Erweiterung) ▪ Persistente Identifikatoren („GeoNetwork“) ▪ Downloads („GeoNetwork“, CMS) ▪ Suche („GeoNetwork“, CMS) ▪ Datenschema („GeoNetwork“) ▪ Daten Pre- und Postprocessing (Programmiersprachen)* ▪ Zeitreihenanalyse (Programmiersprachen, CMS-Erweiterung)* ▪ Prognose der Konzentrationen mit Szenarien (Programmiersprachen)* ▪ automatische Synchronisation von „GeoNetwork“ Metadaten mit relevanten übrigen Metadaten (Programmiersprachen, CMS-Erweiterung)* ▪ digitale Verbindung aller relevanter Sensoren und Daten mit der VRE (CMS-Erweiterungen, Programmiersprachen)* ▪ Umfragen (CMS-Erweiterung)* ▪ Publikationsverzeichnis (CMS)* ▪ automatische Reports (Programmiersprachen)* ▪ KI-Modul zur Gasprognose (Programmiersprachen)* ▪ Suchmaschinenoptimierung (CMS-Erweiterungen) ▪ gemeinsame Erstellung von Dokumenten („MediaWiki“)
Authentifizierung und SSO	<ul style="list-style-type: none"> ▪ „OpenLDAP“ als Verzeichnisserver zur Benutzerverwaltung ▪ alle Softwarekomponenten können mit der LDAP Schnittstelle umgehen ▪ kleiner Nutzerkreis: Pseudo SSO durch Cookies (Angemeldet bleiben). Alternativ eine richtige SSO verbunden mit deutlich Mehraufwand in der Implementation*.
Report-Layer	<ul style="list-style-type: none"> ▪ Report für Zielgruppen aufbereitet durch Tools realisieren ▪ (Meta)Data-Harvesting über die eingebaute OAI-Schnittstelle von „GeoNetwork“.
Hilfe, Feedback und Dokumentation.	<ul style="list-style-type: none"> ▪ Wiki für: Benutzerhilfe, Dokumentation für Administratoren*, Dokumentation für Entwickler* ▪ Umfragen unter der Bevölkerung (CMS-Erweiterungen)* ▪ Benutzerfeedback (CMS-Erweiterungen) ▪ FAQ (CMS-Erweiterungen, „MediaWiki“)
Zusätzliche Erweiterungen, Methoden und Technologien	<ul style="list-style-type: none"> ▪ Alle mit „*“ markierten Elemente <p>Generell sind folgende Gebiete für die spezielle VRE von Interesse:</p>

- Langzeitdatenspeicherung
- Visualisierungstechnologien
- Datamining
- Monitoring / Sensornetzwerke
- Modellierung, Simulation, Prognose
- KI Methoden für mehrere Bereiche

Tabelle B.2: Komprimiertes „Rapid L-TER VRE“ Konzept für „Hühnerwasser“

„Rapid L-TER VRE“ Bestandteile	Beschreibung, (*) Zusätzliche Erweiterungen
Open Source	VRE wird mit Basis-Open-Source-Software implementiert: <ul style="list-style-type: none"> ▪ (Turnkey) Linux Web-/Datenbank Server mit „Joomla“ (CMS) und diversen Erweiterungen ▪ „GeoNetwork“ (Metadatenkatalog und WebGIS) ▪ (Turnkey) „MediaWiki“ mit Visual Editor Erweiterung (schlüsselfertige „Wikipedia“- Software für Teamwork an Dokumenten) ▪ (Turnkey) „OpenLDAP“ (schlüsselfertiger Verzeichnissever) ▪ (Turnkey) „Collabative“ (schlüsselfertiges Projektmanagement) ▪ (Turnkey) „OwnCloud“ (schlüsselfertiger Cloud-Service) ▪ „R-Statistics“ Server mit „Shiny“ Web-Oberfläche
ICT-Basistechnologien	<ul style="list-style-type: none"> ▪ Direkte Webtechnologien (HTML, http, XML, Web 2.0, etc.) ▪ Servertechnologien (Betriebssysteme, OS Services, etc.) ▪ Technologien zur Authentifizierung / SSO (Verzeichnissever, Kryptographie, Token, etc.) ▪ Datenbanken und Datenbank Management Systeme: MySQL (CMS und Services), PostgreSQL (GIS Daten), NoSQL (Sensordaten und verschiedene unstrukturierte Daten) ▪ Content Management Systeme (CMS) ▪ Programmiersprachen (z.B. Java, Java-Script, Python, Perl, C, PHP, etc.) ▪ Metadaten und Auszeichnungssprachen ▪ Semantik Web Technologien ▪ Remote-Desktop-Services ▪ Cloud Computing
Schlüsselfertige Server und Virtualisierung	<ul style="list-style-type: none"> ▪ Installation von „Virtual-Box“ (Virtualisierung) und Erzeugung Virtueller Maschinen (VM) ▪ Installation und Konfiguration von schlüsselfertigen Servern auf den VM ▪ Start-/Stopp-Scripte für VM-Steuerung ▪ Backup-Scripte für automatische VM-Snapshots und -Backups
Homepage und CMS	<ul style="list-style-type: none"> ▪ „Joomla“ CMS mit Erweiterungen: ▪ Installation Basissystem

	<ul style="list-style-type: none"> ▪ Installation der Erweiterungen ▪ Konfiguration ▪ Design und Layout ▪ Optimierung für Mobilgeräte ▪ Struktur (Menü und andere Gliederungen) ▪ Integration der nicht zum CMS gehörenden Service und Tools ▪ Basis-Inhalte einpflegen ▪ CMS-Systemsicherheit und -Backups ▪ Suchmaschinenoptimierung
Services	<ul style="list-style-type: none"> ▪ klassische CMS-Funktionalitäten („Joomla“) ▪ Social-Media-Community (CMS-Erweiterung)* ▪ Wiki („MediaWiki“ mit „VisualEditor“, etc.) ▪ Metadaten-Management ▪ Projektmanagement ▪ Remote-Desktop-Services ▪ Cloud-Services (OwnCloud) ▪ Screen-Sharing, Videokonferenzen ▪ Statistik-Service
Tools (Middleware)	<ul style="list-style-type: none"> ▪ klassische CMS-Funktionalitäten ▪ interaktive Auswertung von Forschungsdaten ▪ Datenaufbereitung ▪ Einbindung mobiler Technologien ▪ Sensornetzwerke ▪ Automatisierung des Forschungsprozesses und der Workflows ▪ Langzeitdatenspeicherung (Server-Tools, individuell)* ▪ Downloads („GeoNetwork“, CMS) ▪ Suche („GeoNetwork“, CMS) ▪ Datenschema („GeoNetwork“) ▪ Automatische Synchronisation von „GeoNetwork“ Metadaten mit relevanten übrigen Metadaten (Programmiersprachen, CMS-Erweiterung)* ▪ Suchmaschinenoptimierung (CMS-Erweiterungen)
Authentifizierung und SSO	<ul style="list-style-type: none"> ▪ „OpenLDAP“ als Verzeichnisserver zur Benutzerverwaltung mit SSO*
Report-Layer	<ul style="list-style-type: none"> ▪ Report für Zielgruppen aufbereitet durch Tools realisieren* ▪ (Meta)Data-Harvesting über die eingebaute OAI-Schnittstelle von „GeoNetwork“*.
Hilfe, Feedback und Dokumentation.	<ul style="list-style-type: none"> ▪ Wiki für: Benutzerhilfe, Dokumentation für Administratoren*, Dokumentation für Entwickler* ▪ Benutzerfeedback (CMS-Erweiterungen) ▪ FAQ (CMS-Erweiterungen, „MediaWiki“)
Zusätzliche Erweiterungen, Methoden und Technologien	<ul style="list-style-type: none"> ▪ Alle mit „*“ markierten Elemente <p>Generell sind folgende Gebiete für die spezielle VRE von Interesse:</p>

- Langzeitdatenspeicherung
- Visualisierungstechnologien
- Datamining
- Web-Labor für: Modellierung, Simulation, Prognose

Tabelle B.3: Komprimiertes „Rapid L-TER VRE“ Konzept für „HydroEurope“

„Rapid L-TER VRE“ Bestandteile	Beschreibung, (*) Zusätzliche Erweiterungen
Open Source	VRE wird mit Basis-Open-Source-Software implementiert: <ul style="list-style-type: none"> ▪ (Turnkey) Linux Web-/Datenbank Server mit „Joomla“ (CMS) und diversen Erweiterungen. ▪ „GeoNetwork“ (Metadatenkatalog und WebGIS) ▪ (Turnkey) „MediaWiki“ mit Visual Editor Erweiterung (schlüsselfertige „Wikipedia“- Software für Teamwork an Dokumenten) ▪ (Turnkey) „OpenLDAP“ (schlüsselfertiger Verzeichnissever) ▪ (Turnkey) „Mantis“ (schlüsselfertiges Code Repository) ▪ (Turnkey) „Moodle“ (schlüsselfertiges Lern-Management-System) ▪ (Turnkey) „Collabtive“ (schlüsselfertiges Projektmanagement) ▪ (Turnkey) „OwnCloud“ (schlüsselfertiger Cloud-Service)
ICT-Basistechnologien	<ul style="list-style-type: none"> ▪ Direkte Webtechnologien (HTML, http, XML, Web 2.0, etc.) ▪ Servertechnologien (Betriebssysteme, OS Services, etc.) ▪ Technologien zur Authentifizierung / SSO (Verzeichnissever, Kryptographie, Token, etc.) ▪ Datenbanken und Datenbank Management Systeme: MySQL, PostgreSQL ▪ Content Management Systeme (CMS) ▪ Programmiersprachen (z.B. Java, Java-Script, Python, Perl, C, PHP, etc.) ▪ Metadaten und Auszeichnungssprachen ▪ Semantik Web Technologien ▪ Remote-Desktop-Services ▪ Cloud Computing
Schlüsselfertige Server und Virtualisierung	<ul style="list-style-type: none"> ▪ Installation von „Virtual-Box“ (Virtualisierung) und Erzeugung Virtueller Maschinen (VM) ▪ Installation und Konfiguration von schlüsselfertigen Servern auf den VM ▪ Start-/Stopp-Scripte für VM-Steuerung ▪ Backup-Scripte für automatische VM-Snapshots und -Backups
Homepage und CMS	<ul style="list-style-type: none"> ▪ „Joomla“ CMS mit Erweiterungen: ▪ Installation Basissystem ▪ Installation der Erweiterungen ▪ Konfiguration,

	<ul style="list-style-type: none"> ▪ Design und Layout ▪ Optimierung für Mobilgeräte ▪ Struktur (Menü und andere Gliederungen) ▪ Integration der nicht zum CMS gehörenden Service und Tools ▪ Basis-Inhalte einpflegen ▪ CMS-Systemsicherheit und -Backups ▪ Suchmaschinenoptimierung
Services	<ul style="list-style-type: none"> ▪ klassische CMS-Funktionalitäten („Joomla“) ▪ Social-Media-Community (CMS-Erweiterung)* ▪ Wiki („MediaWiki“ mit „VisualEditor“, etc.) ▪ Metadaten-Management, WebGIS, (Meta)Data-Harvesting ▪ Lern-Management-System (LMS) ▪ Code Repositorien ▪ Projektmanagement ▪ Remote-Desktop-Services ▪ Cloud-Services (OwnCloud) ▪ Screen-Sharing, Videokonferenzen, Chat mit automatischer Übersetzung von Fremdsprachen
Tools (Middleware)	<ul style="list-style-type: none"> ▪ klassische CMS-Funktionalitäten ▪ Automatisierung des Forschungsprozesses und der Workflows (allgemein) ▪ Langzeitdatenspeicherung (Server-Tools, individuell)* ▪ Downloads („GeoNetwork“, CMS) ▪ Suche („GeoNetwork“, CMS) ▪ Datenschema („GeoNetwork“) ▪ automatische Synchronisation von „GeoNetwork“ Metadaten mit relevanten übrigen Metadaten (Programmiersprachen, CMS-Erweiterung)* ▪ Suchmaschinenoptimierung (CMS-Erweiterungen)
Authentifizierung und SSO	<ul style="list-style-type: none"> ▪ „OpenLDAP“ als Verzeichnisserver zur Benutzerverwaltung mit SSO
Report-Layer	<ul style="list-style-type: none"> ▪ Report-Layer und Datenverbünde sind vorerst nicht erforderlich
Hilfe, Feedback und Dokumentation.	<ul style="list-style-type: none"> ▪ Wiki für: Benutzerhilfe, Dokumentation für Administratoren*, Dokumentation für Entwickler* ▪ Benutzerfeedback (CMS-Erweiterungen) ▪ FAQ (CMS-Erweiterungen, „MediaWiki“)
Zusätzliche Erweiterungen, Methoden und Technologien	<ul style="list-style-type: none"> ▪ Alle mit „*“ markierten Elemente <p>Generell sind folgende Gebiete für die spezielle VRE von Interesse:</p> <ul style="list-style-type: none"> ▪ Langzeitdatenspeicherung ▪ Visualisierungstechnologien ▪ Datamining ▪ Web-Labor für: Modellierung, Simulation, Prognose

C Beispiele für VRE Basis-Softwarekomponenten

Tabelle C.1: Open-Source-Software für eine Basis-VRE nach dem „Rapid L-TER Konzept“.

Name	Beschreibung	Turnkey VM
„Linux“ Webserver, MySQL	OS mit Web-Server und DBMS	Ja
„Joomla“ oder anderes CMS	CMS und deren Erweiterungen	Ja
„GeoNetwork Open Source“	Metadatenkatalog, WebGIS, Datenaustausch	bedingt
„MediaWiki“	Wiki (Software für Teamwork bei Dokumenten)	Ja
„Moodle“	E-Learning-Plattform	Ja
„Virtual-Box“ / „VM-Ware“	Virtualisierung und Virtuelle Maschinen	Plattform
„OpenLDAP“	Verzeichnisserver zur zentralen Authentifizierung	Ja
„OwnCloud“	Cloud Server	Ja
„R-Statistics“-Server	Statistik-Software „Shiny“-Server zur Einbindung in Websites (Basisversion von „Shiny“ ist kostenlos)	Nein
„Netlogo“	Modellierung und Simulation, PC-Software erzeugt Java Applets, die in Websites eingebunden werden können	-

Tabelle C.2: Verbreitete Datenbank-Management-Systeme (Open Source)

Klasse	Namen
SQL	MySQL, PostgreSQL (mit GIS-Funktionalität),
noSQL	Riak, Apache Cassandra, CouchDB, MongoDB, Redis

Tabelle C.3: Software für Virtualisierung

Produkt	Hersteller	Beschreibung
„Virtual Box“	Oracle	Einfache und Leistungsstarke VMs
Diverse Produkte	VMware	Professionelle Software für VMs und Virtualisierung
„Hyper-V“	Microsoft	Integrierte Virtualisierung bei MS-Betriebssystemen
Diverse Produkte	Citrix	Professionelle Software rund um Virtualisierung
Diverse Produkte	Virtuozzo	Professionelle Software rund um Virtualisierung

Tabelle C.4: Software für Remote-Desktop und Screen-Sharing

Software / Protokoll	Betriebssystem	Bemerkung
Remote-Desktop-Protokoll (RDP)	MS Windows, höhere Versionen, Varianten (Client: in allen Varianten; Server: Professional, Enterprise, etc. – nicht in Home, Core, Starter)	Es können auch lokale Ressourcen des Clients (Festplatte, Drucker, etc.) auf dem Server eingebunden werden.
Server-Software: Xrdp Client-Software: rdesktop	Linux	Leistungsstarke Software für Linux-Systeme.
Apple Remote Desktop (ARD)	Apple	Zugriff über VNC auch auf Windows, und Linux.
Skype mit integriertem Screen-Sharing	plattformübergreifend	Beaufsichtigter Zugriff
HTML5 Funktionen	plattformübergreifend	Screen-Sharing, gute Basis für eigene Web-Projekte
Virtual Network Computing (VNC): UltraVNC, Real VNC	plattformübergreifend	Sitzungsabhängig
Teamviewer	plattformübergreifend	Sitzungsabhängig, beaufsichtigter und unbeaufsichtigter Zugriff. Sehr umfangreich und einfach zu benutzen. Kommerzielle Nutzung ist kostenpflichtig.

Tabelle C.5: Verbreitete Content-Management-Systeme nach Breitenstrom et al. (2013)

Name	Beschreibung
Joomla	zeichnet sich durch seine einfache Installation und Bedienung, die große Anzahl an Erweiterungen und eine große Community mit regelmäßigen Updates aus. Es kann sehr gut für Standard-Websites, Communities und Informationsportale eingesetzt werden.
Typo 3	ist ein sehr umfangreiches, professionelles CMS für große Internetplattformen. Die Installation und Konfiguration des CMS erfordert jedoch besonderes Fachwissen. Individuell skalierbare Benutzeroberflächen ermöglichen eine einfache Bedienung für den Anwender.
Plone	eignet sich besonders für dokumentenlastige Websites. Umfangreiche Möglichkeiten für Benutzer- und Rechteverwaltung unterstützen einen schnellen Aufbau von Kollaborations- und Kommunikationsplattformen. Es bietet als professionelles CMS mit Barrierefreiheit eine sehr hohe Stabilität und Skalierbarkeit.
Drupal	ist ein zeitgemäßes, stabiles CMS, das sich auf Communities mit viel von Benutzer erzeugten Inhalten spezialisiert hat (wie z.B. Foren). Es blickt auf eine langjährige Entwicklungszeit zurück und hat eine große Community.
Wordpress	ist von Haus aus ein Weblog-Publishing-System, das sich besonders für häufig zu aktualisierende Inhalte eignet. Zunehmend findet es durch die zahlreichen Erweiterungen Verwendung als normales CMS für einfache Websites.

Tabelle C.6: Verbreitete Verzeichnisdienste vgl. Klünter und Laser (2008)

Name	Hersteller
389 Directory Server	früher „Fedora Directory Server“ und „Netscape Directory Servers“, freie Software
Active Directory (AD)	von Microsoft in Microsoft Windows Server und Samba
Apache Directory Server	Apache Software Foundation
Atos DirX	früher „Siemens DirX“, von Atos
Critical Path Directory Server	früher „Novell eDirectory“ und „Novell Directory Services“ – „NDS“, von Novell
NetIQ eDirectory	
Network Information Service (NIS)	von Sun Microsystems, nicht LDAP-basiert
Open Directory	von Apple in macOS Server
OpenLDAP	Open Source für diverse Betriebssysteme
Oracle Directory Server	früher „Sun Directory Server“ und „Sun ONE LDAP Server“, von Oracle
SAP Solution Manager	SAP
Tivoli Directory Server	IBM
ViewDS Directory Server	von „eNitiatives“

Tabelle C.7: Verbreitete SSO-Produkte und Standards vgl. Perske (2012)

Name	Beschreibung
Liberty Alliance Project	Anmeldedienst (dezentralisierte Lösung)
Shibboleth	Anmeldedienst (dezentralisierte Lösung)
OpenID	Anmeldedienst, dezentrales Protokoll in der Informationstechnik
Security Assertion Markup Language	Single Sign-on Protocol für Webdienste
Kerberos	verteilter Authentifizierungsdienst (Netzwerkprotokoll)
IDpendant Single Sign -on	unterstützt sehr viele Authentisierungsarten, ermöglicht Mehrbenutzer und Sitzungsweiterleitung
Central Authentication Service (CAS)	eine auf Servlets basierende SSO-Lösung für Webapplikationen
Lightweight Third-Party Authentication (LTPA)	wird in den Produkten IBM Websphere und Lotus Domino verwendet

Tabelle C.8: „MediaWiki“ Erweiterungen für VREs

Name	Beschreibung
Visual Editor	WYSIWYG-Editor mit zahlreichen Funktionen, wie z.B.: Formeleditor
Google Charts	Einbindung für automatisch generierte statistische Grafiken (Charts) durch den Google Service (Google 2017).
R	„R-Statistic“ Scripte mit interaktiven Parametern werden im Wiki ausgeführt (Klinke 2014)
Shiny	„Shiny“ Oberfläche für „R-Statistik“ in „MediaWiki“ (Klinke 2016)

D „Joomla“ CMS: ausgewählten Erweiterungen für VREs

In dieser Arbeit wird aufgrund der Flexibilität, der einfachen Installation und Bedienung sowie der jahrelangen eigenen Erfahrungen, speziell auf das „Joomla“ CMS und dessen Komponenten eingegangen. Für den Aufbau einer VRE können je nach Anforderungen auch andere CMS verwendet werden.

Das „Joomla“ CMS zeichnet sich besonders durch seine Vielzahl an Erweiterungen mit Komponenten (**C**), Modulen (**M**) und Plug-Ins (**P**) aus, die für sehr viele Einsatzzwecke auf der Homepage von „Joomla“ zur Verfügung stehen (joomla.org 2017). Durch die Vielzahl ist es trotz der Suchfunktionen und Kategorien nicht immer einfach, geeignete Erweiterung auf Anhieb zu identifizieren. Neben dem technischen Reifegrad der jeweiligen Erweiterung ist auch entscheidend, inwieweit sie mit der CMS-Version und den anderen eingesetzten Erweiterungen kompatibel ist und wie die Software weiterentwickelt, bzw. gepflegt wird. Dabei spielen besonders sicherheitsrelevante Aspekte eine Rolle (Breitenstrom et al. 2013). Die Flexibilität der Erweiterungen für unterschiedliche Aufgabenstellungen ist nicht immer gleich ersichtlich. Manche Erweiterungen lassen sich für viele Einsatzgebiete verwenden, wie z.B. Katalog-Erweiterungen.

Nachfolgend sind einige besonders erwähnenswerte Erweiterungen für den Einsatz in einer Virtuellen Forschungsumgebung mit einer kurzen Beschreibung zusammengestellt. Viele davon wurden in jahrelanger Erfahrung des Autors praktisch getestet. Bei der Auswahl wurden die am geeignetsten erscheinenden Komponenten für den jeweiligen Einsatzzweck ausgewählt, auch wenn sie nicht frei verfügbar (FLOSS) sind, sondern als kostenpflichtige Lizenz erworben werden müssen. Die Kosten für die Lizenzen der Erweiterungen sind jedoch meist sehr gering:

Externe Authentifizierung:

- **„Single Sign-on“** ist ein Framework, um externe Authentifizierungen durchzuführen. Folgende Authentifizierungsmethoden werden unterstützt: CAS, openid, AOL OpenAuth, Google AuthSub, Microsoft Windows Live Contacts API, Yahoo BBAuth. (**C, M, P**, Joomla 2, frei)
- **„JMapMyLDAP Tools“**: Bindet LDAP Authentifizierung ein für: OpenLDAP, Microsoft Active Directory, Novell eDirectory. (**C, M, P**, Joomla 2.5/3, frei)
- **„Shimla SSO“**: Shibboleth SAML Standard für SSO auf Joomla-Websites (9starinc 2016)
- **„MediaWiki Login“**: Dieses Plug-In ermöglicht den gleichzeitigen Login auf Joomla- und „MediaWiki“-Websites mit Single-Sign-on. (**P**, Joomla 2.5/3, frei)
- **„OpenSSO login“** nutzt auf der Basis von OpenSSO Token Cookies für den Login. (**P**, Joomla 3, frei)
- **„Shmanic LDAP“**: LDAP Anbindung für „Joomla“. (**C, P**, Joomla 3, frei)

- Für „Facebook“ und andere Social-Networks finden sich zahlreiche Erweiterungen.

Lernmanagementsysteme (LMS):

- **„Online Virtual Classroom“:** Cloud-basiertes Lern-Management-System (LMS) mit Live-Classrooms, Meetings, Webinars, Video-/Audio-Conferences, Whiteboards. (C, Joomla 2.5/3, frei)
- **„Joomla“ („Moodle“ für „Joomla“):** Dieses Plug-In ermöglicht den gleichzeitigen Login auf Joomla- und Moodle-Websites (Lernplattform) mit Single-Sign-on. Gleichzeitig kann es die Inhalte aus Moodle darstellen und z.B. online-Kurse kommerziell mit der Komponente „VirtueMart“ anbieten. Communitys von Drittanbietern, wie „Jomsocial“ und „Community Builder“ werden unterstützt. (C, M, P, Joomla 2.5/3, frei)

Entwicklung eigener Komponenten:

- **„Just Another Component Creator (JACC)“:** Mit JACC können installierbare Pakete für Komponenten, Module, Plug-Ins und Templates einfach entwickelt werden. (C, Joomla 2.5/3, frei)
- **„Component Architect“:** Webbasiertes Entwicklertool für Komponenten, Module, Plug-Ins und Templates. (C, Joomla 3, frei)
- **Anything Studio Pro:** Datenbankanwendungen für Joomla mit SQL Abfrage ohne Programmierung. (C, Joomla 3, kostenpflichtige Lizenz)

Datenvisualisierung und Statistik:

- **„Art Data“:** Diese Sammlung von Plug-Ins ist auf die Präsentation von Daten spezialisiert. Dazu gehören die Darstellung von Tabellen, Bildern, Diagrammen, Statistiken. Es werden unterschiedliche Datenbankformate unterstützt. (C, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„ARI Data Tables“:** Konvertiert HTML-Tabellen und bietet Filter, Sortierung und Paginierung auch für SQL-Datenbanken. (M, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„Fancy TagCloud“:** Erzeugt aus den Joomla-Tags eine optisch ansprechende Darstellung von Tags (Keywords). (M, P, Joomla 2.5/3, frei)
- **„Joes Word Cloud“:** Sucht automatisch die Keywords aus Joomla Artikeln und bereitet sie optisch als Modul auf. Arbeitet auch mit zahlreichen anderen Komponenten zusammen. (M, Joomla 2.5/3, frei)

Community / Social Networks:

- **„JomSocial“:** Bietet sehr komfortable und professionelle Community-Funktionalitäten im „Facebook Style“ mit Posts, Multimedia, Events, Gruppen, Benutzerprofilen, Privatsphäre-Einstellungen. Dazu gibt es zahlreiche Erweiterungen von Drittanbietern. (C, M, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„Community Builder“:** Diese Komponente ist ähnlich umfangreich, wie JomSocial, aber frei erhältlich. (C, M, P, Joomla 2.5/3, frei)
- **„JChatSocial“:** Erweiterung zu „JomSocial“ für Videokonferenzen und Chats mit Simultanübersetzung. (C, P, M Joomla 3, kostenpflichtige Lizenz)
- Weiterhin existieren zahlreiche Komponenten, um Inhalte von Joomla in sozialen Netzwerken zu teilen oder mit ihnen zu synchronisieren.

Sonstige:

- **„JSON Export“:** exportiert jede Tabelle in das besonders für mobile Endgeräte geeignete JSON-API-Format. Damit lassen mit wenig Aufwand mobile Apps an Joomla-Datenbanken koppeln. (C, Joomla 3, kostenpflichtige Lizenz)
- **„NoNumber ReReplacer“:** Komfortables Suchen und Ersetzen von Texten und Strings auf der Website, z.B. bei der Anpassung von fremdsprachigen Komponenten sehr hilfreich. (C, P, Joomla 2.5/3, frei)
- **„DB Replacer“:** Komfortables Suchen und Ersetzen von Texten und Strings in Datenbanken, z.B. bei der Übernahme von Datenstrukturen sehr hilfreich. (C, Joomla 2.5/3, frei)
- **„ImproveMyCity“** ist eine Komponente, um Bürger durch Meldungen mit Fotos, Karteneintrag und Abstimmungen in umweltbezogene städtische Belange einzubeziehen. Schwerpunkte sind bauliche Verbesserungen / Reparaturen, illegale Müllentsorgung, Lärm. (C, Joomla 2.5, frei)
- **“Add Custom CSS”:** Individuelle CSS-Layout-Einstellungen, die zentral gepflegt werden. Besonders geeignet, um das CSS-Layout verschiedener Komponenten nachträglich permanent zu ändern. (M, P, Joomla 2.5/3, frei)
- **„Conference and Abstract Manager“:** Plattform für das Management von Konferenzen mit Registrierungsverwaltung, Tickets, Abstract-Submission, Committee Reviewers. (C, Joomla 2.5, kostenpflichtige Lizenz)
- **„Calc Builder“:** Erzeugt Formulare für dynamische Berechnungen auf der Website mit Feedback und Statistik. (C, M, Joomla 2.5/3, kostenpflichtige Lizenz)
- **“OS PDF Indexer”** indiziert automatisch vorhandene PDF-Dokumente und erstellt eine Joomla Datenbank für deren Suchbegriffe mit Link zu den relevanten Dokumenten. (C, P, Joomla 2/3, kostenpflichtige Lizenz) [Joomla.org]

- **„jQuery Easy“** hilft bei der Verwendung der jQuery Libraries für dynamische Websites. (P, Joomla 2.5/3, frei)
- **„OSE Anti-Hacker“** und weitere Produkte aus dem Hause OSE sichern Joomla umfassend auf dem Stand der Technik gegen Hacker-Angriffe ab. (C, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„Community Polls“** erzeugt Umfragen mit vielseitigen Funktionalitäten und Auswertungsstatistiken. (C, M, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„iSEO“** optimiert Websites für die Suchmaschinen, vereinfacht das Management von Metatags und erzeugt eine Statistik für Keyword-Positionen in Google. (C, P, Joomla 2.5/3, frei/ kostenpflichtige Lizenz)
- **„XCloner“** ist eine professionelle Komponente für Backup und Restore von kompletten Joomla-Websites inklusive Datenbanken mit zahlreichen Einstellungsmöglichkeiten. (C, Joomla 2.5/3, frei)
- **„JoomGallery“** hebt sich durch seine Funktionsvielfalt und Benutzerfreundlichkeit von vielen anderen Galerie-Erweiterungen ab. (C, M, P, Joomla 3, frei)
- **„JoomSEF“** erzeugt und verwaltet suchmaschinenfreundliche (SEF) und Anwenderfreundliche Links mit Metatag-Funktionalitäten. (C, P, Joomla 2.5/3, frei/ kostenpflichtige Lizenz)
- **„K2“** eignet sich als Katalogsystem und erweitert Joomla-Artikel um optionale Felder. Zahlreiche Erweiterungen von Drittanbietern arbeiten mit K2 zusammen. (C, M, P, Joomla 2.5/3, frei)
- **„Sobi pro“** ist ein sehr anpassungsfähiges Katalogsystem mit vielen Erweiterungen von Drittanbietern (C, M, P, Joomla 2.5/3, frei)
- **„Map Locations“** erstellt benutzerdefinierte Karten auf der Basis von Google Maps mit vielen unterschiedlichen Layouts und Funktionen. (C, M, Joomla 3, kostenpflichtige Lizenz)
- **„WISroGIS mapping“** bietet die Möglichkeit, GIS-Karten mit Layer-Technologie in Joomla zu veröffentlichen. Folgende Formate und Schnittstellen werden unterstützt: KML, GPX, Google Map, Yahoo Map, Virtual Earth, Open Street Map, WMS Server. (C, M, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„AcyMailing“** ist eine umfangreiche Erweiterung zum Management von Newslettern mit verschiedenen Listen. User-Management, Templates, Vorschau, Integration von Joomla-Inhalten, Bounce-Management und einstellbare Parameter beim Newsletter Versand zählen zu den Optionen. (C, M, P, Joomla 2.5/3, kostenpflichtige Lizenz)
- **„Paper Manager“** organisiert wissenschaftliche Publikationen. (C, P, Joomla 3, frei)

-
- **„JS Jobs“**: Umfangreiche Jobbörse auch für Praktika und Abschlussarbeiten nutzbar. . (C, P, M, Joomla 3, kostenpflichtige Lizenz)
 - **„Projectfork“**: Einfaches Projektmanagement mit Meilensteinen . (C, P, M, Joomla 3, frei)
 - **„Geek Elasticsearch“** ist eine skalierbare Volltext-Suchmaschine für Joomla auf der Basis von Elasticsearch, skalierbar. (C, M, P, Joomla 2.5/3, kostenpflichtige Lizenz)
 - **„Web Push Notification“ (Chrome and Firefox)**: Push-Nachrichten über den Browser
 - **„Rest-API“** erweitert das CMS für Funktionen mit verteilten Systemen in Form einer Programmierschnittstelle auf der Basis von REST. (P, Joomla 3, kostenpflichtige Lizenz)
 - **„cAPI“** erweitert das CMS für Funktionen mit verteilten Systemen in Form einer Programmierschnittstelle auf der Basis von REST. (C, P, Joomla 3, frei)
 - Für das **Responsive Layout** (Anpassung an Displaygröße und Auflösung des Endgerätes, z.B. Smartphone / Tablet) eignen sich z.B. die freien und kommerziellen Templates von „Joomlart“ (joomlart 2017) und „Joomlashine“ (Joomlashine 2017) mit deren Erweiterungen.

E Beispiele für Befehle zur Steuerung und Backup von VMs

Die einzelnen Befehle können in einer Datei (z.B. *.bat oder *.cmd für Windows oder *.sh für Linux) zu einem Script zusammengefasst und durch einen Zeitplan zur Ausführung automatisch gestartet werden.

Start einer VM mit GUI-Interface (Windows, Oracle Virtual Box):

```
"c:\Program Files\Oracle\VirtualBox\VBoxManage.exe" startvm  
  "TURNKEY JOOMLA3 VRE" --type gui
```

Start einer VM ohne GUI-Interface (Windows, Oracle Virtual Box):

```
"c:\Program Files\Oracle\VirtualBox\VBoxManage.exe" startvm  
  "TURNKEY JOOMLA3 VRE" --type headless
```

Stop einer VM (Windows, Oracle Virtual Box):

```
"c:\Program Files\Oracle\VirtualBox\VBoxManage.exe"  
  controlvm "TURNKEY JOOMLA3 VRE" acpipowerbutton
```

Erstellen eines VM-Snapshots (Windows, Oracle Virtual Box):

```
"c:\Program Files\Oracle\VirtualBox\VBoxManage.exe"  
  snapshot "TURNKEY JOOMLA3 VRE" take WeeklyBackup
```

Netzwerkpfad verbinden und Daten auf Netzwerkspeicher kopieren (Windows):

1. net use \\nas.Domain.de\Backup_VRE_VMs /persistent:yes
 /user:admin <PASSWORD>
2. xcopy /s/h "C:\Users\Administrator\VirtualBox VMs"
 "\\ nas.Domain.de\Backup_VRE_VMs"

F Schlüsselfertige „Linux“-Server für L-TER VREs

Die folgende Übersicht zeigt über 100 „schlüsselfertige“ Open-Source-Linux-Server für bewährte Software zu verschiedensten Einsatzzwecken. Die Liste ist nicht speziell für VREs zusammengestellt, aber dennoch sind die meisten Server gut für die vielfältigen VRE-Services und -Tools in L-TER VREs einsetzbar. Sie können als CD/DVD-Image oder als VM von www.turnkeylinux.org kostenlos heruntergeladen werden. Für das „Rapid L-TER VRE“ Konzept wurden daraus einige Beispiele verwendet, wie in der Arbeit beschrieben. Alternativ können z.B. auch „Docker“-Images genutzt werden, die bei Abschluss dieser Arbeit populär wurden (Docker.com 2017).



LAMP Stack

Web Stack (MySQL)



WordPress

Blog Publishing Plattform



Observium

Network Management and Monitoring



File Server

Simple Network Attached Storage



OpenVPN

Open Source VPN solution



Domain Controller

free Active Directory server



ownCloud

Share files, music, calendar



Redmine

Integrated SCM & Project Management



OpenLDAP

Open Source Directory Services



OTRS

Ticket Request System



TurnKey LXC Linux Containers

1 host, multiple TurnKey apps



Odoo

From ERP to CRM, eCom-merce to CMS

**GitLab**

Self Hosted Git Management

**Joomla 3**

Mobile-ready user-friendly content management

**MediaServer**

Simple Network Attached Media Storage

**Nginx PHP FastCGI Server Configuration**

with Adminer

**PostgreSQL**

Advanced Object-relational Database System

**Torrent Server**

File download and sharing server

**MySQL**

Relational Database Management System

**Node.js**

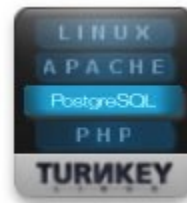
Asynchronous Javascript Framework

**Moodle**

Course Management System

**Drupal 7**

Content Management Framework

**LAPP**

Web Stack (PostgreSQL)

**Jenkins**

Continuous integration

**MongoDB**

NoSQL database

**Ruby on Rails**

Web Application Framework

**OrangeHRM**

Human Resource Management (HRM) software

**Ansible**

Radically simple IT automation platform



SugarCRM

Business & Social CRM software



Magento

Flexible Shopping Store eCommerce Platform



DokuWiki

Documentation Wiki Platform



Canvas

Learning Management System



Bugzilla

Bug Tracking System



SiT!

Support Incident Tracker



vTiger CRM

Customer Relationship Management



ProcessMaker

Workflow & BPM software



Mibew

Live Support Chat



SuiteCRM

CRM for the world



Laravel

PHP Framework for Web Artisans



web2py

Python framework



Revision Control

All-in-one code repository



CodeIgniter

Web application framework



TYPO3 CMS

Enterprise CMS



Tracks

Getting Things Done (GTD) Application



**ASP .NET on Apache with
Mod Mono**
Free .NET hosting



TomatoCart
Shopping cart



Tomcat on Apache
Java Servlet and JSP
Platform



phpBB
Community Forum Solution



b2evolution
Content management system



SimpleMachines
Forum system



CakePHP
Rapid development PHP
framework



SimpleInvoices
Invoicing system



TurnKey Core
Debian GNU/Linux with
Batteries Included



Zurmo
Gamified, Social, Mobile
CRM system



XOOPS
Content Management and
Web Application Platform



Yii Framework
PHP framework



Django
High-level Python Web
Framework



Mambo
Content Management system



Zen Cart
online store management
system



Mahara
Electronic portfolio and
social networking



Ghost

A publishing platform for professional bloggers



Lighttpd PHP FastCGI Configuration

with Adminer



PrestaShop

Easy to use online shop



iceScrum

Agile collaborative development



GNU social

Open, distributed microblogging



Foodsoft

Web management for non-profit food cooperatives



Standalone Tomcat

Java Servlet and JSP Platform



Omeka

Serious web publishing for cultural collections



MoinMoin

Wiki Engine



Trac

Integrated SCM & Project Management



EspoCRM

lightweight customer relationship manager



osCommerce

Online shop



Roundup

Issue Tracking System



LimeSurvey

Survey application



Mantis

Bug Tracking System



Plone

Open Source Content Management

**PunBB**

Forum software

**TKLDev**TurnKey Development
Toolchain and Build System**Sahana Eden**

Humanitarian platform

**Gallery**

Photo Album Organizer

**SilverStripe**

CMS and framework

**eZ platform**Enterprise Content
Management System**Kliqqi**

Social publishing CMS

**Etherpad Lite**Real-time document
collaboration**Piwik**Self Hosted Real-Time Web
Analytics**Elgg**

Social networking engine

**e107**

Content Management system

**phpList**Email campaign
management**CouchDB**

JSON based Web database

**PhreeBooks**Enterprise Resource
Planning**TWiki**

Enterprise Wiki Platform

**Concrete5**Next generation Content
Management System



Vanilla forum

Community forums evolved



ProjectPier

Easy Online Collaboration



Collabtive

Open Source Project Management



Ushahidi

Crowdsourcing Crisis Information



Symfony

PHP Web Framework



MediaWiki

Wikipedia's Wiki Engine



Mattermost

Self-hosted Slack compatible team messaging



Mumble Server

the open source VoIP solution



Nextcloud

Share files, music, calendar



Mayan-EDMS

An electronic vault for your documents



Drupal 8

Content Management Framework

Danksagung

An dieser Stelle möchte ich meinen besonderen Dank nachstehenden Personen entgegenbringen, ohne deren Mithilfe die Anfertigung dieser Promotionsschrift niemals zustande gekommen wäre:

Zuerst danke ich meinem Doktorvater, Prof. Dr. Albrecht Gnauck, für die fundierte Ausbildung im Studium, die langjährige Zusammenarbeit am Lehrstuhl „Ökosysteme und Umweltinformatik“ und für die Begutachtung im Ruhestand. Er motivierte und unterstützte mich stets bei allen Ideen und Vorhaben seit den Anfängen des Studiums. Seine Expertise aus der Kombination von Umweltinformatik, Ökosystemtheorie und Mathematik, sowie das Engagement in wissenschaftlichen Netzwerken, waren für diese Arbeit von besonderer Bedeutung.

Ebenso danke ich meinem Chef, apl. Prof. Dr. Frank Molkenthin, für die Übernahme des Erstgutachtens, nach dem Prof. Dr. Gnauck in den Ruhestand gegangen war. Er half mir, den roten Faden wieder aufzugreifen und gab mir sehr hilfreiche und pragmatische Unterstützung bei der Fertigstellung der Arbeit in einer für mich sehr schweren Zeit, in der auch mein Vater verstorben ist.

Prof. Dr. Gerhard Wiegler danke ich besonders für die mehrfache Durchsicht der Arbeit, die Übernahme der Begutachtung im Ruhestand und für die langjährige Zusammenarbeit am Lehrstuhl „Allgemeine Ökologie“. Sein fast tägliches Mentoring war mir über die Jahre eine sehr große Hilfe und Stütze.

Bei Dr. Werner Gerwin bedanke ich mich für die differenzierten Anmerkungen sowie die zweckdienlichen Diskussionen in Bezug auf eine VRE für das Forschungsprojekt „Hühnerwasser“.

Dr. Andreas Degkwitz als ehem. Leiter des „Informations- und Kommunikationszentrums“ der BTU (IKMZ) gilt der Dank für die Inspiration und Motivation zu diesem Thema. Durch seine Anfrage habe ich mich erstmals intensiver mit VREs beschäftigt.

Herrn Rolf de Vries danke ich für zahlreiche Hintergrundinformationen im Umweltmanagement und die lehrreichen Gespräche über die Praxis.

Für die interessanten Einblicke zum Datenmanagement des Bundes und zu Umweltinformationssystemen bei zahlreichen Konferenzen und Workshops danke ich Frau Dr. Gerlinde Knetsch aus dem Umweltbundesamt.

Dr. Werner Pillmann und Dr. Kristina Voigt aus dem Komitee der „EnviroInfo“ Konferenz danke ich für das Engagement im Forschungsnetzwerk und für die weiterführenden Gespräche im Bereich der Umweltinformatik.

Meinen Arbeitskollegen, gilt der Dank für die Unterstützung bei fachbezogenen Themen und geduldigen Tests der Prototypen.

Ein allgemeiner Dank geht an das Rechenzentrum der BTU für die Unterstützung mit den Basisdiensten und die technische Beratung.

Von ganzem Herzen möchte ich mich bei meiner Mutter und der Familie für ihre lebenslange Unterstützung, Geduld, Motivation und Zuversicht bedanken.

Bei meinen Freunden bedanke ich mich herzlich für die moralische Unterstützung und die Unternehmungen, wenn der Kopf sonst nicht mehr frei wurde.