

wibaklidama
wissensbasiertes **Klima-Datenmanagement**

Projektbericht B3

Analyse und Darstellung des State-of-the-Art von Werkzeugen für E-Infrastrukturen in der Wissenschaft

Projektkürzel	wibaklidama	
Typ und Nr.	B3 State-of-the-Art-Bericht	
Arbeitspaket(e)	AP3	
Datum	22.4.2010	
Dateiname	wibaklidama-B3- V02.pdf	Version 2
Status	Öffentlich	
Autoren	Georg Mannsperger	
Abstract	Analyse, Beschreibung und Empfehlungen zu Tools zur Unterstützung des Datenmanagements in Orientierung am Data Curation Lifecycle.	
Stichwörter	Datenmanagement, Tools	



LAND
BRANDENBURG

Ministerium für Wissenschaft,
Forschung und Kultur



Inhaltsverzeichnis

Einleitung	3
Kap. 1: Information Retrieval.....	7
Relevancy Retrieval	7
IDOL Retina.....	10
Ultraseek.....	13
Johns Hopkins Institutions	13
FAST ESP	17
Coveo Enterprise Search	19
Kap. 2: Kollaborationssysteme	22
BSCW.....	22
<i>kleine Firma oder Projektgruppe</i>	26
<i>mittlere Firma / mehrere Projekte / externe Teilnehmer</i>	26
<i>große Firma</i>	26
Opengroupware.org.....	27
MS Sharepoint.....	30
Sonexis Conference Manager.....	32
Kap. 3: Workflow-Management-Systeme.....	35
Taverna.....	35
Kepler	37
Pegasus.....	39
Kap. 4: Datenvisualisierung	40
Matlab.....	40
DataGraph	44
Cartographica	46
GSI 3D.....	49
InstantAtlas.....	52
Paraview	54
Rockworks 15.....	57
Visual Thesaurus	59
Kap. 5: Speichersysteme	63
LOCKSS (Lots of Copies Keep Stuff Safe).....	63
Kopal	66
kopal Library for Retrieval and Ingest	68
kopal-Solution	70
kopal-Tools.....	70
DIAS-Core	70
DIAS.....	73
Tivoli Storage Manager.....	75
Kap. 6: Datenmanagement-Systeme.....	78
FuD (Forschungsnetzwerk und Datenbanksystem).....	78
eSciDoc	80



Einleitung

Der Rohstoff der Wissenschaft ist ein heikles Gut: Wissen. Dieses zu erwerben, zu strukturieren, vorzuhalten und auszutauschen: Das ist das Tagesgeschäft des Wissenschaftlers. Gleichzeitig ist es die Realität der Forschung, dass in einer globalen Welt die Zusammenarbeit in Projektgruppen immer häufiger orts- und zeitübergreifend stattfindet. Die Unterstützung der Forschungsarbeit durch elektronische Hilfsmittel ist daher längst Alltag. Es gibt jedoch nur wenige Konzepte, die einen Gesamtüberblick über die Vielzahl der dafür erhältlichen Softwareprodukte systematisch in den Zusammenhang des Einsatzzwecks Forschungsarbeit stellt. Die vorliegende Zusammenstellung soll interessierten Anwendern eine Entscheidungshilfe für die Auswahl geeigneter Systembausteine zur Unterstützung ihrer Forschungsarbeit geben. Zu diesem Zweck wurden sechs Kategorien von Anwendungen gebildet, die innerhalb einer forschungsnahen IT-Architektur – im Folgenden „E-Infrastruktur“ – jeweils eine bestimmte Funktion erfüllen:

- Information-Retrieval-Systeme
- Kollaborationssysteme
- Workflow-Management-Systeme
- Datenvisualisierungssysteme
- Speichersysteme
- Datenmanagementsysteme

Mit Hilfe dieser Systemtypen lassen sich die typischen wissenschaftlichen Vorgänge elektronisch unterstützen. In den folgenden Kapiteln werden jeweils einige ausgewählte Systeme aus diesen Klassen bezüglich ihrer Eigenschaften und Funktionalitäten vorgestellt und hinsichtlich ihres Einsatzes in E-Infrastrukturen bewertet. Ein wesentliches Kriterium zur Einordnung der Systeme bildet dabei das „Curation Lifecycle Model“ des schottischen Digital Curation Center – ein verbreiteter Ansatz zur Abbildung der wesentlichen Workflows digitaler Datenhaltung (vgl. <http://dcc.ac.uk/resources/curation-lifecycle-model>). Jedes System wurde danach bewertet, für welche Prozessschritte des Lifecycles es besonders hilfreich ist. Diese Einordnung ist einer bei jedem System abgebildeten Tabelle zu entnehmen, in der die Prozessschritte abgebildet sind. In Rot sind diejenigen Prozessschritte hervorgehoben, für die das jeweilige Instrument besonders nützlich ist.

Einleitend sei ein Modell vorgestellt, das alle o. g. Systemtypen in einem integrierten Einsatzszenario demonstriert. Es handelt sich dabei um eine web-basierte Plattform, welche die verschiedenen an einer Forschungsinstitution üblichen Prozesse zur Schaffung von Wissen elektronisch abbildet und transparent macht („integriertes Wissensverarbeitungs-Framework“). Die Plattform hat drei verschiedene Bereiche, die sich an den verschiedenen prototypischen Phasen des Forschungsprozesses orientieren. Am Anfang steht die Ideenfindung. Ganz wie bei einem System des betrieblichen Vorschlagswesens können hier Ideen für mögliche Forschungsprojekte recherchiert und zusammengetragen werden. Zu diesem Zweck müssen die Wissenschaftler wissen, was in ihrem Forschungsgebiet und in verwandten Bereichen schon an Ergebnissen vorliegt, um diese als Basis für neue Forschungsfelder nehmen zu können. Dafür kommen in dem Wissensverarbeitungs-Framework Information-Retrieval-Systeme zum Einsatz (siehe Kap. 1), die mit Hilfe linguistisch verfeinerter Suchtechnologien die unterschiedlichsten Datenquellen, Datenformate und wissenschaftlichen Datenbanken abfragen. Hat ein Forscher ein neu zu erschließendes Feld entdeckt, publiziert er es als Idee in einer einfach zu bedienenden, wiki-ähnlichen Benutzeroberfläche. Nun kann sich eine Forschungsinstitution entschließen, eine



oder mehrere der so publizierten Ideen innerhalb eines Forschungsprojektes weiterzuverfolgen. Dann eröffnet sie innerhalb des Frameworks ein neues Projekt. Für die Projektumsetzung – Phase 2 des Modells – stehen der Forschungsinstitution nun innerhalb eines Kollaborationssystems (siehe Kap. 2) zahlreiche Funktionalitäten für wissenschaftlichen Austausch und Kooperation zur Verfügung. Außerdem werden in dieser Phase Workflow-Management-Systeme eingesetzt, um Experimente elektronisch zu steuern (siehe Kap. 3) sowie Visualisierungssysteme (siehe Kap. 4), um die Projektergebnisse darzustellen. Ist ein Projekt abgeschlossen, so geht es in die letzte Phase über – die Publikation und Archivierung der Ergebnisse. War die zweite Phase noch einer geschlossenen Benutzergruppe – dem Projektteam – vorbehalten, so findet die letzte, wie die erste, wieder öffentlich sichtbar statt, so dass die wissenschaftliche Community im Sinne des Open-Content-Gedankens von den Ergebnissen der Forschungsprojekte profitieren kann. Das Wissensverarbeitungs-Framework einer anderen Forschungseinrichtung würde in der Phase der Ideenfindung mit seinen eigenen Information-Retrieval-Systemen wieder auf die gespeicherten Ergebnisse der erstgenannten Institution zurückgreifen können. In der E-Infrastruktur des Frameworks kommen in der Phase der Publikation und Archivierung Speicher- und Datenmanagementsysteme zum Einsatz (siehe Kap. 5 und 6).

Neben der Klassifizierung nach Systemtypen enthält die folgende Tabelle, die das dem Wissensverarbeitungs-Framework zugrunde liegende Schichtenmodell illustriert, eine Überklassifizierung der Systemtypen hinsichtlich ihrer Einsatzweise: Die explizit benutzerorientierten Tools werden vom Endanwender zur unmittelbaren Erfüllung seines Forschungsinteresses eingesetzt. Die implizit benutzerorientierten Tools dienen als sekundäres Werkzeug zur elektronischen Unterstützung von Prozessen, die ohne computertechnische Unterstützung nicht oder nur sehr aufwändig durchführbar wären. Die Schritte, die für den Einsatz der rein datenorientierten Tools notwendig sind, bekommt der Endanwender schon gar nicht mehr mit, da sie von der IT-Abteilung der Forschungsinstitution durchgeführt werden. Die Überklassifizierung illustriert also die während der einzelnen Phasen des Schichtenmodells abnehmende Involvement des Endnutzers in die technischen Abläufe der E-Infrastruktur.



Schichtenmodell einer E-Infrastruktur für ein integriertes Wissensverarbeitungs-Framework

	Idee 1	Idee 2	Idee 3	alle Ideen	Phase 1 Ideenfindung Information Retrieval
	Aktuelles Projekt A	Aktuelles Projekt D	Aktuelles Projekt G	alle aktuellen Projekte	Phase 2 Projektumsetzung Visualisierungssysteme Kollaborationssysteme Workflow-Management-Systeme
	Aktuelles Projekt B	Aktuelles Projekt E	Aktuelles Projekt H		
	Aktuelles Projekt C	Aktuelles Projekt F	Aktuelles Projekt I		
	Abgeschlossenes Projekt B	Abgeschlossenes Projekt D	Abgeschlossenes Projekt I	alle abgeschlossenen Projekte	Phase 3 Publikation und Archivierung Datenmanagement-Systeme Speichersysteme

- Farbcode der Überklassifizierung:

Explizit benutzerorientierte Tools

Implizit benutzerorientierte Tools

Rein datenorientierte Tools

Für dieses Modell kommt die Metapher des Schichtenmodells zum Einsatz: Lose Ideen (visualisiert durch die Grashalme) verdichten sich im Forschungsprozess zu Wissensbausteinen (visualisiert durch die Steine unten in der Abbildung). Im Sinne der Prozesstransparenz sind in der Phase der Projektumsetzung noch die Vorgänge der Ideenfindung gespeichert. In Phase 3 werden schließlich alle drei Vorgänge publiziert. Auf diese Art und Weise werden nicht nur die Forschungsergebnisse, sondern auch der Weg dorthin vollständig dokumentiert, was zusätzliche Erkenntnisse für Wissenschaftler und wertvolle Lerneffekte für Studierende bringen kann.



Ein Anwendungsbeispiel für das Wissensverarbeitungs-Framework soll das Prinzip verdeutlichen:

Phase 1: Studenten, Dozenten und Externe schlagen fachübergreifende und fachbezogene Ideen vor.

Idee 1: Konservierung unserer aktuellen Lebenssituation in Wort, Bild und Ton für künftige Generationen.

Idee 2: Wortschatzvergleich zwischen verschiedenen Sprachen.

Idee 3: Einfluss des Wetters auf Arbeitsdauer, Stimmung und Art des Abendessens.

Die Ideen werden nach Fachbereichen und fachübergreifend nach Kategorien sortiert dargestellt und können kommentierend weiterentwickelt werden. Außerdem stehen für Universitätsangehörige Suchtools bereit.

Phase 2: Aus den Ideen bzw. Ideensträngen wählen Dozenten Projekte aus und lassen Mitglieder mit unterschiedlichen Rechten zu. Die Projektmitglieder arbeiten mittels Kollaborationssystemen zusammen. Als Hilfsmittel stehen den universitätsangehörigen Mitgliedern fachspezifische und allgemeine Visualisierungstools zur Verfügung. Neu hinzugekommene Mitglieder können die Projektentwicklung von der Ideengenerierung hin zum momentanen Projektstand verfolgen.

Phase 3: Das fertige Projekt wird für die Veröffentlichung aufbereitet und der Entstehungsprozess transparent gemacht, indem sowohl das fertige Produkt als auch der Produktprozess von der Idee zum Produkt gespeichert werden.

Dieses Modell, das der folgenden Systemübersicht vorangestellt ist, soll als Einführung in die Nutzungsmöglichkeiten von E-Infrastrukturen dienen und Denkanstöße geben für den Einsatz und die Kombination entsprechender Systeme. In den folgenden Kapiteln werden nun existierende Lösungen vorgestellt, mit denen sich solche oder ähnliche Plattformen aufbauen lassen. Natürlich erfüllen die Systeme auch jeweils für ihre Systemklasse spezifische Aufgaben und können für sich genommen spezielle Zwecke innerhalb einer Forschungsinstitution erfüllen. Auch dies soll im Folgenden deutlich werden.



Kap. 1: Information Retrieval

Relevancy Retrieval

Die Information Retrieval-Software Relevancy unterstützt alle Formen des inhaltsbasierten Informationszugriffs wie interaktive Suche, themenorientiertes Navigieren, Kategorisierung und Informationsfilterung. Mit der Standard-Software Relevancy bietet Eurospider ein kundenspezifisch konfigurierbares Produkt für hohe Ansprüche an. Eurospider entwickelt als Spin-off Firma der ETH-Zürich innovative Suchtechnologie für Bibliotheken, Archive und Informationsvermittlungsstellen.

Studien:

Enterprise-Search-Systeme im internen Wissensmanagement: Ergebnisse einer Studie zu Perspektiven der Unternehmen:

<http://www.eurospider.com/pdf/Datenbank-Spektrum%20302009.pdf>

Nutzung von Relevancy Retrieval aus Sicht der Online-Kommunikation:

http://www.zhaw.ch/fileadmin/php_includes/popup/publikation-detail.php?publ_id=1359

Anbieter:

Eurospider (ETH Zürich)

Referenzprojekte:

Relevancy Retrieval verschafft im Rahmen des Projekts KISS Knowledge Transfer («Knowledge Information Sharing System») den beteiligten Institutionen des ETH-Bereichs Zugriff auf ihre Forschungsprojekte und -resultate.

Das Projekt KISS umfasst alle relevanten Forschungsinformationen im ETH-Bereich. relevancy corporate retrieval erschliesst Daten in den verschiedensten Formaten aus Filesystemen, dem Intranet und externen Websites. Die einheitliche Benutzeroberfläche ermöglicht den geografisch verteilten Forschungsbereichen den Zugriff auf das vorhandene Wissen und unterstützt dessen Verteilung.

Funktionsumfang:

Relevancy Retrieval erschliesst alle unternehmensrelevanten Daten aus heterogenen Quellen. Das einheitliche Indexieren und intelligente Analysieren aller erschlossenen Daten gewährleistet eine zielgerichtete, benutzerspezifische Suche. Das System ermöglicht die natürliche Formulierung von Suchanfragen und stellt sicher, dass selbst Sprachwechsel, Beugungsformen oder einzelne Bestandteile zusammengesetzter Wörter erkannt und verarbeitet werden.

Interaktive Suchfunktionen unterstützen den Anwender bei der Verdichtung der Ergebnisse:



Der Search Guide analysiert eingegebene Begriffe, stellt Kontextfragen und präzisiert die Suchanfrage. So erkennt Relevancy Retrieval beispielsweise, dass der Suchbegriff „Emission“ in den verschiedenen Zusammenhängen „Bank“ und „Umwelt“ erwähnt wird und stellt dem Benutzer Kontextfragen. Oder der Benutzer markiert gefundene, relevante Dokumente und schickt so eine erneute, verfeinerte Suchanfrage ans System zurück.

Relevancy Retrieval übersetzt Anfragen automatisch in verschiedene Sprachen und identifiziert selbst Phrasen, Synonyme oder benutzerspezifisch definierte Einheiten. Anwender können zudem prüfen, ob ein Dokument auch in einer anderen Sprache vorhanden ist. Die sprachübergreifende Suche und Kategorisierung bieten wir in Deutsch, Englisch, Französisch, Italienisch und Spanisch; weitere Sprachen auf Wunsch.

Entscheidend für den schnellen Zugriff ist zudem die Gliederung der Informationen: In Kategorien werden Daten analog zugrunde liegenden Taxonomien, Topic Maps oder Thesauri des Unternehmens gebündelt. Relevancy Retrieval strukturiert die Informationen automatisch nach Themen, Organisationen oder Prozessen und ordnet neue Dokumente entsprechend zu. Individuell entwickelte, hochsensible Konzeptsensoren erkennen selbst Feinstrukturen von Daten und Inhalten und kategorisieren Informationen entsprechend.

Die offene Schnittstellenarchitektur von Relevancy Retrieval stellt die einfache Integration der Funktionalitäten in bestehende Informationsportale, CRM-Lösungen oder andere Applikationen sicher.

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * **Curate and Preserve**

Sequential Actions

- * Conceptualise
- * **Create or Receive**
- * Appraise and Select
- * Ingest
- * Preservation Action
- * **Store**
- * **Access, Use and Reuse**
- * **Transform**

**Systemvoraussetzungen:**

Betriebssystem

- Unix: Sun Solaris Sparc 8 und 10
- Linux: Kernel 2.2.x, 2.4.x oder 2.6.x, glibc 2.1 oder neuer

Hardware

- Server mit mindestens 1 GB RAM und 10 GB Harddisk (250 MB für Software; Platz für Daten richtet sich nach Anzahl Dokumente). Die Software ist je nach Hardware in der Anzahl Dokumente sowie in der Anzahl Anfragen skalierbar.

Fremdsoftware

- Webserver: Apache 1.3.x/2.x, Netscape/IPlanet, Microsoft IIS 5.x
- Datenbank: PostgreSQL 7.x oder Oracle 8/9, andere Datenbanken per ODBC3-Treiber, für den Webroboter ist der Einsatz einer Datenbank zwingend erforderlich
- Outside In 8.1.0 von Stellent für die Dokumentenkonversion
- PHP4 oder PHP5 für den Einsatz des Standard Interface
- JDK oder JRE 1.4.x für den Einsatz des Filterarchitekten

Lizenzierungsmodell:

Die Lizenzierung der Relevancy Software beinhaltet eine Einmallyzenz sowie eine jährliche Betriebslizenzen. Die Einmallyzenz ist eine Vergütung für die Programmierung der Relevancy-Software.

Einmallyzenz

Die Höhe der Einmallyzenz hängt von folgenden Faktoren ab:

- 1) Aktivierte Zusatzmodule
- 2) Anzahl erschlossener und auffindbarer Dokumente
- 3) Anzahl Zugriffe auf Relevancy in einem bestimmten Zeitraum

Weitere Zusatzmodule, mehr Dokumente, bzw. mehr Zugriffe erfordern eine aufwändigere Software und führen zu höheren Einmallyzenzen.

Betriebslizenzen

Die Betriebslizenzen ist eine Vergütung für die Pflege und Weiterentwicklung der Relevancy-Standardsoftware. Die jährliche Betriebslizenzen beträgt 18% der Einmallyzenz und beinhaltet das Recht, kostenlos neue Releases mit dem gleichen Funktionsumfang zu beziehen (exkl. Installations- und Konfigurationskosten). Die Unterstützung für den Betrieb und die Pflege sowie Anpassung kundenspezifischer Konfigurationen und Spezialentwicklung werden in einem separaten Support-Vertrag geregelt.

Dokumentation:

Siehe

<http://www.eurospider.com/relevancy/suche.htm>

<http://www.eurospider.com/relevancy/kategorisierung.htm>

<http://www.eurospider.com/relevancy/aktualisierung.htm>

<http://www.eurospider.com/relevancy/informationssteuerung.htm>



<http://www.eurospider.com/relevancy/kontextanalyse.htm>

Weitere Informationen:

http://www.eurospider.com/pdf/whitepaper_de.pdf

Empfehlung:

Der Hersteller Eurospider ist als Tochterunternehmen der ETH Zürich auf wissenschaftliche Bibliotheken und Archive spezialisiert und kann auf nachweisbare Projekterfolge in der Implementierung umfassender Information-Retrieval-Lösungen in verteilten Forschungsbereichen verweisen. Die Technologie zeichnet sich besonders durch die geführte Suche aus, die dem Nutzer dynamisch Hilfestellungen beim Auffinden eines Dokuments gibt.

IDOL Retina

IDOL Retina ist eine einfache Enterprise-Search-Lösung, die das volle Spektrum von Retrieval-Methoden unterstützt. Die Out-of-the-box-Software ist aufwandsarm zu administrieren und die Benutzeroberfläche intuitive zu bedienen.

Anbieter:

Autonomy Corporation PLC

Kooperationspartner:

Die Suchtechnologie basiert auf Forschungsergebnissen der University of Cambridge

Referenzprojekt:

Videosuche der Harvard Business School

<http://video.hbs.edu/videotools.html>

Larry Bouthillier, head of

multimedia production at HBS: "Simply put, video that is not searchable is not very useable,"

Funktionsumfang:

Benutzer-Identifizierung und -Authentifizierung durch login oder optional durch Verzeichnisdienste (LDAP, SSL LDAP).



Weitere Funktionen:

Retrieval

Conceptual Queries

Boolean Search

Parametric Search

Federated Search

Conversation

Real-Time Predictive Query

Automatic Summarization

Automatic Query Guidance

Advanced Analytics

Automatic Categorization

Automatic Clustering

Visualization

2D and 3D Cluster Maps

Spectrograph

Personalization

Agents

Profiling

Expertise Locator

Implicit Query

Während des laufenden Suchprozesses erhält der Nutzer kontinuierlich automatisch ermittelte Vorschläge für Inhalte, die zu den auf dem Bildschirm angezeigten Inhalten in Beziehung stehen.

Active Folders

Active Folders sind lernfähige Informationsstrukturen. Zudem kann der Nutzer selbst Inhalte in Active Folders ablegen, so dass die bevorzugten Inhaltstypen des Nutzers dem System bekannt werden.

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

**Systemvoraussetzungen:**

- Windows XP
- Macintosh OS X (Version 10.3 und höher)

Minimale Server-Spezifikationen:**Windows:**

- 1GHz Pentium Prozessor
- 256 MB RAM
- 400 MB min freier Festplattenspeicher

Mac OS X:

- Mac OS X Version 10.3 oder höher
- G4
- 256 MB RAM
- 400 MB min freier Festplattenspeicher

Lizenzierungsmodell:

Proprietär, Preise a. A. (Individualpreise)

Dokumentation (registrierungspflichtig):

<http://publications.autonomy.com/pdfs/Power/Product%20Briefs/Enterprise%20Search/Enterprise%20Desktop%20Search%20Technical%20Brief.pdf>

Weitere Informationen:

<http://www.autonomy.com/content/Products/enterprise-search/index.en.html>

Empfehlung:

IDOL Retina ist eine aufwandsarm zu installierende Software, die keine hohen Anforderungen an die IT-Infrastruktur stellt. Sie bringt dennoch einige Besonderheiten mit; neben der Ausrichtung auf Multimedia (Videosuche, Active Folders) wäre hier die besondere Stärke in der Suchvisualisierung zu nennen (Cluster Maps, Spektrographen). Durch die Kooperation mit der University of Cambridge ist ein Abgleich der Funktionalitäten mit wissenschaftlichen Anforderungen gewährleistet.



Ultraseek

Einer der Hauptvorteile von Ultraseek ist die zügige Installation in der vorkonfigurierten Standardvariante. Diese kann einfach und schnell an spezielle Bedürfnisse angepasst werden. Für die Integration in bestehende Infrastrukturen bietet Ultraseek eine Java API, eine Webservice-Schnittstelle und Portlets für die wichtigsten Portalhersteller an.

Stimme zum System:

PSECU

www.psecu.com

“Even though Ultraseek produces great search results out of the box, we like the ability to display specific Quick Link results based upon specific keywords. For example, earlier this year we ran a 'beat zero' keyword advertising campaign. By searching for the Quick Link keyword 'beat zero', you received a link to a page explaining why it costs more to take an auto manufacturer's 0% APR loan, rather than our 4.99% APR loan with the manufacturer's cash-back rebate. This brought current members, as well as potential members, to our site where we were able to advertised and explain even additional products and services we offer.”

Craig Petrou, Webmaster

Anbieter:

Ultraseek
5758 W. Las Positas Blvd., Suite 100
Pleasanton, CA 94588

Referenzprojekte:

Johns Hopkins Institutions

Die Johns Hopkins Institution hat Ultraseek innerhalb von 36 Stunden installiert, um mehr als 350.000 Dokumente aus rund 1.000 externen and internen Websites zu indexieren. Die renommierte medizinische Bildungseinrichtung brachte Ultraseek innerhalb eines engen Zeitplans online, um den Besucher-Traffic auf der Website www.hopkinsmedicine.org zu kapitalisieren.

http://publications.autonomy.com/pdfs/Ultraseek/Case%20Studies/MK0496_Johns_Hopkins_UIA.pdf



CERN

Physiker des CERN setzen Ultraseek ein, um relevante Informationen aus neun großen wissenschaftlichen Datensammlungen aus rund 5.000 Websites zu finden. Ultraseek führt dabei bis zu 3.000 Suchanfragen pro Tag aus, mit Spitzen von 200 Anfragen pro Stunde.

http://publications.autonomy.com/pdfs/Ultraseek/Case%20Studies/MK0481_CERN_UIA.pdf

Funktionsumfang:

Linguistisch optimierte Relevanzbewertung

Die Trefferliste wird nach einer Relevanzbewertung sortiert, die linguistische Techniken nutzt (z. B. Stemming, Linkanalyse sowie Wortgewichtung nach Suchbegriffslokalisierung und Frequenz innerhalb des Volltextdokuments). Der Suchindex kann durch individuelle Benutzerwörterbücher erweitert werden.

Sicherheitsfeatures

Mit Hilfe von Rollenmodellen und Filtermechanismen kann der Benutzerzugriff auf Dokumentenebene gesteuert werden. Jeder Benutzer kann nur die Dokumente anzeigen, für die er autorisiert ist. Von allen anderen weiß er nicht mal, dass sie existieren.

Page Expert

Der Indexierungsmechanismus von Ultraseek erkennt irrelevante Webseitenbestandteile wie Navigationselemente, Menüs oder Werbung und nimmt diese vom Index aus. Die Filter können an die Spezifika und Konventionen der Webseiten einer Organisation angepasst werden.

Multi-spider

Die Informationen einer Organisation liegen typischerweise in hunderten von Datenformaten vor und sind verteilt auf Webservern, Dateisystemen und Datenbanken. Der Multi-spider der Ultraseek-Software führt alle diese Inhaltstypen in einem einheitlichen Suchindex zusammen. Ultraseek kann sogar JavaScript spidern und damit dynamisch generierte Websites indexieren.

Contextual Summaries

Automatisch erstellte Textzusammenfassungen ermöglichen es den Endanwendern, sich einen Eindruck vom Inhalt eines Dokuments zu verschaffen, ohne es öffnen zu müssen. Für den Anbieter reduziert sich so die Netzlast auf dem Server, da Volltextdokumente weniger oft heruntergeladen werden müssen.

Spelling Suggest ("Meinten Sie"-Funktion)

Ultraseek nutzt ein dynamisch aus dem Index erstelltes Wörterbuch, um alternative Schreibweisen zur Nutzereingabe vorzuschlagen. So können viele Suchen, die wegen



fehlerhafter Schreibweisen sonst erfolglos gewesen wären, noch zum Ziel gebracht werden.

“No-hits” Forwarding

Anwender, die auf ihre Anfrage keine Suchergebnisse erhalten, können ihre Suche erweitern, indem sie weniger restriktive Suchterme anklicken, die ihnen dynamisch von Ultraseek vorgeschlagen werden.

Automatische Erzeugung von Dokumenttiteln

Das System, nach dem Dokumente, die im Tagesgeschäft einer Organisation erzeugt werden, benannt werden, ist in der Regel nutzerspezifisch und folgt keinen organisationsweiten Regeln. Ultraseek kann Dokumenttitel nach einheitlichen, automatisch generierten Konventionen neu vergeben und hilft den Anwendern damit, das richtige Dokument schneller zu finden.

Multilinguale Suche

Die multilinguale Suche von Ultraseek unterstützt die vollständige grammatikalische und lexikalische Analyse folgender Sprachen:

- Chinesisch (Traditional & Simplified)
- Dänisch
- Holländisch
- Englisch
- Finnisch
- Französisch
- Deutsch
- Italienisch
- Japanisch
- Koreanisch
- Norwegisch
- Portugiesisch
- Spanisch
- Schwedisch

Die folgenden Schnittstellen und Funktionen helfen bei der Integration von Ultraseek in bestehende Infrastrukturen:

- Web Services (.NET und J2EE)
- Java API
- Vorkonfigurierte Portlets
- Upgrade Wizard
- Layout Manager

**Funktionen bezogen auf Phasen des Data-Lifecycle:****Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * **Curate and Preserve**

Sequential Actions

- * Conceptualise
- * **Create or Receive**
- * Appraise and Select
- * Ingest
- * Preservation Action
- * **Store**
- * **Access, Use and Reuse**
- * **Transform**

Systemvoraussetzungen:

Ausführliche Beschreibung der wichtigsten technischen Voraussetzungen, die bei der Einführung von Ultraseek zu beachten sind:

http://www.ultraseek.com/support/docs/pdf/Ultra5.6_Implement.pdf

Lizenzierungsmodell:

Proprietär, Preise a. A. (Individualpreise)

Dokumentation:

http://www.ultraseek.com/support/docs/pdf/Ultra5.6_Implement.pdf

Weitere Informationen:

<http://www.ultraseek.com>

Empfehlung:

Ultraseek ist eine umfassende Information-Retrieval-Lösung, die vor Allem für große Institutionen mit umfangreichen Daten-Repositories Sinn macht. Dort kommen die sprachtechnologischen Features wie die linguistisch optimierte Relevanzbewertung oder die „Meinten Sie“-Funktion zur Geltung. Weitere Besonderheiten, die andere Retrieval-Systeme nicht bieten, sind die automatisch erzeugten Textzusammenfassungen und Dokumenttitel. So kann der indexierte Content selbst auf optimale Auffindbarkeit hin angepasst werden.



FAST ESP

Die geführte Suche von FAST gestaltet die Recherche als Konversationsprozess zwischen Nutzer und Suchmaschine. Dieser Prozess ist an die Bedürfnisse der einsetzenden Organisation anpassbar.

Die Lösung von FAST ist hochskalierbar und daher auch für sehr große Infrastrukturen und Informationsmengen geeignet.

Strukturierte Inhalte können mit unstrukturierten Dokumenten in einem Index zusammengeführt werden. Auch Dokumente außerhalb der eigenen Organisation können indexiert werden.

Die Lösung unterstützt die folgenden linguistischen Funktionalitäten:

Spell Checking	Detects and corrects miss-spelled words and phrases
Lemmatization	Matches regular and irregular grammatical inflections
Phrase Detection	Detects names and phrases and either automatically rewrites the query or provides a search tip
Phonetic Search	Enables approximate matching based on phonetic similarities
Anti-Phrasing	Removes words that do not contribute to the query's meaning
Natural Language	Recognizes constructs such as who, what, where, when, and why
Thesaurus	A custom synonym dictionary to define your specific business terms

Vgl. http://download.microsoft.com/download/7/3/A/73AB39D7-52FE-4EB0-8941-AFF78BDD35F0/FAST_ESP_Data_Sheet.pdf

Anbieter:

FAST, A Microsoft Subsidiary

Referenzprojekte:

Standardisiertes Beispielprojekt: http://download.microsoft.com/download/7/4/0/740F293C-0869-4A02-9D44-36021D4636B8/Best_Practices_in_a_Search_Project.pdf

**Funktionsumfang:**

- **Skalierbarkeit**
Geeignet für extreme Anfragenlast und Inhaltsmengen. Lineare Skalierbarkeit für High-End-Performance.
- **Robustheit des Content Retrieval**
Strukturierte und unstrukturierte Informationen aus hunderten von Dokumenttypen können mit Hilfe eines konfigurierbaren, flexiblen Web Crawlers aggregiert und transformiert werden. Über 85 Sprachen werden unterstützt.
- **Konfigurierbare Inhaltsmodule zur Verbesserung der Suchergebnisse**
Suchinhalte können zum Zeitpunkt der Indexierung in vorgefertigten oder nutzerdefinierten Modulen gruppiert werden. Dafür stehen umfangreiche linguistische Textanalysefunktionalitäten zur Verfügung.
- **An den Content anpassbare Suchergebnis-Ausgabe**
Die Verarbeitung von Sucheingaben und –ergebnissen sowie die Relevanzbewertung kann durch die einsetzende Organisation an den Charakter ihrer indexierten Inhalte angepasst werden. Individuell erstellte Navigationspfade, Taxonomien und Informationscluster erleichtern dem Nutzer den Weg zum Antwort auf seine Frage.
- **Programmierbare Personalisierung**
Das “FAST Recommendations”-Modul erweitert die Suche um ein umfassendes personalisierbares Empfehlungssystem auf Basis dynamischer Benutzerprofile.
- **Exprierte und Dienstleistungen des FAST-Partner-Netzwerks**
FAST bietet ein umfangreiches Partner-Netzwerk an, über das zahlreiche Dienstleistungen rund um die Implementierung und Anpassung der FAST-Suchtechnologie verfügbar sind.

Funktionen bezogen auf Phasen des Data-Lifecycle:**Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Je nach System-Infrastruktur

Lizenzierungsmodell:

Preise a. A.

Dokumentation:



FAST Technical Resources: <http://www.microsoft.com/enterprisesearch/en/us/FAST-technical.aspx>

Weitere Informationen:

Produktbroschüre: http://download.microsoft.com/download/B/F/D/BFD2EBD0-1100-4821-B0EA-4BAC1535B3A6/FAST_ESP_Brochure.pdf

FAQ: <http://www.microsoft.com/pathways/fast/FAQ.htm>

Unterlagen zum Workshop „Einsatz von FAST-Suchtechnologie im bibliothekarischen und wissenschaftlichen Umfeld“ des Hochschulbibliotheksentrums NRW : <http://www.hbz-nrw.de/dokumentencenter/produkte/dlk/aktuell/vortraege/FAST-Workshop/>

Empfehlung:

Das norwegische Unternehmen FAST Enterprise Search, seit 2008 Teil des Microsoft-Konzerns, zeichnet sich durch die hohe Skalierbarkeit seiner Lösungen aus. Sie eignen sich daher vor allem für große Organisationen, die aufgrund großer Datenmengen und hoher Zugriffszahlen erhöhte Anforderungen an die Performanz ihrer Suche haben. FAST bietet zudem eines der umfassendsten am Markt befindlichen Sets an linguistischen Textanalyse-Features. Zu erwähnen ist außerdem die erweiterte Sprachunterstützung (85 Sprachen).

Coveo Enterprise Search

Coveo Solutions Inc. ist ein auf Enterprise Search spezialisierter Anbieter, der seit 2005 das Produkt Coveo Enterprise Search (CES) anbietet. Das Produkt ist mit über 700 Kunden in Nordamerika und Europa inzwischen zur festen Größe im Bereich der organisationsweiten Suche geworden.

Das Produkt basiert auf der .NET-Technologie von Microsoft und läuft ausschließlich auf windowsbasierten Systemen. Gleichwohl ist die Indexierung von Inhalten aus externen Systemen wie Oracle-Datenbanken, SAP oder Webservern möglich.

Die Coveo Enterprise Search Platform kombiniert strukturierte und unstrukturierte Daten in einem einheitlichen Index ohne die Ausgangsdaten zu verschieben. Dadurch kann eine Suche über unterschiedliche Datenquellen innerhalb einer einheitlichen Suchoberfläche implementiert werden. Coveo unterstützt Integration in E-Mail-Systeme (Mail-Indexierung und Outlook-Sidebar) und Desktop-Oberflächen (Floating Desktop Sidebar) sowie Elemente der Social Search (Jive- und Confluence-Integration).

Anbieter:

Coveo Solutions Inc



Funktionsumfang:

Coveo Enterprise Search zeichnet sich durch eine besonders einfache Installation und Administration aus. Das Deployment ist typischerweise innerhalb von unter einer Stunde möglich. Dazu muss man noch die Zeit rechnen, die zur Indexierung der Dokumente benötigt wird. Je nach Hardware ca. 20.000 - 50.000 pro Stunde (bei sehr leistungsfähiger Hardware auch mehr).

Da die Suche browserbasiert ist, muss auf dem Client nichts installiert werden. Als Webbrowser werden IE 5.5+ oder Firefox 2+ unterstützt.

Trotz der einfachen Installation und Administration skaliert Coveo Enterprise Search von 100.000 bis hin zu mehreren 100 Millionen Dokumenten. Durch die Möglichkeit, Indizes auf mehrere (auch geografisch getrennte) Server zu verteilen, lassen sich die Anforderungen von Unternehmen flexibel abbilden.

Die Dokumentensicherheit wird basierend auf der Active-Directory-Sicherheit automatisch als Metadaten zu jedem Dokument indexiert. Damit werden die über Active Directory eingerichteten Zugriffsrichtlinien direkt in den Index übernommen und damit sichergestellt, dass jeder Anwender nur die Dokumente findet, auf die er auch Zugriff hat.

Neben der eigentlichen Anwendung steht eine leistungsfähige API zur Verfügung, über die man Suchfunktionalitäten nahtlos in eigene Anwendung integrieren kann. Hier stehen als Schnittstelle COM+ oder Webservices bereit.

Neben dem einfachen, google-ähnlichen Suchinterface integriert sich Coveo in Microsoft SharePoint, WSS 2.0 und 3.0 sowie MOSS 2007 und ersetzt die normale SharePoint-Suche. Damit kommen Anwender von SharePoint-Systemen neben anderen Leistungsmerkmalen ebenfalls in den Genuss von Wildcards und komplexeren logischen Verknüpfungen in Suchabfragen.

Neben den üblichen Quellen wie Filesystem und Webserver liefert Coveo Enterprise Search ebenfalls Konnektoren zu Microsoft Exchange, Salesforce.com, Enterprise Vault, Documentum, Lotus Notes, Novell und Datenbanken über ODBC sowie zu allen Systemen, die Google OneBox unterstützen.

Alle gängigen Dokumentenformate wie MS Office, PDF, HTML, XML, ZIP und mehr sind im Lieferumfang enthalten. Erweiterbar ist der Zugriff über iFilter, die auch für exotischere Dokumentenformate verfügbar sind. Außerdem ist über einen Open Crawler die Erschliessung völlig eigener Dokumentenformate möglich.

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

**Systemvoraussetzungen:**

- Windows 2000 Server SP3 oder höher
- IIS 5.0 oder höher
- Prozessor 2 GHz
- Hauptspeicher 1 GB
- .NET Framework 2.0
- 70 MB Festplattenspeicher für Applikation
- 200 MB Festplattenspeicher oder mehr abhängig von der Zahl der indexierten Dokumente

Lizenzierungsmodell:

Die Lizenzierung der Coveo Enterprise Search basiert auf der Anzahl der indexierten Dokumente. Es können beliebig viele Anwender auf das System zugreifen. Ebenfalls unbegrenzt ist die Anzahl der Suchserver-Instanzen. Für ein System, das 100.000 Dokumente indexiert, fallen Kosten in Höhe von 8.000,- Euro an. Die Lizenz ist zeitlich unlimitiert. Zusätzlich gibt es einen Support- und Updateplan. In diesem Servicevertrag sind sämtliche Releasewechsel und zusätzlich freigegebene Konnektoren enthalten. Für den Support- und Updateplan fallen jährliche Kosten in Höhe 20% des Lizenzpreises an.

Dokumentation:

http://www.coveo.com/en/services-and-support/support/knowledge-base/~media/Files/Support/Knowledge-Base/information-articles/Installation-Guide-Typical-Setup_Coveo-Enterprise-Search-6.ashx

Weitere Informationen:

<http://www.coveo.com>

Empfehlung:

Coveo Enterprise Search ist eine für Windows-Infrastrukturen empfehlenswerte Lösung. Sie integriert sich gut in andere Microsoft-Applikationen wie z. B. Sharepoint oder Exchange. Als Besonderheit ist die Indexierbarkeit von E-Mails über Outlook sowie die Integration auf dem Desktop über eine Floating Sidebar zu nennen. Coveo bietet neben offenen Standardschnittstellen auch Integrationsmöglichkeiten zu weit verbreiteten proprietären Lösungen wie Lotus Notes oder Novell an. Hinsichtlich der indexierbaren Datenformate ist Coveo besonders flexibel, was die Lösung für Organisationen mit speziellen wissenschaftlichen Dokumenten interessant macht.



Kap. 2: Kollaborationssysteme

BSCW

Mit BSCW können Anwender auf einfache Weise Dokumente in gemeinsamen Arbeitsbereichen einstellen und so anderen berechtigten Mitarbeitern den Zugriff ermöglichen. Versionsverwaltung und Sperrmechanismen erlauben eine komfortable Dokumentverwaltung.

Online-Umfragen

Über Umfragen können Meinungsbilder zu bestimmten Themen ermittelt und anschaulich in grafischer Form präsentiert werden.

Kontaktlisten

Mit BSCW sind Kontakte für alle Gruppenmitglieder immer aktuell verfügbar. Mittels Kontaktlisten in gemeinsamen Arbeitsbereichen lassen sich etwa Kundenkontakte schnell und einfach austauschen und bearbeiten (Import/Export in das Standard-Format vCard möglich).

Termine und Aufgaben

Die Kalenderkomponente von BSCW beinhaltet neben Gruppenkalendern auch Serientermine und eine Erinnerungsfunktion. Eine Aufgabenfunktion ermöglicht das Anlegen, Zuweisen und Nachverfolgen von Aufgaben sowie das Verknüpfen von Dokumenten mit Aufgaben.

Diskussionsforen

Benutzer können Diskussionen führen, das System verwaltet die Argumentationsketten (threads) und präsentiert sie in übersichtlicher Form.

Annotationen

Dokumente können von verschiedenen Benutzern mit Notizen und Bewertungen versehen werden.

Stimme zum System:

"Auf dem BSCW-Server haben wir unsere gesamte Schulstruktur, Schulorganisation abgebildet. Es hat zwar einige Zeit gedauert, bis die Kollegen den Vorteil gemeinsamer Teamarbeit auch im Internet und Intranet erkannt haben, aber jetzt benutzen wir diesen Server für die pädagogische Dokumentenverwaltung unserer Kollegen. Dokumente, die bisher individualisiert waren, werden nun auch dem Team sichtbar gemacht."

Mercator Berufskolleg, Moers

Anbieter:



Fraunhofer FIT & Orbi-Team Software GmbH

Referenzprojekte:

ETH Zürich

An der ETH Zürich wird BSCW erfolgreich zur webbasierten Teamarbeit mit Studierenden eingesetzt. Die neuen Informations- und Kommunikationstechnologien eröffnen der Ausbildungsmethodik dabei neue und vielversprechende Dimensionen.

http://www.bscw.de/eth_zuerich.html

Funktionsumfang:

Authentifizierung

Nur durch Name und Passwort identifizierte Benutzer erhalten Zugang zu BSCW-Arbeitsbereichen.

Versionsmanagement

Dokumente in verschiedenen Versionen verwalten — zur optimalen Unterstützung gemeinsamer Dokumentenbearbeitung.

Dokumentübertragung

Es stehen verschiedene Mechanismen zur komfortablen Dokumentübertragung zur Verfügung – Anwender können Dokumente einfach per Drag and Drop in das System einstellen.

Mechanismen zum Sperren von Dokumenten

Dokumente in Bearbeitung können für den Bearbeitungszeitraum für Schreibzugriffe durch andere gesperrt werden.

Diskussionsforen

Benutzer können Diskussionen führen, das System verwaltet die Argumentationsketten (threads) und präsentiert sie in übersichtlicher Form.

Annotationen

Dokumente können von verschiedenen Benutzern mit Notizen und Bewertungen versehen werden.

Ereignisbenachrichtigung

BSCW bietet vielfältige Möglichkeiten, um die Benutzer über aktuelle Vorgänge in den gemeinsamen Arbeitsbereichen zu informieren. Dazu zählen die direkte Benachrichtigung über Ereignisse per Email sowie die Versendung eines täglichen Berichts. Benutzer können genau spezifizieren, über welche Ereignisse und in welcher Form sie informiert werden möchten.



Differenzierte Zugriffsrechte

Einzelne Dokumente können für bestimmte Benutzer z. B. als nur lesend nutzbar definiert werden, für andere kann z.B. auch anonymer Zugang aus dem WWW zugelassen werden. Für alle Benutzeraktionen und Datentypen können Zugriffsrechte gesetzt werden.

Suchfunktionen

Objekte in Arbeitsbereichen können nach Inhalt oder Eigenschaften wie Autor oder Datum der letzten Änderung gefunden werden; außerdem können Suchergebnisse aus dem Web direkt in einen Arbeitsbereich übernommen werden. BSCW unterstützt eine inhaltsbezogene Suche in Dokumenten verschiedenster Formate mittels einer indexbasierten Suche.

Archivierungsfunktionen

Speicherung einer Anzahl von Dokumenten in einem Archiv, z.B. zum Austausch zwischen Client und Server.

Versenden von Dokumenten

Auswahl, Adressierung und Versand von Dokumenten (per Email) direkt aus dem Arbeitsbereich.

Zugang über mobile Endgeräte

Mit BSCW wird Wissensaustausch und Zusammenarbeit auch über mobile Endgeräte wie Smartphones oder PDAs möglich. Es stehen Schnittstellen für mobile Web-Browser zur Verfügung, die jeweils eine angepasste Benutzerschnittstelle für den mobilen Zugang bereitstellen.

Integrierter HTML-Editor

Ein integrierter Editor erlaubt die direkte Bearbeitung von Web-Dokumenten (im HTML-Format) in der Dokumentenverwaltung des BSCW-Systems.
Serientermine und Erinnerungsfunktion

Schnittstellen

Zur Anbindung von Fremdsystemen stehen verschiedene Schnittstellen zur Verfügung. Das System bietet u.a. Schnittstellen zu synchronen Kooperationsformen wie etwa Audio/Video-Konferenzen. Über Schnittstellen zu Konvertierungsdiensten können Dokumente in das von dem Benutzer bevorzugte Dokumentformat transformiert werden.

Individualisierte Benutzeroberflächen

Verschiedene Individualisierungsoptionen und Benutzerprofile ermöglichen dem Anwender die Anpassung der Benutzeroberfläche an seine Wünsche.



Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Systemanforderungen Endbenutzer (Client)

Endbenutzer greifen direkt über ihren Web-Browser auf die für die Erledigung ihrer Aufgaben erforderlichen Daten zu und können diese schnell und einfach bearbeiten. Die Systemanforderungen sind entsprechend gering:

- * Standard-Web-Browser: Internet Explorer, Opera, Firefox, Safari, Konqueror usw.
- * JavaScript/Java für komfortable Bedienung (optional)
- * Plattformen: PC (Windows), Unix/Linux, MacOS
- * alternativer Zugang über WebDAV-Applikationen und MS Office
- * Es ist keine zusätzliche Software-Installation notwendig.

Systemanforderungen Server

Die BSCW-Server-Software unterstützt zahlreiche Server-Plattformen und ist einfach zu installieren. Der laufende Betrieb erfordert einen geringen Wartungsaufwand und garantiert so einen kostengünstigen Einsatz. Die Systemanforderungen sind wie folgt:

- * UNIX-Betriebssysteme: Linux, Solaris, HP-UX, AIX, IRIX, MacOS X, BSD
- * Apache-Web-Server oder Standard-HTTP-Web-Server
- * Windows-Betriebssystem: Windows 7/2008/Vista/2003/XP
- * Apache-Web-Server oder Microsoft Internet Information Server (IIS)
- * Email-Anbindung: SMTP-Server (UNIX/Windows) oder Sendmail (UNIX)
- * Zur Installation erforderliche Software-Pakete:
 - o Python Runtime Environment - version 2.5 oder 2.6
(www.python.org , Open Source)
 - o Auf einer Windows-Server-Plattform benötigen Sie zudem die Python-Erweiterungen für Windows ab Build#210
(sourceforge.net/projects/pywin32)
- * Eine zusätzliche Datenbank ist nicht erforderlich.

**Lizenzierungsmodell:**

Shared Group Hosting:

- für den **1. bis 20.** Benutzer **6,- €** pro Benutzer,
- für den **21. bis 50.** Benutzer **5,- €** pro Benutzer und
- ab dem **51.** Benutzer **4,- €** pro Benutzer.

Siehe http://www.bscw.de/shared_group_hosting.html

Server-Lizenzen (Preisbeispiele):

kleine Firma oder Projektgruppe

20 Benutzer: 999,- €

z.B. Architekturbüro, das über BSCW ein Bauvorhaben organisiert.

mittlere Firma / mehrere Projekte / externe Teilnehmer

50 Benutzer: 2.499,- €

z.B. Firma, die über BSCW mit externen Vertriebspartnern kommuniziert.

große Firma

300 Benutzer: 8.999,- €

z.B. Dienstleister mit verteilten Standorten in ganz Europa, Ablage wichtiger Firmen-Infos in BSCW, Arbeitsgruppen für Abteilungen und Standorte ..

Siehe http://www.bscw.de/bscw_server.html**Dokumentation:**<http://www.bscw.de/documentation.html>*Weitere Informationen:*<http://www.bscw.de/>

**Empfehlung:**

BSCW ist für Organisationen empfehlenswert, die einen Schwerpunkt auf zentrale Dokumentenverwaltung legen. Zu diesem Zweck bietet BSCW eine differenzierte Zugriffsteuerung sowie eine sinnvolle dokumentorientierte Suche. Über mobile Endgeräte kann man auch unterwegs auf seine Dokumente zugreifen, was die Lösung für Organisationen mit vielen externen Mitarbeitern, z. B. Lehrbeauftragten, interessant macht. Die Möglichkeit, BSCW über Schnittstellen mit Audio- und Videokonferenzsystemen zu verbinden, rundet das Angebot ab. Da die Anwendung vollständig browserbasiert ist, erfordert sie nur einen geringen Installations- und Wartungsaufwand. Auch die Lizenzpreise (Mehrbenutzerlizenzen) sind überschaubar, so dass auch kleinere Organisationen die Lösung wirtschaftlich abbilden können.

Opengroupware.org

Die Philosophie der Opengroupware-Initiative stellt den Einsatz offener Protokolle noch über das Bereitstellen offener Quellen. So wird eine sehr hohe Flexibilität für die Integrierbarkeit der Lösung in bestehende IT-Infrastrukturen erreicht. Auch andere OpenSource-Groupware-Projekte wie phpGroupWare oder Kolab werden nicht als Konkurrenz gesehen, sondern als Kooperationspartner.

In die Bedienbarkeit der web-basierten Benutzeroberfläche wird die gleiche Sorgfalt investiert wie es bei einer nativen Desktop-Software der Fall wäre. Für den Zugriff ohne Internetzugang werden Offline- und Mobile-Lösungen bereitgestellt. So wird an einer Integration in Outlook, Glow und Evolution gearbeitet.

Anstatt proprietärer Datenbanksysteme kommen Standard-SQL-Datenbanken zum Einsatz, deren Schema gut dokumentiert ist. Außerdem gibt es einen Adapter-Mechanismus, der eine freie Wahl des Datenbanksystems ermöglicht.

Anbieter:

Open Source

Funktionsumfang:

Adressmanager

Speichert und organisiert tausende von persönlichen und firmenbasierten Kontaktdaten. Einfach konfigurierbar und mit erweiterten Such- und Kategorisierungsfunktionalitäten.



Gruppenkalender

Verwaltung von Sitzungs- und Ereignisterminen für Gruppen und Individuen. Termine können annotiert oder mit Kontakt- und Projektdaten verknüpft werden.

Ressourcenplaner

Einsatzplanung für organisationseigene Ressourcen wie Besprechungs- oder Seminarräume, Projektoren oder Automobile. Automatisches Auffinden verfügbarer Ressourcen, Prüfung von Ressourcenkonflikten bei der Erstellung eines Termins.

Task Manager

Aufgaben können nach Person, Gruppe oder Projekt kategorisiert werden. Todo-Listen können nach Priorität, Enddatum oder Bearbeitungsstatus sortiert werden. Ein Überblick über alle Tasks ist über die Anwendung „Projekte und Dokumente“ (s. u.) verfügbar. Auf der Benutzerseite finden sich alle Tasks des jeweiligen Benutzers.

E-Mail-Client

Ein integrierter, IMAP4-basierter E-Mail-Client stellt die standardmäßigen Funktionen einer E-Mail-Software zur Verfügung.

Projekte und Dokumente

Dokumente und Dateien können lokal oder remote sowie in Gruppen oder individuell in einer projektorientierten Umgebung gespeichert werden. Projekte können mit Kontaktdaten oder Tasks mit Projekten verknüpft werden. E-Mails und Dokumente wie z. B. Faxe können in einem Dokumentenarchiv abgelegt werden, das mit jedem Projekt verknüpft werden kann.

Newsboard

Das Newsboard ermöglicht die Publikation wichtiger Informationen innerhalb der virtuellen Projektumgebung. Außerdem gehen aus dem Newsboard kommende Termine und Tasks hervor und dient somit als individuelle Startseite des Benutzers.

Palm Sync

Informationen, die in den Opengroupware-Anwendungen gespeichert sind, können über die Palm Sync-Applikation mit dem Palm-PDA synchronisiert werden.

Administration

Die feingranuliert anpassbaren Optionen sowie die Benutzerverwaltung können über ein einfaches und intuitives Web-Interface gesteuert werden.



Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * **Community Watch and Participation**
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * **Create or Receive**
- * **Appraise and Select**
- * **Ingest**
- * Preservation Action
- * **Store**
- * **Access, Use and Reuse**
- * **Transform**

Systemvoraussetzungen:

Evaluation und Test

- Intel, AMD, oder PowerPC CPU 1 GHz
- 512 MB RAM (ohne Funambol, sonst 1 GB RAM)
- 1 GB Festplattenplatz

Echtbetrieb

- Intel, AMD oder PowerPC CPU 3 GHz
- 2048 MB RAM
- 10 GB Festplattenplatz (ohne Mailspeicher)

Desktop

- Intel, AMD, oder PowerPC CPU 1.5 GHz
- 1024x768 Bildschirmauflösung
- 512 MB RAM
- Netzwerkverbindung

Microsoft Windows

- Microsoft Windows XP SP2 oder Vista

Apple Mac OS X

- Apple Mac OS X 10.2 bis 10.6

Linux

- Jede GNU/Linux-Distribution

Mobile Endgeräte

- Jedes Endgerät, das den SyncML 1.0 or 1.1-Standard unterstützt.

Empfohlen

- Palm OS-basierte Endgeräte mit Synthesis SyncML-Client
- "Research In Motion (RIM)"-BlackBerry- Geräte mit Funambol-Client
- Microsoft Windows Mobile PocketPC or SmartPhone mit dem Funambol-Client
- Apple iPhone / iPod mit Apple iPhone OS 3.0 or 3.1

**Lizenzierungsmodell:**

General Public License (GPL)

Dokumentation:

Entwicklerdoku: <http://www.opengroupware.org/en/devs/docs/index.html>

Installationsdoku: <http://www.scalablego.org/files/docs/SOG%20Installation%20Guide.pdf>

Weitere Informationen:

<http://www.scalablego.org/english.html>

Empfehlung:

Opengroupware ist eine der verbreitetsten Open Source-Initiativen im Bereich der Kollaborationssysteme. Die Anwendung kann daher als erprobt gelten. Ein Scherpunkt wird dabei auf die Integrierbarkeit in bestehende IT-Infrastrukturen über Standardschnittstellen gelegt. So ist die Lösung auch für alle drei großen Betriebssysteme (Windows, Linux, Mac) verfügbar und bietet auch mobile Zugriffsmöglichkeiten an. Funktional bietet die Software alle wichtigen Funktionalitäten, die für die Koordination und Zusammenarbeit von Arbeitsgruppen über das Internet Standard sind. Lediglich eine kollaborative Echtzeit-Dokumentbearbeitung fehlt.

MS Sharepoint

SharePoint ist eine integrierte Suite mit Serverfunktionen, die dazu beitragen können, die Effektivität einer Organisation zu verbessern. Hierzu zählen umfassende Content Management-Funktionen, unternehmensweite Suchläufe, Beschleunigung gemeinsamer Prozesse und einfache, Bereichsgrenzen überschreitende gemeinsame Datennutzung für einen besseren Einblick in die Abläufe innerhalb der Organisation. Sharepoint unterstützt alle Intranet-, Extranet- und Webanwendungen in der gesamten Organisation innerhalb einer einzigen integrierten Plattform, sodass sich separate fragmentierte Systeme erübrigen.

Darüber hinaus stellt dieser Kooperations- und Content Management-Server für IT-Fachleute und Entwickler die Plattform und die Tools bereit, die für die Serververwaltung, die Erweiterung von Anwendungen und die Interoperabilität benötigt werden.

Anbieter:

Microsoft



Funktionsumfang:

- Verwenden von Websitevorlagen zur Zusammenarbeit oder zum Verwalten von Besprechungen
- Gemeinsames Nutzen von Dokumenten, Kontakten Aufgaben und Kalendern
- Einfaches Brainstorming mithilfe von Wikis
- Weitergeben von Ideen mithilfe von Blogs
- Empfangen von Aktualisierungen an Listen und Bibliotheken mithilfe von RSS (Really Simple Syndication)
- Verwalten von Projekten (Grafische Darstellung von Projektaufgaben und Arbeitspaketen als Gantt-Diagramm)
- Mobiler Zugriff auf Inhalte
- Offlineverwaltung von Dokumenten für den Zugriff ohne Internetverbindung

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Unterstützte Betriebssysteme: Windows 7; Windows Server 2003; Windows Server 2008; Windows Vista; Windows XP

Für Web Front End Farm Server, Application / Index Server, Database Server siehe <http://geekswithblogs.net/MainaD/archive/2007/10/17/116101.aspx>

Lizenzierungsmodell:

Servers	Estimated Price
Office SharePoint Server 2007 (CALs also required)	\$4,451
Office SharePoint Server 2007 for Internet Sites (no CALs required)	\$41,134
Microsoft Search Server 2008 (no CALs required)	\$14,420
Microsoft Search Server 2008 Express (no CALs required)	Free Download
Client Access Licenses	Estimated Price
Office SharePoint Server 2007 Standard CAL	\$94
Office SharePoint Server 2007 Enterprise CAL ¹	\$75

NOTE CALs for Office SharePoint Server 2007 can also be purchased as part of the Microsoft Server CAL Suites (Microsoft Core CAL Suite and Microsoft Enterprise CAL Suite), which offer increased flexibility, simplified licensing across a variety of products, and increased value. For more information on these licenses, discuss the details with your Microsoft Representative or reseller or [visit the CAL Suites site](#).

¹To be licensed for the Enterprise Edition functionality of Office SharePoint Server 2007, both the Standard and Enterprise CALs are required.



vgl-.<http://office.microsoft.com/en-us/sharepointserver/FX102176831033.aspx>

Dokumentation:

Allgemeine Sharepoint-Dokumentation: <https://sp01.uibk.ac.at/Seiten/Default.aspx>

Weitere Informationen:

Developer Evaluation Guide:

<http://www.microsoft.com/downloads/details.aspx?FamilyID=cffb14e8-88a9-43bd-87aa-4792ab60d320&displaylang=en>

Sharepoint im Einsatz: http://www.evocom.de/Portals/0/MS-SharePoint_byJochenBaur.pdf

Empfehlung:

Sharepoint richtet sich an Anwender, die ein umfassendes Kollaborationssystem aus einer Hand wünschen. Über die Standardkomponenten wie Gruppenkalender und Dokumentenverwaltung hinaus legt die Lösung einen Schwerpunkt auf die Verwaltung von Projekten. Zudem bringt die Suite Social-Software-Elemente wie Wikis und Blogs gleich mit. Die Komfortabilität einer integrierten Lösung erkaufte man sich allerdings durch eine stark eingeschränkte Systemoffenheit: Sharepoint läuft nur in Microsoft-Umgebungen.

Sonexis Conference Manager

Der Sonexis ConferenceManager ist eine Kollaborationsplattform, die in der einsetzenden Organisation gehostet wird und somit keine regelmäßigen Betriebsgebühren erfordert. Die Plattform ist, besonders in den USA, in hunderten führenden Organisationen im Einsatz und ist speziell ausgerichtet auf die einfache Integration in bestehende Telefonie- und Netzwerk-Infrastrukturen.

Stimme zum System:

"We looked at two additional in-house solutions in addition to the Sonexis system," said Marty Alsip, Senior Telephony Engineer for General Cable. "We chose the Sonexis system because it was very price competitive and scaled to meet our needs now and in the future. It integrated seamlessly into our existing voice and data networks and the Outlook integration was very easy to use. The installation went smoothly and the training and support we received were excellent. My opinion is that Sonexis has one of the best support staff I have ever worked with. The Sonexis Customer Care team deserves a big pat on the back for the job they do, and I would recommend Sonexis to anyone."

General Cable

**Anbieter:**

Sonexis
Tewksbury, Massachussets 01876, USA

Referenzprojekt:**Datatec**

Datatec Systems ist ein IT-Unternehmen, das sich auf die Bereitstellung und Implementierung von Netzwerk-Technologien für Großunternehmen konzentriert. Für das Anrufvolumen seiner rund 400 Außendienstmitarbeiter fielen über 8.000 \$ pro Monat an Telefonkosten an. Datatec wählte Sonexis ConferenceManager, um zusätzliche Web-Conferencing-Funktionen nutzen zu können und gleichzeitig Kosten zu reduzieren. Die Lösung zahlte sich innerhalb von vier Monaten aus. Die kundenspezifischen virtuellen Konferenzräume ermöglichen es Datatec, bis zu 50 Experten in Ad-hoc-Konferenzen zusammenarbeiten zu lassen.

Funktionsumfang:

Conferencing-Funktionalitäten

- Gemeinsame Bearbeitung und Annotation von Spreadsheets, Präsentationen und anderen Dokumenten
- Kollaboratives Browsen von Websites
- Whiteboards für Diagramme and Notizen
- Auch während bereits laufender Konferenzen können Dokumente oder zusätzliche Teilnehmer hinzugefügt werden
- Chatfunktionalität

Sicherheitsfeatures

- System und Dokumente verbleiben hinter der Firewall der Organisation
- Eindeutige Konferenz-IDs und PINs
- Optionaler Passwortschutz
- Verschlüsselter Login am Host

Benutzerfreundlichkeit

- Konferenzen können via Telefon oder via Web-Interface gesteuert werden
- Dokumente, die in jeder beliebigen Desktop-Anwendung erstellt wurden, können ins System hochgeladen und dort eingesehen werden
- Vorgeplante Konferenzen und Ad-Hoc-Konferenzen



Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * **Community Watch and Participation**
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * **Access, Use and Reuse**
- * Transform

Systemvoraussetzungen:

- Mounting System: 2U 19"
- Betriebssystem: Windows 2003 server
- Prozessor: Dual 2.4GHz Xeon
- Arbeitsspeicher: 1GB RAM
- Festplattenspeicher: 73 GB Disk
- Redundanz: Dual disk, RAID 1 und redundante Stromzufuhr (optional)
- Netzwerk: Dual Gigabit Ethernet (10BASE-T, 100BASE-TX und 1000BASE-T, RJ45 output)
- Erweiterbarkeit: 3 PCI-X Erweiterungslots

Spec Sheet (Base System) (registrierungspflichtig):

http://sonexis.com/downloadable_materials/02-collateral/02-02%20-%20Spec%20Sheet%20-%20ConferenceManager%209.0%20%2803-02-09%29.pdf

Lizenzierungsmodell:

Ab 950 \$ pro Lizenz. Individuelle Preise a. A.

Dokumentation:

Platform Overview: http://omnipresence.com/PDFs/Sonexis_Data_Sheet.pdf

Weitere Informationen:

Evaluating Conferencing Solutions: A Buyer's Guide:

http://sonexis.com/downloadable_materials/white_papers/Conferencing_Buyers_Guide.pdf

**Empfehlung:**

Sonexis Conference Manager ist für Organisationen relevant, die ihre Telefonie-Infrastruktur um Audio- und Webconferencing-Funktionalitäten erweitern wollen. Darüber hinaus bietet Sonexis als Alleinstellungsmerkmal gegenüber anderen Kollaborationssystemen auch eine Echtzeit-Dokumentenbearbeitung und kollaboratives Web-Browsing, was die Lösung für gemeinsame wissenschaftliche Recherchen in Projektgruppen interessant macht. Andere Standardkomponenten der Projektsteuerung wie Gruppenkalender oder gemeinsame Adressbücher fehlen allerdings; die Anwendung konzentriert sich auf die direkt kommunikationsbezogenen Aspekte der Kollaboration.

Kap. 3: Workflow-Management-Systeme**Taverna**

Anwendung von automatisierten Analysemethoden auf Daten aus diversen Fachgebieten (z. B. Biologie, Chemie, Medizin, Meteorologie, Sozial- und Musikwissenschaft)

Zusammenführung von Daten aus unterschiedlichen Quellen mit automatisierter Formatkonvertierung

Anbieter:

University of Manchester, Teil des myGrid-Projekts, finanziert durch die Open-Source-Initiative OMII-UK

Referenzprojekte:

- Jiaotong Universität Shanghai: Paigos – automatisierte Zusammenstellung von Lerneinheiten nach benutzerspezifischen Kriterien auf der Basis einer multimedialen Bibliothek aus Kursmaterialien
- University of Bath: Einsatz von web-service-basierten Workflows für die Komposition synthetischer Musik
- University of Manchester: MIASGrid – Einsatz von web-service-basierten Information-Retrieval-Workflows auf der Basis medizinischen Bilddatenmaterials

**Funktionsumfang:**

Taverna Engine: Softwarekern, der die Geschäftslogik enthält

Taverna Workbench 2.1: Client, der dem Benutzer die Bedienungsoberfläche zur Verfügung stellt

- o Copy/paste, shortcuts, undo/redo, drag and drop
- o Animiertes Workflow-Diagramm
- o Datenexport nach CSV und MS Excel

Taverna Server 1.7 (Version 2.1 geplant für die 2. Jahreshälfte 2010)

- o Ermöglicht das Ausführen von Workflows auf entfernten Systemen
- o Die serverbasierten Workflows können entweder clientbasiert über ein Plugin in Taverna Workbench oder webbasiert über den Browser gesteuert werden

Funktionen bezogen auf Phasen des Data-Lifecycle:**Full Lifecycle Actions**

- * Description and Representation Information
- * **Preservation Planning**
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * **Conceptualise**
- * **Create or Receive**
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * **Access, Use and Reuse**
- * **Transform**

Systemvoraussetzungen:

- Microsoft Windows XP / Vista / 7, Mac OS X 10.4+, Linux
- 1 GB RAM
- Java 1.5+
- GraphViz dot

Lizenzierungsmodell:

GNU Lesser General Public License (GPL) 2.1

Dokumentation und weitere Informationen:

<http://www.taverna.org.uk/documentation/taverna-2-1>

Empfehlung:

Im wissenschaftlichen Tagesgeschäft erleichtert Taverna den Umgang mit der typischen Vielzahl von Online-Inhalten, die mit Hilfe des Tools automatisch und ohne mühsames manuelles Copy&Paste zu einer einheitlichen Datenbasis zusammengeführt werden können.



Somit vereinfacht die Anwendung den Kurationsschritt der Auswahl langfristig zu archivierender Daten.

Kepler

Anwendung zur Analyse und Modellierung wissenschaftlicher Daten

Die Software wurde auf der Basis des Ptolemy II-Frameworks der University of California entwickelt.

Anbieter:

Kepler/CORE team (UC Davis, UC Santa Barbara, UC San Diego), finanziert durch die NSF (National Science Foundation)

Referenzprojekte:

- **UC Davis:** pPOD – eine Extension zur Kepler-Software, die es Biologen ermöglicht, stammesgeschichtliche Analysen zu automatisieren, aufzuzeichnen und zu einem späteren Zeitpunkt zu rekonstruieren
- **NSF Cyberinfrastructure for Environmental Observatories:** COast-to-Mountain Environmental Transect Project (COMET) – die Kepler-Software wird eingesetzt, um Umwelt- und Klimafaktoren zu analysieren sowie um Wechselwirkungen zwischen Kohlendioxidausstoß und Meeresauftrieb zwischen der Küstenregion Kaliforniens und der Sierra Nevada zu untersuchen
- **Scientific Data Management Center des US Department of Energy:** Mit Hilfe der Kepler-Software werden Lösungen und Anwendungen entwickelt, um wissenschaftliche Workflows auf einfache und effiziente Art und Weise modellieren, ausführen und wiederverwenden zu können

Funktionsumfang:

- Run-Time-Engine, die eine Benutzung der Software wahlweise über eine grafische Benutzeroberfläche oder über die Kommandozeile ermöglicht
- Kepler-Workflows können modular aufgebaut und verschachtelt werden, so dass sich Sub-Workflows bilden lassen, die für viele verschiedene Anwendungen wiederverwendet werden können
- Kepler unterstützt Standards der Grid-Technologie (z. B. Globus, SRB, Web und Soaplab Services) und den Einsatz von Parallelcomputern
- Dateiaustausch möglich über das Kepler Archive Format (KAR)



- Über 350 mitgelieferte, anpassbare Workflows, z. B. für R und Matlab (statistische Analysen), für die Integration von WSDL-Webservices, für den Datenimport aus Excel-Sheets und zur Ausführung von Kommandozeilen-Anwendungen innerhalb eines Workflows
- Serverbasiertes Component Repository für das zentrale Vorhalten von Workflows
- Unterstützte Protokolle und Standards: Ecological Metadata Language (EML), DiGIR, OPeNDAP, GridFTP, JDBC, SRB

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * **Conceptualise**
- * **Create or Receive**
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * **Access, Use and Reuse**
- * **Transform**

Systemvoraussetzungen:

- 300 MB Festplattenplatz 512 MB of RAM Minimum, 1 GB oder mehr empfohlen
- 2 GHz CPU Minimum
- Java 1.5.x (nicht kompatibel mit Java 1.6)
- Netzwerkanbindung (optional)
- R Software (optional)

Lizenzierungsmodell:

Berkeley Software Distribution License (BSD)

Dokumentation und weitere Informationen:

<https://kepler-project.org/users/documentation>

Empfehlung:

Kepler ist spezialisiert auf die naturwissenschaftliche Forschung. Dank der Unterstützung der wichtigsten Standardschnittstellen dieses Bereichs ist es ein zur Datenanalyse in naturwissenschaftlichen Disziplinen besonders geeignetes Tool.



Pegasus

Skalierbares Workflow-Management-System, das sich als Middleware in zahlreiche Forschungsanwendungen (z. B. aus den Bereichen Astronomie, Biologie, Erdbebenforschung und Gravitationsphysik) integrieren lässt

Ziel des Projekts: Die Entwicklung benutzerfreundlicher Tools für das Ausführen, Überwachen und Debuggen wissenschaftlicher Workflows

Anbieter:

University of Southern California und University of Wisconsin Madison

Stimmen zum System und Anwendungsbeispiele:

„Pegasus automatically chains dependent tasks together, so that a single scientist can complete complex computations that once required many different people“

Ewa Deelman vom Information Sciences Institute der University of Southern California, siehe <http://www.isgtw.org/?pid=1000664>

Mapping Workflows on Grid Resources: Experiments with the Montage Workflow, siehe <http://pegasus.isi.edu/publications/2009/coregrid09.pdf>

Adaptive Workflow Processing and Execution in Pegasus, siehe <http://pegasus.isi.edu/publications/2008/lee-pegasus-wage.pdf>

Funktionsumfang:

- **Pegasus Mapper:** Entscheidungssystem, mit dem Strategien zur Abbildung von Workflow-Beschreibungen auf verteilte Infrastrukturen (z. B. TeraGrid, Open Science Grid) entwickelt. So können Workflows abstrakt und ohne Abhängigkeit von der zugrundeliegenden e-Infrastruktur konstruiert werden.
- **Condor Schedd:** Batch-System für rechenintensive Operationen
- **DAGMan (Directed Acyclic Graph Manager):** Anwendung zur Koordination von Workflow-Schritten, deren Input, Output oder Ausführung von mehreren unterschiedlichen Programmen abhängt



Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * **Community Watch and Participation**
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * **Create or Receive**
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * **Access, Use and Reuse**
- * **Transform**

Lizenzierungsmodell, Systemvoraussetzungen: N. N.

Dokumentation und weitere Informationen:

<http://pegasus.isi.edu/wms/docs/quickch1/quickch1.html>

Empfehlung:

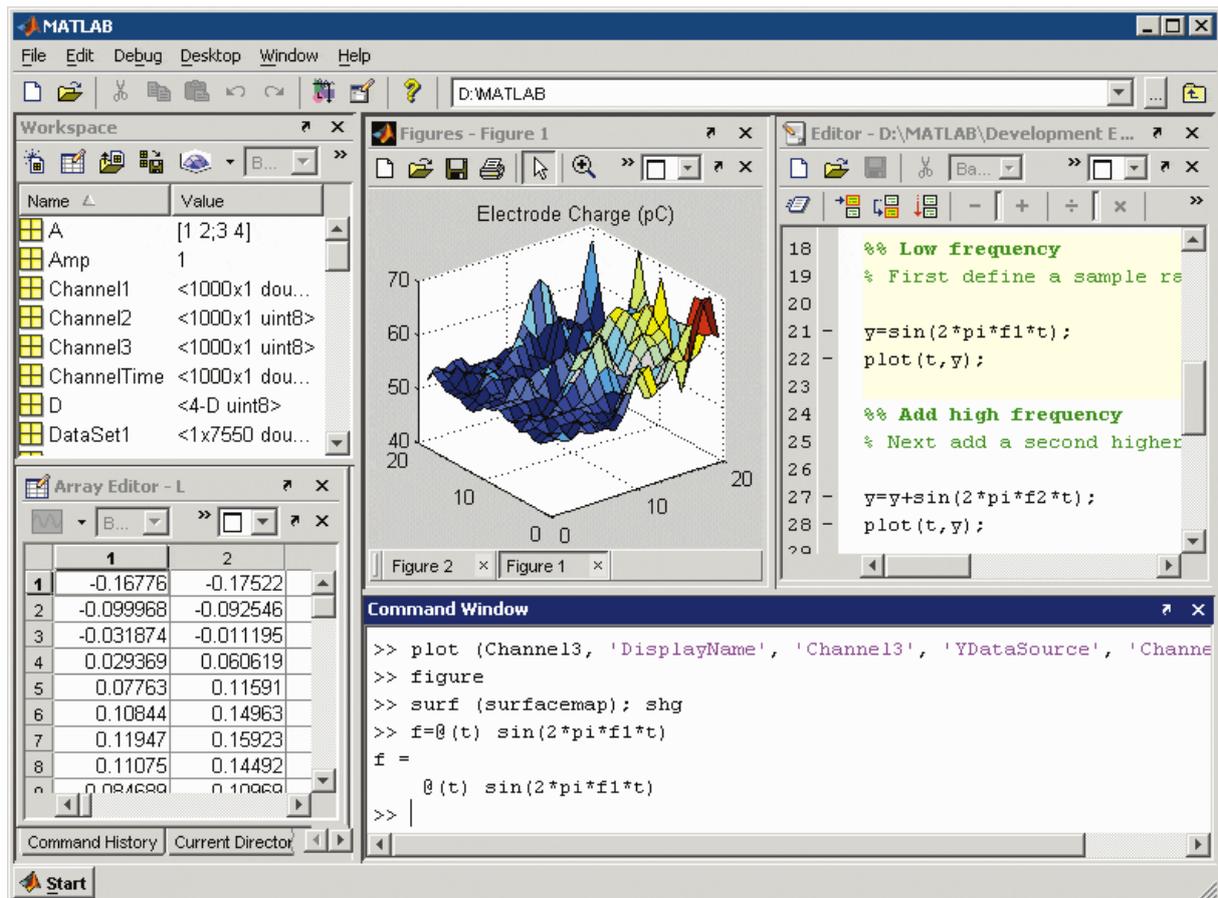
Pegasus richtet sich an Software-Entwickler und Integriatoren, die ein leistungsstarkes und dank umfangreicher Grid-Unterstützung hochskalierbares Workflow-Management-System als Middleware in eine bestehende e-Infrastruktur einbinden wollen.

Kap. 4: Datenvisualisierung

Matlab

MATLAB ist eine hochentwickelte Programmiersprache und Software für die Visualisierung und Analyse von Daten mittels mathematischer Modelle.

Neben den Basisfunktionen ist es möglich, für die Lösung bestimmter Problemklassen Toolboxes mit speziellen MATLAB-Funktionen zu erwerben.



Quelle: http://www.mathworks.com/products/matlab/images/matlabdesktop_lg.jpg

Anbieter:

The MathWorks GmbH
Adalperostraße 45
85737 Ismaning

Referenzprojekte:

Max-Planck-Institut

Problem: Erzeugung von hochwertigen 3D-Bildern von Proteinen

Lösung: Einsatz von MATLAB und MathWorks-Tools, um Daten von digitalen Aufnahmen eines Elektronenmikroskops zu sammeln, zu analysieren, zu filtern, zusammenzuführen und als Modell darzustellen.

Verbesserung dank der MATLAB-Software: Zeitersparnis bei Forschung (um Jahre), Entwicklung (Tage statt Wochen) und Workflow.

Quelle:

<http://www.mathworks.de/products/matlab/userstories.html?file=45628&title=The%20Max%20Planck%20Institute%20Reconstructs%20Key%20Protein%20Complexes%20Using%20MATLAB%20and%20Parallel%20Computing%20Toolbox%E2%84%A2>



MIT

Problem: Verbesserung der Diagnosefähigkeit bei Krebserkrankungen durch die Identifikation von Proteinen und die Analyse der Interaktion dieser Proteine.

Lösung: Einsatz von MATLAB und MathWorks-Tools, um Massenspektrometrie-Daten zu analysieren sowie um Modelle von Protein-Interaktionen zu erstellen und zu visualisieren.

Verbesserungen dank der MATLAB-Software:

- Einbeziehung von Studenten in die Forschung
- Kürzere Rechenzeit
- Forschungspreis

Quelle:

<http://www.mathworks.de/products/matlab/userstories.html?file=45694&title=Massachusetts%20Institute%20of%20Technology%20Integrates%20Cancer%20Research%20in%20the%20Lab%20and%20Classroom%20with%20MathWorks%20Tools>

Funktionsumfang:

- Hochentwickelte Programmiersprache für wissenschaftlich-technische Berechnungen
- Entwicklungsumgebung zur Verwaltung von Code, Dateien und Daten
- Interaktive Werkzeuge für iterative Untersuchungen, Entwürfe und die Lösung von Problemen
- Mathematische Funktionen für die lineare Algebra, die Statistik, die Fourieranalyse, das Filtern, Optimieren und die numerische Integration
- 2D- und 3D-Grafikfunktionen zur Visualisierung von Daten
- Tools zur Erstellung eigener grafischer Benutzeroberflächen
- Funktionen zur Integration auf MATLAB basierender Algorithmen in externe Anwendungen und Sprachen wie C/C++, Fortran, Java, COM und Microsoft Excel

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform



Systemvoraussetzungen:

Windows

Operating Systems	Processors	Disk Space	RAM
32-Bit and 64-Bit MathWorks Products			
Windows XP Service Pack 3	Any Intel or AMD x86 processor supporting SSE2 instruction set*	1 GB for MATLAB only, 3-4 GB for a typical installation	1024 MB (At least 2048 MB recommended)
Windows XP x64 Edition Service Pack 2			
Windows Server 2003 R2 Service Pack 2			
Windows Vista Service Pack 1 or 2			
Windows Server 2008 Service Pack 2 or R2			
Windows 7			

Macintosh

Operating Systems	Processors	Disk Space	RAM
32-Bit and 64-Bit MathWorks Products			
Mac OS X 10.5.5 (Leopard) and above	All Intel-based Macs*	1 GB for MATLAB only, 3-4 GB for a typical installation	1024 MB (At least 2048 MB recommended)
Mac OS X 10.6.x (Snow Leopard)			

* Intel Core 2 or later required to run 64-bit MATLAB.

Linux

Operating Systems	Processors	Disk Space	RAM
32-Bit and 64-Bit MathWorks Products			
Qualified distributions*: Ubuntu 8.04, 8.10, 9.04, and 9.10	Any Intel or AMD x86 processor supporting SSE2 instruction set**	1 GB for MATLAB only, 3-4 GB for a typical installation	1024 MB (At least 2048 MB recommended)
Red Hat Enterprise Linux 5.x			
SUSE Linux Enterprise Desktop 11.x			
Debian 5.x			

* The listed distributions are those Linux distributions that MathWorks products have been validated against. It is likely that other distributions with kernel version 2.6 or later and glibc version 2.5 or later can successfully run MathWorks products, but The MathWorks will be in a limited position to provide technical support for those distributions.

** You can find more information in [Solution 1-B3MR75](#).

Quelle: http://www.mathworks.com/support/sysreq/current_release/index.html

Lizenzierungsmodell:

Basis-Software: 500 €(Individuallizenz), 800 €(Concurrent Licence)

Je Spezialmodul: 200 €(Individuallizenz), 325 €(Concurrent Licence)

Dokumentation:

<http://www.mathworks.com/access/helpdesk/help/techdoc/>

Weitere Informationen:

Beispiele: http://www.mathworks.com/access/helpdesk/help/techdoc/demo_example.html

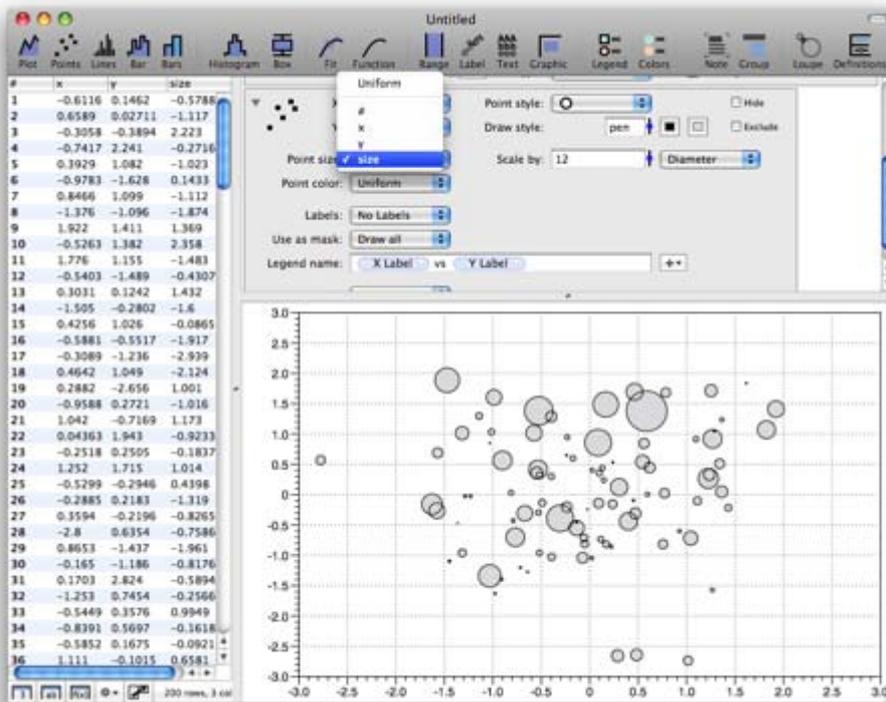


Empfehlung:

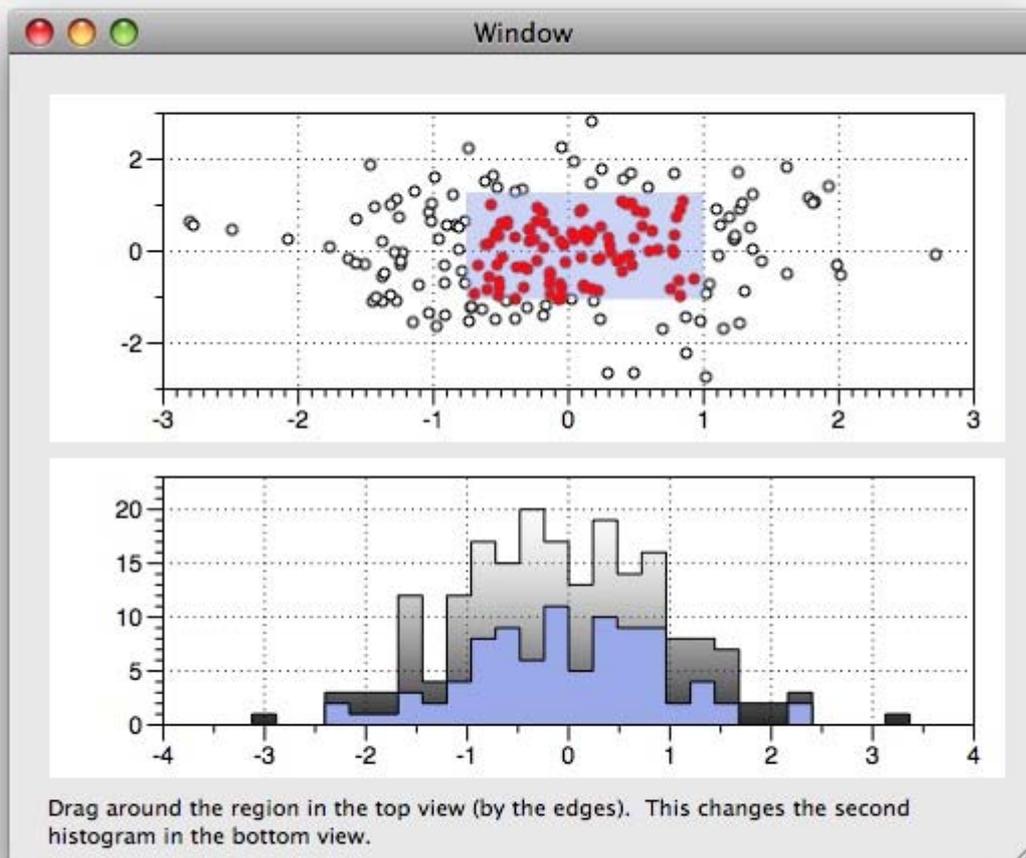
Ist es möglich, seine Daten so sehr zu abstrahieren, dass ein mathematisches Modell greift, bietet sich der Einsatz von MATLAB an. MATLAB ist eine hoch entwickelte Programmiersprache und Software für die Visualisierung und Analyse von Daten. ZU bestimmten Problemklassen existieren bereits Toolboxes, die zusätzlich erworben werden können.

DataGraph

DataGraph ist eine dynamische Statistiksoftware für Mac-Computer zum Erstellen von Diagrammen: Während man die Daten eingibt, wird das Diagramm dargestellt. DataGraph ist gut mit Excel und anderen Statistikpaketen kombinierbar.



Quelle: <http://www.visualdatatools.com/DataGraph/ScreenShots/index.html>



Quelle: <http://www.visualdatatools.com/DataGraph/Framework/index.html>

Anbieter:

Visual Data Tools, Inc.

Funktionsumfang:

Analyse:

- Zoomfunktion
- einfache Überprüfung von Modellhypothesen
- Variation von Parametern durch Slider

Darstellung:

- Charts, Boxplots, Histogramme, Ausgleichsrechnung
- Diagramm mittels Labels kommentier- und beschriftbar
- Original- und Modelldaten in einem Diagramm darstellbar
- Individuelle Darstellung der Diagramme durch command line Befehle

Exportformate:

pdf, eps, svg, tiff, jpg, png



Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Mac OS X 10.4.11 (oder höher)

Lizenzierungsmodell:

400 \$ (Netzwerklicenz)

Dokumentation:

<http://www.visualdatatools.com/phpBB2/faq.php>

Weitere Informationen:

<http://www.visualdatatools.com/phpBB2/>

Empfehlung:

Für Mac-User, die oft Statistiken erstellen, bietet DataGraph eine gute Lösung. Durch seine Interaktivität – während man die Daten noch eingibt, wird das Diagramm bereits grob erstellt – und die Möglichkeit, Parameter mit einem Slider zu variieren, ermöglicht DataGraph ein effizientes Arbeiten. DataGraph ist gut mit Excel und anderen Statistikpaketen kombinierbar.

Cartographica

Cartographica ist ein GIS (Geographisches Informationssystem) für Macintosh-Computer. Mit Cartographica können aus räumlichen Geodaten wissenschaftliche Landkarten und Analysen erstellt werden. Bei der Entwicklung der Software stand das möglichst unkomplizierte Importieren, Extrapolieren, Analysieren und Transformieren der Daten in visuelle Modelle im Vordergrund.

**Anbieter:**

Clustrust
11654 Plaza America Drive
Suite 214
Reston, VA 20190

Kooperationspartner:

- Geospatial Data Abstraction Libraries (GDAL)—X/MIT license
- Xerces XML Parser—Apache License
- LizardTech MrSID® SDK—Read-only license
- Growl notification system—BSD License
- Sparkle update system—MIT License

Referenzprojekte:

Geographica gibt es auch für Iphone und das Ipad:
<http://www.clustrust.com/Press/2010-0415-Cartographica-Mobile-1.0.pdf>

Funktionsumfang.

Datenimport/Datenexport:

- unterstützte Datenimport-Formate: GeoTIFF, DEM, MrSID, etc.
- unterstützte Vektorgrafikformate-Formate: ArcInfo, Shape Files, pre-2007 TIGER/Line files, etc.
- unterstützt Import und Export von ESRI® Shape-Dateien
- unterstützt Import von Daten aus Datenbanken mit ODBC Standard
- unterstützt Datenimport mittels pen tablet und puck devices (z.B. von Wacom)
- Datenimport von Internetseiten
- intelligenter Datenimport mit Ausfiltern von doppelten Koordinatenangaben

Drucker:

- unterstützt das Ausdrucken mittels großformatiger Drucker

Datenanalysefeatures:

- Kernel Density Maps
- Point in Polygon

Ergonomie:

- Ausgearbeitete Undo- sowie Drag-and-Drop-Funktionalität

Filtering:

- multi-variable searching
- precise filtering

Layout:



- Editierfunktionen für die Diagramme

Direct Editing:

- Editiermöglichkeit der geometrischen Datenmodelle mittels control points

Layer Transparency:

- Visuelle Hervorhebung unterschiedlicher Datenebenen

Location-Based-Services durch GPS-Unterstützung

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * **Description and Representation Information**
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * **Create or Receive**
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * **Access, Use and Reuse**
- * **Transform**

Systemvoraussetzungen.

Minimal:

Betriebssystem: MacOS X 10.5.6

CPU: Intel or PowerPC

RAM: 512MB

Festplatte: 1GB

empfohlen:

Betriebssystem: MacOS X 10.5.6

CPU: Intel Core2 Duo

RAM: 2GB

Festplatte: 1GB

Lizenzierungsmodell:

Kauflizenz 395 \$ (Limitierter Aktionspreis), regulär 495 \$

Dokumentation:

<http://www.macgis.com/help/index.html>

Weitere Informationen:

<http://www.cluetrust.com/kb/categories/About+Cartographica>



<http://www.macgis.com/quicktours.php>

Empfehlung:

Cartographica ist ein Geographisches Informationssystem (GIS) für Macintosh User. Mittels Cartographica können wissenschaftliche räumlichen Geodatenmodelle erstellt und analysiert werden. Dank der Unterstützung großformatiger Ausdruckoptionen können Plakate erstellt werden.

GSI 3D

Mittels des Softwarepakets GSI 3D können Oberflächen dreidimensional modelliert und visualisiert werden. Die so erhaltenen Modelle können mit anderen geographischen Informationssystemen (GIS) ausgetauscht und weiterbearbeitet werden. GSI 3D kann beispielsweise dazu in den Bereichen Wassermanagement, Archäologie und Stadtplanung zur Analyse von Überflutungswahrscheinlichkeiten und Umweltschäden, zur Einschätzung von Bodenqualitäten sowie zur Evaluierung von Bodenbedingungen für Müllablagerplätze und zur Analyse von Bodeneingesetzt werden.

Voraussetzung zur Nutzung der GSI 3D-Software ist die Mitgliedschaft in einem Konsortium. Die Mitglieder erhalten Softwareupdates. Die Software wird entsprechend der Bedürfnisse der Mitglieder weiterentwickelt.

Daten können in Form von digitalen Landkartenmodellen in das System eingespeist werden oder aus Kartenmaterial, Luft- und Satellitenbildern digitalisiert werden.

Anbieter:

British Geological Survey
Keyworth
Nottingham
NG12 5GG

Referenzprojekte:**Universität Köln**

Geologische Schnitte zur Abbildung archäologischer Schichten
http://www.geosum3d.de/cms/Poster_Deuqua_2004.pdf

**Funktionsumfang:**

Es kann auf Datenmodelle der British Geological Survey (BGS) zurückgegriffen werden:

- Scale independent (no maximum project area)
- Unlimited* structural (i.e. geological) units
- Unlimited* cross sections
- Local grid coordinate system
- User-defined stratigraphy and legend colours
- Non-faulted 3D volumes
- Medium structural complexity, including non-overtured folds
- Subcrops
- Onlap/overlap and offlap
- Lenses within units

*Subject to available computer memory

Folgende Standard-Import-/Export-Formate werden unterstützt:

- Open project file format (XML)
- ASCII grids
- GOCAD TIN .ts files
- ESRI Shapefiles
- Geokodierte JPEG-Rasterkarten und vertikale Schnitte
- Spreadsheet borehole import (tab-delimited)
- Point data (tab-delimited)

Quelle: <http://gsi3d.org/technicalData.html>

**Funktionen bezogen auf Phasen des Data-Lifecycle:****Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

- Betriebssystem: Windows (mindestens XP)
- Andere Betriebssysteme: eventuell (Kontaktaufnahme mit Hersteller)
- RAM: mindestens 1GB
- Festplatte: ca. 200 MB
- Grafikkarte: NVIDIA oder ATI mit OpenGL (1.3 oder höher)
- NICHT UNTERSTÜTZT werden: Thin-Client-Architekturen, Web- und Netzwerkinstallation

Lizenzierungsmodell:

Academic teaching	
One-off joining fee	Annual membership fee
€2,000	Unlimited users €1,000
<i>Licence conditions</i>	
This level of membership allows for academic teaching purposes only. It does not include teaching as part of a commercial training course. The annual membership fee is not based on the number of users, and should therefore be an attractive option for universities and colleges who will have the freedom to deploy the software across their teaching programmes.	
<i>Example</i>	
<ul style="list-style-type: none"> • A university wishes to equip its earth and environmental science departments with GSI3D to support a series of teaching modules in which the students would also use the software in practical classes. In the first year they would pay €3,000 (including the €2,000 joining fee), and thereafter would pay an annual membership fee of €1,000. 	

Quelle: <http://gsi3d.org/pricelist.html>

Dokumentation:

http://www.gsi3d.org.uk/downloads/GSI3D_manual_V2_6_3_WebVersion.pdf

**Weitere Informationen:**

http://nora.nerc.ac.uk/9444/1/GSI-3D_corporate_flyer_finalversion.pdf

Empfehlung:

GSI 3D ist ein geographisches Informationssystem (GIS), dessen Einsatz die Mitgliedschaft in einem Konsortium voraussetzt. Mittels des Softwarepakets können Oberflächen dreidimensional modelliert und visualisiert werden und so die Bodenbedingungen in vielfältigen Kontexten wie Archäologie, Stadtplanung oder Umweltschutz analysiert werden. Dank der Möglichkeit, Kartenmaterial, Luft- und Satellitenbildern zu digitalisieren, ist GSI 3D Effizienz steigernd einsetzbar.

InstantAtlas

Tool zur web-basierten Visualisierung ortsbasierter statistischer Daten

Darstellung von Schlüsselindikatoren und Performance-Daten auf lokaler, regionaler und nationaler Ebene

Nutzer: Wissenschaftler in der Forschung, Analysten und Entwickler von Web-Applikationen
Einsatzzweck: Verbesserung von Geschäftsprozessen und Qualitätsoptimierung interner und externer Berichterstattung

Anbieter:

GeoWise Ltd, a subsidiary of ESRI Holdings Ltd.

Referenzprojekte:

Institut für Krebspidemiologie, Lübeck: Darstellung von Erkrankungs- und Mortalitätsraten in Schleswig-Holstein auf Landkreisebene (<http://www.krebsregister-sh.de/atlas/Inzidenz/atlas.html>)

Statistisches Amt für Hamburg und Schleswig-Holstein: Strukturentwicklung in den Kreisen und kreisfreien Städten (<http://www.statistik-nord.de/fileadmin/download/kreismonitor/atlas.html>)

**Funktionsumfang:****InstantAtlas Desktop:** Erstellen von Reports mit Auswahl von Geografien

- o Publizieren (Publisher, Karten, Vorlagen)
- o Design (StyleEditor, Designer)
- o Daten hinzufügen (Daten Manager)

InstantAtlas Server:

- o Webbasierte, datenbankgestützte Applikationslösung für statistisches Datenmanagement
- o Erzeugung von Reports aus (Teil-)Selektionen der Datenbasis in Echtzeit
- o Kollaborative Nutzung ortsbezogener statistischer Daten Intelligence-Systeme, Datenbeobachtungs-Features und Bereichsprofil-Anwendungen
- o Live-Demo: http://213.171.206.130/IAS_Demo/

Funktionen bezogen auf Phasen des Data-Lifecycle:**Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Server:

- Microsoft Windows Server 2003 / Microsoft Internet Information Server (v6)

Datenbank:

- Microsoft SQL Server 2005
- Microsoft SQL Server 2000
- Oracle 10g

Client:

- PC mit Pentium 3 Prozessor (oder entsprechend) und 512 MB RAM
- Microsoft Excel 2003+
- Ein Kartenpaket oder Ihre eigenen digitalen Kartendaten im SHP- TAB- oder MID-/MIF- Format
- Browser: Kompatibel mit Adobe Flash Player 9+

Lizenzierungsmodell:

Endnutzerlizenzen lt. Preisliste

http://www.instantatlas.com/downloads/PriceList%28EURO%29_de.pdf



Dokumentation und weitere Informationen: <http://www.instantatlas.com/de/index.xhtml>

Empfehlung:

Wenn es um die web-basierte Publikation von auf geographische Kontexte bezogenen Forschungsdaten geht, ist Instant-Atlas die Lösung der Wahl. Die Visualisierungen werden in flexibel benutzbarer und zeitgemäß gestalteter Form ausgegeben und lassen sich gut in bestehende Web-Infrastrukturen integrieren.

Paraview

ParaView ist eine Open-Source-Software zur Analyse und Visualisierung von Daten. ParaView stellt qualitative und quantitative Techniken zur Analyse der Daten zur Verfügung. ParaView wurde entwickelt, um große Datenmengen aus verschiedenen (räumlich verteilten) Datenquellen aufzubereiten. ParaView läuft sowohl auf Supercomputern (große Datenmengen) als auch auf Laptops (kleine Datenmengen).

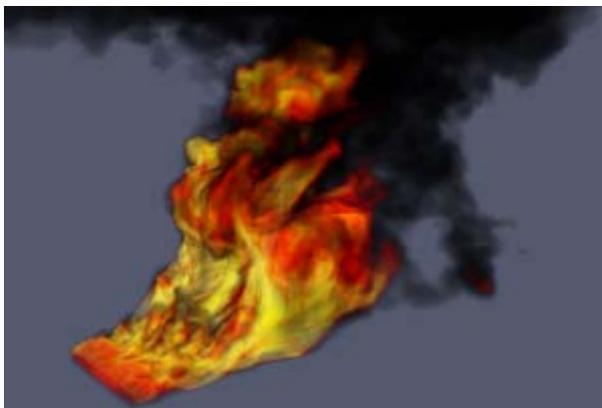
Anbieter:

Kitware Inc.
28 Corporate Drive
Clifton Park, NY 12065
USA

Referenzprojekte:

Feuer unter Windeinfluss

Simulation eines Feuers unter Windeinfluss mit Berücksichtigung der Temperaturverteilung mittels 150 Millionen Freiheitsgraden. Berechnet mittels Supercomputer der Sandia National Laboratories.

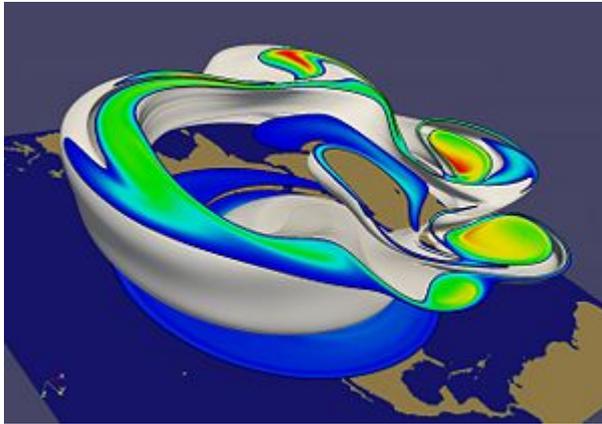


Quelle: <http://www.paraview.org/paraview/resources/applications.html>



Strömungsmodell

Strömungsmodell, das den Abriss eines Luftstroms darstellt mittels dessen Ozongehalt und Wetter vorhergesagt werden können.



Quelle: <http://www.paraview.org/paraview/resources/applications.html>

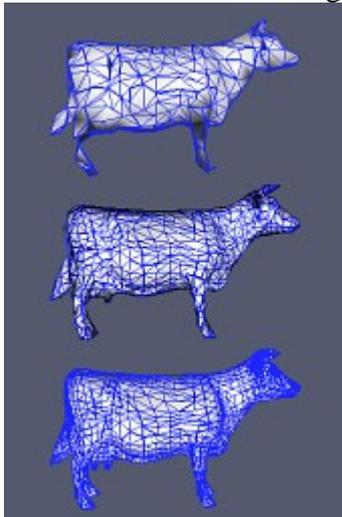
Funktionsumfang:

Unterstützte Datenformate:

- VTK, EnSight 6, EnSight Gold, Plot3D, STL and BYU

Visualisierung:

- verschiedene Modelle von Koordinatensystemen stehen zur Verfügung
- interaktives thresholding (User bestimmt Auflösung des Modells, siehe Bild)



Datentypen:

- uniform rectilinear, non-uniform rectilinear und curvilinear
- polygonal, image, multi-block, AMR
- unstrukturierte Datentypen

**Daten**

- Jede Operation erzeugt Datensets, die ebenfalls zur Weiterverarbeitung benutzt werden können.
- Teilmengen von Datensets (wie beispielsweise die Oberflächenstruktur) können isoliert werden und ebenfalls zur Weiterverarbeitung verwendet werden.

Analyse

- interaktive Parametervariation
- Erzeugung von Streamlins innerhalb der Modelle
- Berechnung von neuen Variablen
- Generierung von statistischen Daten

Berechnung

- Möglichkeit, berechnungen mittels der Programmiersprache Python durchzuführen.
- Berechnung in Form einer verteilten Anwendung möglich.

Funktionen bezogen auf Phasen des Data-Lifecycle:**Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Windows (XP/Vista/7)

Linux

Software erhältlich in 32 and 64 Bit.

Lizenzierungsmodell:

Open Source

Dokumentation:

<http://paraview.org/OnlineHelpCurrent/>

Weitere Informationen:

Kitware Public Wiki: <http://paraview.org/Wiki/ParaView:FAQ>

**Empfehlung:**

ParaView bietet sich insbesondere dann an, wenn man ein komplexes (dynamisches) Objekt modellieren, analysieren oder darstellen möchte. ParaView wurde entwickelt, um große Datenmengen aus verschiedenen (räumlich verteilten) Datenquellen aufzubereiten. ParaView läuft sowohl auf Supercomputern (große Datenmengen) als auch auf Laptops (kleine Datenmengen).

Rockworks 15

Rockworks ist eine Software zur Visualisierung und Analyse von Subsurface-Daten und ist seit Langem Standard in den Disziplinen Umwelt und Geotechnologie sowie der Rohstoffbranche.

Stimme zum System:

“As a student I found the RockWorks program to be very easy to learn. Using one of the sample datasets I found that I could simply edit the sample to match my data and I was done. The whole process took about 10 minutes and the printout looks much better than anything I could draw- thanks!”

Mike Fidler, student at Ohio State University

Anbieter:

RockWare Europe
Vicolo dei Saroli 1
Cureglia CH-6944

Funktionsumfang:

Viele Import/Export-Formate

Datentypen:

lithology, stratigraphy, geophysical, geochemical, fracture und water level data.

Darstellung

- 2D und 3D
- Log-Sections, Cross-Sections (2D, 3D)

Modelle:

- geochemistry, geophysical, geotechnical
- point, contour, isopach, 3D surface, land grid, survey maps
- continuous models



Bedienung der Software:

- Auch mittels RockWare Command Language (command line-Befehle)
- Interaktive Ansichten
- Detaillierte Reports

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

- Windows XP, Vista oder Windows 7
- 1 GB RAM (2+ empfohlen)
- CPU: 1,4 GHz oder mehr
- Bildschirmauflösung höher als 800x600 Pixel
- RockWorks 15 ist ein 32-Bit-Programm, das auf 64-Bit-Windows läuft

Lizenzierungsmodell:

Universitäten: 2.499 US-\$ pro Jahr

Weitere Preise siehe <http://www.rockware.com/product/pricing.php?id=165>

Dokumentation:

http://www.rockware.com/assets/products/165/downloads/documentation/41/rw15_manual.pdf

Weitere Informationen:

<http://www.rockware.com/product/faq.php?id=165>

Empfehlung:

Rockworks ist eine Software zur Visualisierung und Analyse von Subsurface-Daten und ist seit Langem Standard in den Disziplinen Umwelt und Geotechnologie sowie der Rohstoffbranche.



"A classic in this category is the Visual Thesaurus... Related words are linked with "virtual springs" that pull them together into a cluster. To navigate through the Visual Thesaurus, a user types in a word which pulls up a tree-like shape with the word in question at the centre, and related terms clustered around it... Definitions appear when the cursor is moved over dots next to the words. To find out more about a synonym, a user can click on it, which makes it move to the centre of a new cluster of words."

The Economist

Quelle: <http://www.visualthesaurus.com/praise/>

Anbieter:

Thinkmap, Inc.
New York, NY

Referenzprojekte:

http://www.thinkmap.com/casestudies_noaa.jsp

Funktionsumfang:

Treffende englischsprachliche Wörter finden. (Wortmaterial: 145.000 englischsprachige Wörter und 115.000 Wortbedeutungen)

Möglichkeit, bei unbekanntem englischsprachigen Wörtern die Bedeutung mittels bedeutungsverwandter Wörter zu erschließen

Verbesserung der verbalen Fähigkeiten durch das Lesen von Mustersätzen aus denen sich die Bedeutung und Verwendung unbekannter Wörter erschließt

Wahl zwischen 2D- und 3D-Darstellung der Wortfelder.

Die Wörter können auch akustisch ausgegeben werden.

Internetanbindung für weitere Wortverwendungsmöglichkeiten.

Funktionen bezogen auf Phasen des Data-Lifecycle:



Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:



OS: 2000, XP, Vista, Windows 7
Memory: 128MB
Browser: Internet Explorer 7+, Firefox
Java: Virtual Machine 1.4+



OS: 2000, XP, Vista, Windows 7
Memory: 128MB
Disk Space: 100MB



OS: OSX 10+
Memory: 128MB
Browser: Safari, Firefox
Java: Virtual Machine 1.4+



OS: OSX 10+
Memory: 128MB
Disk Space: 100MB

Lizenzierungsmodell:

Desktop Edition:

Lizenzen:

- 1-4: \$39.95
- 5-9: \$26.50
- 10-99: \$24.00
- 99+: a. A.

Online Edition:

	BEST VALUE 	
Individual Monthly	Individual Yearly	Group
A great low-cost way to get started. Billed monthly to your credit card.	Billed annually to your credit card, a 44% savings from the monthly rate.	Each member gets their own personal account, but all billing is done through one credit card.
\$2.95 per month	Save 44% \$19.95 per year	Starting at \$29.95 per year

Quelle: <https://www.visualthesaurus.com/store/>

Dokumentation:

<http://www.visualthesaurus.com/howitworks/manual/>



Weitere Informationen:

<http://www.visualthesaurus.com/howitworks/faq/>

<http://www.visualthesaurus.com/howitworks/tips/>

**Empfehlung:**

Dank des großen Wortmaterials von 145.000 englischsprachigen Einträgen und der intuitiv verstehbaren Baummetapher bei der visuellen Umsetzung eignet sich Visual Thesaurus für Anglisten und Amerikanisten und vergleichende Sprachwissenschaftler für semantische Analysen sowie für Lerner der englischen Sprache zur aktiven kognitiven Erschließung ausgewählter sprachlicher Bereiche.

Kap. 5: Speichersysteme**LOCKSS (Lots of Copies Keep Stuff Safe)**

LOCKSS ist eine Open Source-Datenhaltungslösung auf Basis einer dezentralen peer-to-peer-Struktur

LOCKSS basiert auf dem klassischen Bibliotheksmodell

Rund 400 Bibliotheken und Verlage haben sich in einer weltweiten Non-Profit-Allianz zusammengeschlossen, die sich zum Ziel gesetzt hat, elektronischen Content langfristig zu archivieren. Die angeschlossenen Verlage stellen nicht mehr lieferbare Titel auf einem zentralen Server auf Basis der LOCKSS-Software zur kostenfreien Nutzung per Internet zur Verfügung.

Das System wird evaluiert durch teilnehmende Bibliotheken, den Council of Library Resources und die Library of Congress

Stimmen zum System:

“NYU Libraries’ decision to join the LOCKSS Alliance was grounded in our sincere commitment to the preservation of digital content. We want to support the broadest possible exploration of methods for insuring such preservation. The LOCKSS model is based on an elegantly simple philosophy, promoting the building of local collections and their preservation on site. We are impressed by the autonomy that the LOCKSS model offers and the minimal effort required to run it. NYU has been very happy with its decision to join LOCKSS, and we encourage others to express their commitment to digital preservation by joining the LOCKSS community.”

Dr. Michael Stoller, Director, Collections & Research Services, New York University Libraries

Having multiple copies of our content on a geographically dispersed network gives us more confidence that it will still be around in five years, ten years, or - given weather conditions down here - next week.



Aaron Trehub, Director of Library Technology, Auburn University, Alabama

Anbieter:

LOCKSS Program
Stanford University Libraries

Projektpartner:

Projektpartner sind die Humboldt-Universität zu Berlin (Projektleitung) und die Deutsche Nationalbibliothek. Im deutschen LOCKSS-Netzwerk werden zudem die Universität Regensburg, die Bayerische Staatsbibliothek, die Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, das Forschungszentrum Jülich, die Niedersächsische Staats- und Universitätsbibliothek Göttingen und das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen teilnehmen. Das Projekt wird außerdem aktiv unterstützt von der Library of Congress und der Stanford University.

(vgl. <http://www.d-nb.de/wir/projekte/lukii.htm>)

Referenzprojekte:

LuKII-Projekt (Gemeinschaftsprojekt der Humboldt-Universität Berlin und der Deutschen Nationalbibliothek finanziert durch die DFG). Ziele:

1. Etablierung eines kostengünstigen LOCKSS-Netzwerks in Deutschland
2. Konzeption und Implementierung der Interoperabilität von LOCKSS und kopal
3. Testen der Interoperabilität durch Archivierung von deutschen Institutsrepositorien

Weitere Informationen:

<http://www.ibi.hu-berlin.de/forschung/digibib/forschung/projekte/LuKII>

Funktionsumfang:

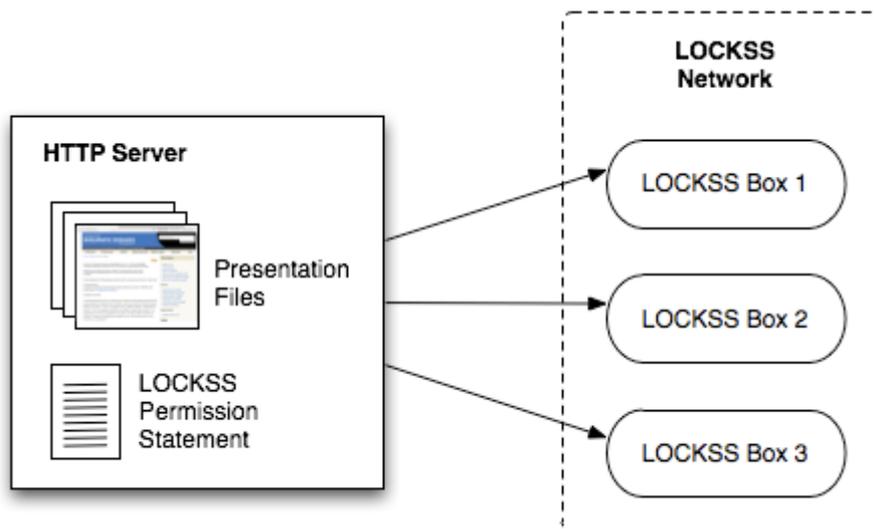
Mit LOCKSS wird ein niedrigpreisiger Einzelplatz-PC in einer Bibliothek zu einer digitalen Datenhaltungseinheit („LOCKSS-Box“), die die folgenden Funktionen erfüllt:

- Sammeln von Web-Inhalten mit Hilfe suchmaschinenähnlicher Crawler.
- Kontinuierlicher Abgleich der Inhalte zwischen den verschiedenen LOCKSS-Boxen, Bereinigung der Unterschiede.
- Die LOCKSS-Box agiert als Proxy oder Cache und ermöglicht den Bibliotheksnutzern Zugriff auf den gespeicherten Content. Das Layout der in LOCKSS gespeicherten Inhalte sieht jederzeit so aus wie das des Anbieters. Die Inhalte stehen auch dann noch



zur Verfügung, wenn die Webseite des Anbieters ausfällt. Liefert eine von LOCKSS an den Anbieter weitergeleitete Anfrage keinen Inhalt zurück, wird die in der LOCKSS-Box gespeicherte Version zurückgeliefert.

- LOCKSS stellt eine web-basierte Administrationsoberfläche bereit, die es den Bibliotheksmitarbeitern erlaubt, neue Inhalte einzuspielen, den Status bestehender Inhalte zu kontrollieren und den Zugriff darauf zu steuern.
- Neben dem Ausgangs-Datenformat werden die Inhalte kontinuierlich auch in aktuelle Datenformate migriert, damit Zukunftssicherheit gewährleistet ist.
- Die LOCKSS-Software basiert auf der Technologie der Association of Computing Machinery (ACM) und ist somit OAIS-kompatibel.



<http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf>

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

- CPU: Mindestens 1GHz empfohlen, 2.4GHz Celeron (Ausbaustufe). W
- Arbeitsspeicher: Mindestens 1GB empfohlen
- Festplattenspeicher: Mindestens 250GB. Verwendbare Festplattentypen ATA (PATA) oder ATA (SATA, Änderungen im BIOS eventuell notwendig)
- CD-Laufwerk oder USB Flash-Laufwerk mit Schreibschutzschalter

**Lizenzierungsmodell:**

Verlage: Offene Lizenz

Bibliotheken: Mitgliedschaft in der Allianz verpflichtend (Jahresbeitrag mindestens 1.080 US-\$)

Dokumentation:

http://lockss.stanford.edu/lockss/Installation_Instructions

Weitere Informationen:

<http://lockss.stanford.edu/lockss/Home>

Empfehlung:

Die Open Source-Datenhaltungslösung LOCKSS ist eine Non-Profit-Lösung für Bibliotheken und Verlage, um ihre Titel in digitaler Form der Öffentlichkeit zur Verfügung zu stellen möchten. Ziel von LOCKSS ist die langfristige und kostenlose Verfügbarmachung insbesondere nicht mehr lieferbarer Titel. Bislang beteiligen sich rund 400 Bibliotheken und Verlage.

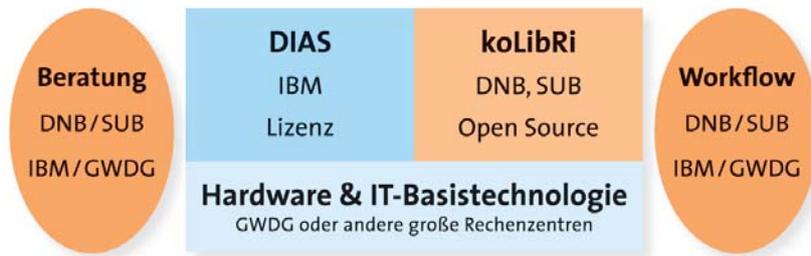
Kopal**Was ist kopal?**

kopal betreibt kooperativ ein Archivsystem für die sichere Langzeitverfügbarkeit digitaler Daten. Dabei ist zu gewährleisten, dass diese nicht nur erhalten, sondern dauerhaft interpretierbar und damit nutzbar bleiben. Der Service richtet sich besonders an Institutionen, die für die Langzeitarchivierung von digitalen Materialien verantwortlich sind, wie Bibliotheken, Archive und Museen sowie Universitäten und Forschungseinrichtungen.

kopal ist relevant für digitale Sammlungen aller Art wie Digitalisate, wissenschaftliche Publikationen und Lehrmaterialien, aber auch für Inhalte, die über Nationallizenzen erworben und nun für einen langfristigen Zugriff archiviert werden sollen.

Servicekomponenten von kopal

kopal bietet ein flexibel konfigurierbares und anpassbares System an. Die Komponenten von **kopal** können je nach Kundenanforderung vereinbart und ausgestaltet werden. Die Modalitäten hängen außerdem vom gewünschten Nutzungsmodell (s.u.) ab.



http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf

DIAS (Digital Information Archiving System) bildet den Kern des kopal-Archivsystems und wurde von IBM für kopal zu einem mandantenfähigen System mit der Möglichkeit des Fernzugriffs weiterentwickelt. Es basiert auf IBM-Standardsoftwarekomponenten sowie ergänzenden, für kopal entwickelten OASIS (Open Archival Information System)-konformen Modulen.

Über die Software selbst hinaus werden hierbei folgende Servicekomponenten angeboten:

- Beratung bzgl. Einsatz und Software-Lizenzen
- Beratung bzgl. Speicherverwaltung
- Installation und Betrieb des Systems.

koLibRI (kopal Library for Retrieval and Ingest) ist eine Open-Source-Softwarebibliothek, die das Erstellen, Einspielen und Abfragen von Archivpaketen, also den gespeicherten Objekten und dazugehörigen Metainformationen, sowie künftig Maßnahmen der Langzeiterhaltung wie Datenformatmigrationen unterstützt. Die für jede Institution spezifischen Arbeitsabläufe für Datenimport und Datenexport werden mittels der Module von koLibRI modelliert und konfiguriert.

koLibRI-Servicekomponenten sind

- Beratung bzgl. des Einsatzes von koLibRI
- Modellierung von Workflows
- Anpassungsentwicklung bei besonderen Kundenanforderungen
- Installation und Support des Betriebs.

Ziel des Projektes kopal ist der Aufbau einer technischen und organisatorischen Lösung, um die Langzeitverfügbarkeit elektronischer Publikationen zu sichern. Dabei spielt die transparente Integration in vorhandene Bibliothekssysteme und die Nachnutzbarkeit durch Gedächtnisorganisationen eine wesentliche Rolle.

Bei der Implementierung des Systems werden internationale Standards in den Bereichen Langzeitarchivierung und Metadaten verwendet. Auf diese Weise werden sowohl die Zukunftsfähigkeit als auch die Möglichkeit zu einer Weiterentwicklung des Systems gewährleistet.

Innerhalb des Projekts werden digitale Materialien aller Art der Partner Deutsche Nationalbibliothek und Niedersächsische Staats- und Universitätsbibliothek Göttingen im Massenverfahren in das Langzeitarchiv eingestellt. Sie sollen von digitalen Dokumenten in Form von PDF, TIFF oder TeX bis hin zu komplexen Objekten wie digitalen Videos reichen.



Der technische Betrieb des Langzeitarchivs ist bei dem Rechenzentrumspartner Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) angesiedelt. Der Entwicklungspartner IBM Deutschland GmbH ermöglicht eine professionelle Anpassung der Softwarekomponenten und bietet eine langfristig stabile Unterstützung.

Das Projekt kopal möchte die sichere Speicherung und langfristige Verfügbarkeit digitaler Daten gewährleisten. Eine Nachnutzung durch weitere Kulturerbeinstitutionen sowie durch sonstige Institutionen, die Langzeitarchivierung benötigen, ist ausdrücklich erwünscht.

Die kopal-Solution ist von vornherein auf unterschiedliche Bedürfnisse ausgerichtet. Es gibt einerseits die Möglichkeit, als Mandant mit einem eigenen „Schließfach“ das bestehende System mit abgesichertem Speicherplatz und eigener Verwaltung der Daten zu nutzen. Diese Lösung ist besonders geeignet für kleinere Institutionen oder Institutionen mit geringerem Archivgutaufkommen.

Andererseits besteht die Möglichkeit einer Nachnutzung der kopal-Solution durch eine eigene Installation des DIAS-Core, der zusammen mit den von kopal entwickelten kpal-Tools betrieben werden kann.

Anbieter:

kopal - Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen
Frank Klaproth
Platz der Göttinger Sieben 1
37073 Göttingen

Projektpartner:

Deutsche Nationalbibliothek
SUB Göttingen GWDG IBM (Link)
gefördert vom Bundesministerium für Bildung und Forschung

Referenzprojekte:

kopal Library for Retrieval and Ingest

Die kopal Library for Retrieval and Ingest (koLibRI) stellt eine Bibliothek von Java-Tools dar, die im Projekt kopal für das Zusammenspiel mit dem DIAS-System der Firma IBM entwickelt wurden. Sie wurde bewusst so angelegt, dass sie als Ganzes oder in Teilen auch in anderen Zusammenhängen nachnutzbar ist.

Das Projekt kopal stellt die koLibRI-Software (kopal-Tools) der interessierten Öffentlichkeit zur Verfügung in der Hoffnung, dass andere Projekte bzw. Entwicklungen von diesen Ergebnissen profitieren können. Intention ist es, dass Gedächtnisorganisationen und andere an der Langzeitarchivierung elektronischer Materialien interessierte Einrichtungen diese Tools analysieren und evaluieren, um zu sehen, ob und inwieweit diese Tools potentiell zur eigenen Software-Architektur bzw. den eigenen Archivierungskonzepten passen. Da es sich um



Ergebnisse eines mit öffentlichen Mitteln geförderten Projekts handelt, ist es ausdrücklich erwünscht, Ergebnisse und Aussagen, die sich aus dem Evaluieren und/oder Erproben dieser Software ergeben, der Öffentlichkeit weiterzugeben. Die vorliegende koLibRI-Software (kopal-Tools) kann jedoch nicht dazu benutzt werden, auf das im Projekt kopal entstandene und in Göttingen stehende Archivsystem zuzugreifen. Eine Nutzung dieses Systems erfolgt im Projekt zurzeit ausschließlich von der Deutschen Nationalbibliothek und der Niedersächsischen Staats- und Universitätsbibliothek Göttingen.

http://kopal.langzeitarchivierung.de/kolibri/koLibRI_v1_0_dokumentation.pdf

Funktionsumfang:

Unterstützte Standards:

DC (Dublin Core)

LMER (Langzeitarchivierungsmetadaten für elektronische Ressourcen)

METS (Metadata Encoding & Transmission Standard)

OAIS (Open Archival Information System)

URN (Uniform Resource Name)

Vgl. http://kopal.langzeitarchivierung.de/index_standards.php.de

Die beteiligten Institutionen wählen digitale Objekte in Form von PDF, TIFF oder TeX bis hin zu komplexen Objekten wie digitalen Videos unter Berücksichtigung formaler Kriterien und Sammelschwerpunkten aus.

Um die Objekte wieder findbar zu machen, werden diese mit inhaltlichen und technischen Metadaten versehen.

Objekte und Metadaten werden zu Datenpaketen gebündelt und auf Vollständigkeit überprüft.

Die Metadaten werden systematisch erfasst um gezielt auf die Pakete zugreifen zu können.

Auswahl:

Die jeweilige Institution wählt digitale Objekte aus, die langzeitarchiviert werden sollen. Sie wird dabei inhaltliche und formale Kriterien berücksichtigen wie Sammelschwerpunkte, besondere Digitalisate etc.

Sammeln und Erstellen der Metadaten:

Um ein systematisches Ablegen und Wiederfinden von Objekten zu ermöglichen, werden diesen Zusatzinformationen wie bibliographische Daten hinzugefügt. Technische Metadaten sind notwendig, um die Objekte regelmäßig auffrischen und migrieren zu können. Die Metadaten werden zum Teil aus Informationssystemen geholt, die die Objekte bereits erschließen, zum Teil mittels spezieller Software aus den Objekten selbst generiert.

Erstellung eines Einlieferungspaketes:

Die digitalen Objekte werden zusammen mit ihren Metadaten in einem speziellen Format, dem Universellen Objektformat (UOF, siehe unten) als Paket gebündelt.

Einbringen in das Langzeitarchiv (Ingest):



Vor dem Einbringen in das Archiv überprüft die Software die Daten noch auf Vollständigkeit und formale Korrektheit.

Umwandlung in ein Archivierungspaket:

Die Metadaten werden in das Data Management überführt. Die Contentdaten und die zugehörige Metadatendatei wandern auf einen von DIAS verwalteten Massenspeicher.

Abfrage und Anforderung von Informationen (Access):

Über das Datenmanagement kann gezielt auf die Metadaten und damit auf die Archivpakete zugegriffen werden.

Auslieferung: Je nach Bedarf werden Metadaten und/oder das Archivmaterial selbst in einem entsprechenden Paketformat ausgeliefert.

Verwendung der Daten: Nutzer von digitalen Objekten greifen i.d.R. über ein Informationssystem auf die Daten zu. Dabei kann dem Nutzer angezeigt werden, dass er gerade auf langzeitarchivierte Daten zugreift. Zugleich kann ihm die Option eingeräumt werden, die Daten in einem bestimmten – nicht notwendig aktuellen – Format auszuwählen.

kopal-Solution

Die kopal-Solution besteht aus dem weiterentwickelten DIAS-Core der IBM und den von den Partnern SUB Göttingen und der Deutschen Nationalbibliothek entwickelten kopal-Tools. Dabei beruht der DIAS-Core auf Standardsoftwarekomponenten der IBM.

kopal-Tools

Bei der Verwendung des kopal-Langzeitarchivierungssystems sollen unterschiedliche Funktionalitäten bei der Bereitstellung von Metadaten sowie Archivdaten genutzt werden können. Die Deutsche Nationalbibliothek und die SUB Göttingen erstellen auf DIAS-Core abgestimmte Softwareprodukte, die unter einer Open-Source-Lizenz stehen. Die Anpassung an eigene Systeme und Bedürfnisse ist ausdrücklich erwünscht, mit dem Ziel einer kooperativen Nutzung der Software.

Eine Beta-Version der kopal-Tools zur Verdeutlichung der modularen Arbeitsweise inklusive Quellcode, Dokumentation und einiger Beispiele finden Sie unter [koLibRI - kopal Library for Retrieval and Ingest](#).

Die Access-Tools sorgen für den strukturierten Zugriff der Institutionen auf ihren Datenbestand, um sie ihren Nutzern zur Verfügung zu stellen.

Die Ingest-Tools unterstützen die Zusammenstellung und Aufbereitung der Daten für das Archiv als Datenpaket entsprechend der Objektspezifikation.

DIAS-Core

Im Rahmen des kopal Projektes entwickeln die Deutsche Nationalbibliothek, die SUB Göttingen, die GWDG und die IBM eine kooperative Langzeitarchivierungslösung. Diese legt DIAS-Core, den Kern des [Digital Information Archiving System](#) der Firma IBM zugrunde, das bereits von der Koninklijke Bibliotheek, der Nationalbibliothek der Niederlande (KB), genutzt wird. Entsprechend des OAIS-Standards für Archivsysteme verwaltet es die



gespeicherten Daten und stellt eine Umgebung für unterschiedlichste Preservation-Strategien zur Verfügung.

DIAS' Mandantenfähigkeit und präzise definierte Import- und Exportschnittstellen ermöglichen es, eine Archivnutzung in verschiedensten Workflows und Institutionen zu integrieren. Durch den Einsatz von vielfach bewährten Standardsoftwarekomponenten (wie z.B. IBM Content Manager und IBM Tivoli Storage Manager) sind langfristige Stabilität, Performanz und Skalierbarkeit gewährleistet.

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

N. N.

Lizenzierungsmodell:

Open Source

Kostenfaktoren bei einer Nutzung

Abhängig vom individuellen Servicekonzept sind folgende Kostenfaktoren zu berücksichtigen:

- Erstberatung (kostenlos)
- Projektplanung und Consulting
- Hardware: Speichersysteme und infrastrukturelle Betreuung
- Lizenz für DIAS-Software inkl. integrierter Komponenten
- Weitere Aufwendungen: z.B. Anpassungsentwicklung der Open-Source-Software koLibRI, konkreter Dateningest durch DNB/SUB

Dauerbetrieb:

Die individuellen Kosten sind dabei abhängig von

- der Zahl und Komplexität der Workflows bei einer Kundeninstitution
- der Menge, Heterogenität und Komplexität der zu archivierenden Objekte und ihrer Metadaten
- den gewünschten Zugriffsmöglichkeiten und Schnittstellen.



Exemplarische Kostenkalkulation

Bei den verschiedenen Nutzungsmodellen von kopal fällt folgender Aufwand an:

- „kopal-Teilnehmer“: Die Institution betreibt koLibRI selbst und kauft weitere Servicekomponenten bei den kopal-Partnern ein. Für die Einrichtung von Workflows ist in einem einfachen Fall mit einem Aufwand von etwa 10 Personentagen zu rechnen. Die Kosten der kommissarischen Archivierung hängen zudem von der Masse des Archivmaterials ab.
- „kopal-Mandant“: Alle DIAS-Servicekomponenten werden vereinbart. Bezüglich koLibRI ist eine Auswahl zu treffen, welche Komponenten teilweise selbst übernommen oder eingekauft werden. Der Umfang der DIAS-Beratungsleistungen kann den Wünschen des Kunden entsprechend sehr unterschiedlich ausfallen. Hier reicht das Spektrum von einem zweistündigen technischen Vortrag bis hin zu einem mehrtägigen Workshop mit einer Feinanalyse der kundeninternen Prozesse. Dementsprechend gestalten sich die Beratungskosten. Für einen zweitägigen Workshop mit der Zielsetzung „Vorstellung der DIAS-Software und Diskussion technischer Fragen“ werden sich Kosten von ca. 5.000 € ergeben. Für Lizenzierung des DIAS fallen je nach Größe des Kunden Kosten von ca. 96.000 bis 385.000 € an. Für eine Mitnutzung des bestehenden Systems als kopal-Mandant ist mit jährlichen Kosten von 50.000 bis 200.000 € zu rechnen, abhängig vom Beratungs- und Nutzungsumfang (Datenvolumina, Zugriffsanforderungen etc.).
- „kopal-Eigenbetrieb“: Das kopal-Archivsystem wird unter Zukauf von Beratung und ggf. Entwicklung eigenständig betrieben. Hier ist von einem Erstaufwand für Hard- und Software eines Systems mittlerer Größe von ca. 750.000 € auszugehen. Hiervon entfallen 40% auf Softwarelizenzen und 60% auf Systembereitstellung und -betrieb.

Quelle: http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf

Dokumentation:

http://kopal.langzeitarchivierung.de/index_downloads.php.de (siehe „Technische Dokumentation“)

Weitere Informationen:

http://kopal.langzeitarchivierung.de/downloads/kopal_Broschuere_2006.pdf

Kopal Demonstrator: http://kopal.langzeitarchivierung.de/index_demonstrator.php.de

Empfehlung:

Kopal ist ein Service, der sich an Gedächtnisinstitutionen wie Bibliotheken, Archive, Museen, Universitäten und Forschungseinrichtungen richtet, deren gesellschaftlicher Auftrag in der Konservierung von Wissen besteht. Mittels des kooperativ betriebenen Archivsystems kann die Langzeitverfügbarkeit digitaler Daten sichergestellt werden.



DIAS

DIAS (Digital Information Archiving System) ist ein Archivierungssystem, das dazu geschaffen wurde, digitalisierte Objekte über einen Zeitraum von 100 Jahren und länger zu speichern und abrufbar zu machen.

Ursprünglich wurde das IBM System DIAS 2000 für die niederländische Nationalbibliothek (Koninklijke Bibliotheek) entwickelt. DIAS basiert auf dem Open Archival Information System (OAIS) Referenzmodell. Derzeit enthält das DIAS System rund 15 Millionen Artikel.

Anbieter:

IBM

Referenzprojekte:

Koninklijke Bibliotheek Amsterdam

Problem:

Die niederländische Nationalbibliothek (KB) stand vor dem Problem, große Mengen digitaler Dokumente über einen langen Zeitraum verfügbar zu halten. Eine Softwarelösung, auf die man hätte zurückgreifen können, existierte nicht.

Lösung:

Mittels der Systeme IBM Content Manager (Inhalte), IBM Tivoli Storage Manager (automatisches backup) und IBM WebSphere wurde das DIAS System realisiert.

Vorteile:

- Sicherstellung, dass heutige digitale Dokumente auch in 100 Jahren noch zugänglich sind.
- Renommee der KB
- Kapazität 100.000e digitale Dokumente pro Jahr zu integrieren.
- Es können digitale Objekte, die in vielfältigen Formaten vorliegen, in den Bestand aufgenommen werden.
- schnellerer Zugriff für die Benutzer

Vgl. <http://www-935.ibm.com/services/nl/dias/ref/references.html>

Funktionsumfang:

- DIAS basiert auf Open Archival Information System (OAIS).
- DIAS benutzt offene Standards wie XML, SOAP, J2EE
- Damit DIAS zukünftig mit wechselnden Datenformaten Schritt halten kann, wurde dafür gesorgt, dass das core model extendable ist.



- Die bereitgehaltenen digitalen Objekte können sowohl vor Ort als auch über das Internet abgerufen werden.
- Multi-organisation support (mehrere Teilnehmer können sich ein DIAS teilen)
- Distributed AIP storage
- Sicherheitssysteme und Backup

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

N.N.

Lizenzierungsmodell:

IBM DIAS Einzelinstallation

Stufe	Anzahl MA	Preisansatz DIAS Lizenz (BVB ÜL)	Preisansatz DIAS Wartung (BVB Pflege)
1	< 100		
2	100 - 249	250.000,00	70.000,00
3	250 - 499	395.000,00	106.650,00
4	> 500	575.000,00	149.500,00

IBM DIAS als Mandant

Stufe	Anzahl MA	Preisansatz DIAS Lizenz (BVB ÜL)	Preisansatz DIAS Wartung (BVB Pflege)
1	< 100	25.000,00	7.500,00
2	100 - 249	55.000,00	15.950,00
3	250 - 499	150.000,00	42.000,00
4	> 500	455.000,00	113.750,00

Quelle: http://kopal.langzeitarchivierung.de/downloads/kopal-goes-live_IBM_DIAS_Perspektive_und_Preismodell.pdf

Dokumentation:

N. N.

**Weitere Informationen:**

DIAS Perspektive und Preismodell: http://kopal.langzeitarchivierung.de/downloads/kopal-goes-live_IBM_DIAS_Perspektive_und_Preismodell.pdf

Empfehlung:

DIAS ist ein Speichersystem für Gedächtnisinstitutionen, das speziell auf Langzeitarchivierung ausgerichtet ist. Durch eine kontinuierliche Weiterentwicklung von Basistechnologie und Datenformaten sollen Anwender digitalisierte Objekte in DIAS über einen Zeitraum von 100 Jahren und länger vorhalten und abrufbar machen können.

Tivoli Storage Manager

Der Tivoli Storage Manager ist ein Sicherungs-, Archivierungs- und Speicherverwaltungstool für Dateiserver, Workstations, Anwendungen und Anwendungsserver. Bis 1999 war die Software bekannt als ADSTAR Distributed Storage Manager (ADSM).

Anbieter:

IBM

Referenzprojekte:

Bayer. Staatsbibliothek (Leibnitz-Rechenzentrum)

Funktionsumfang:

- Inkrementelles Backup und Archivierung auf Festplatte und auf Band
- Auslagerung von Daten nach definierten Kriterien
- Sichern von Anwendungsdaten (z. B. Oracle-DB, SAP-R/3, MS SQL Server)



Features, advantages and benefits		
Features	Advantages	Benefits
Backup and Recovery Management	Intelligent backups and restores utilizing a revolutionary progressive incremental backup and restore strategy, where only new and used files are backed up	Centralized protection based on smart-move and smart-store technology leading to faster backups and restores with less network and storage resources needed
Hierarchical Storage Management	Policy-based management of file backup and archiving	Ability to automate critical processes relating to the media on which data is stored while reducing storage media and administrative costs associated with managing data
Archive Management	Managed archives	Ability to easily protect and manage documents that need to be kept for a certain period of time
Advanced Data Reduction	Combines progressive-incremental backup, source and target data deduplication, compression and tape management to provide best-in-class data reduction	Reduces the costs of data storage, environmental requirements, and administration

Quelle: <http://www-01.ibm.com/software/tivoli/products/storage-mgr/features.html>

Funktionen bezogen auf Phasen des Data-Lifecycle:

Full Lifecycle Actions

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Server

- IBM AIX®
- HP®-UX
- Windows®
- Sun® Solaris
- OS/390® z/OS
- Linux x86_64
- Linux x86
- Linux zSeries
- Linux on Power
- Linux on Itanium 2 (Linux IA64)

**Clients:**

- AIX
- HP/UX
- HP/UX Itanium
- Linux for IA64
- Linux on Power - iSeries/pSeries
- Linux on x86 / x86_64
- Linux zSeries
- Macintosh
- Novell NetWare
- OS/390, zSeries USS
- OS/400
- Sun Solaris
- Solaris x86 / x86_64
- Windows
- Citrix Presentation Server 3.0 for Windows 2000 and 2003

Vgl. <http://www-01.ibm.com/software/tivoli/products/storage-mgr/platforms.html>

Lizenzierungsmodell:

Client-Lizenz

72,25 US-\$

oder

Processor Value Unit (PVU)-Lizenz

382,50 Us-\$

Dokumentation:

Installationsanleitungen für Linux, Solaris, AIX etc. siehe <http://www-01.ibm.com/support/search.wss?rs=663&tc=SSGSG7&dc=DA410+DA450&dtm>

Weitere Informationen:

IBM Tivoli Storage Manager Features and Benefits: <http://www-01.ibm.com/software/tivoli/products/storage-mgr/features.html>

Empfehlung:

Der Tivoli Storage Manager zeichnet sich durch seine Plattformunabhängigkeit aus, da sich die Software in allen gängigen Server- und Client-Umgebungen betreiben lässt. Die Lösung bringt die gängigsten Funktionen für die Administration von Daten-Repositories mit, bietet aber nicht die Vielfalt speziell am Datenhaltungsbedarf wissenschaftlicher Institutionen ausgerichteter Funktionalitäten wie Kopal oder DIAS. Dennoch lässt sich mit dem Tivoli Storage Manager ein großer Teil des Kurations-Lifecycles zu einem attraktiven Preis-Leistungs-Verhältnis abdecken.



Kap. 6: Datenmanagement-Systeme

FuD (Forschungsnetzwerk und Datenbanksystem)

Integrierte Arbeits-, Publikations- und Informationsplattform für die Geisteswissenschaften

Unterstützt die Zusammenarbeit in räumlich verteilten Arbeitsgruppen während der verschiedenen Phasen des Forschungsprozesses von der Inventarisierung und Erfassung der Primärdaten über ihre Erschließung und Analyse bis hin zur Ergebnispublikation und Datenarchivierung

Dezentrale Arbeitsweise: parallele Dateneingabe in eine relationale Datenbank an verteilten Arbeitsplätzen:

- Online-Verfügbarkeit der Daten
- Umfangreiches Benutzer- und Dokumentenmanagement
- Simultane Forschungsarbeit
- Standardisierte Arbeitsabläufe

Anbieter:

Universität Trier

Referenzprojekte:

- [Corpus der Quellen zur mittelalterlichen Geschichte der Juden im Reichsgebiet](#), Akademie der Wissenschaften und der Literatur, Mainz
- [Bilddatenbank zur antiken Sklaverei](#), Universität Trier, Akademie der Wissenschaften und der Literatur, Mainz
- [Emil-Frank-Institut](#), Wittlich

**Funktionsumfang:**

- **Inventarisierung:** Strukturierte Speicherung von Forschungsdaten und Textquellen (z. B. Urkunden, Verwaltungsschriftgut, Zeitungsartikel, visuelle Medien, archäologische Objekte, Tondokumente); Module für Volltextrecherche, kollaborative Dokumentenverwaltung, Bibliographie sowie Druck und Export
- **Textanalyse und -annotation:** themenspezifische Sachanalyse, Semantikanalyse (Wortfeld- und Argumentationsanalyse) und ein freies, individuell aufbaubares, hierarchisch organisiertes Kategorienschema, das frei von jedem Wissenschaftler für spezielle Fragestellungen entwickelt werden kann
- **Redaktion:** Texteditor mit Freigabeworkflow und Layoutmodul sowie Anmerkungsapparat und Register
- **Publikation und Langzeitarchivierung:** Freie oder zugangsbeschränkte Online-Publikation mit Volltextindex; Anschluss an virtuelle Fachportale und Bibliotheken möglich; Referenzierbarkeit über URNs bzw DOIs; Standardisierter Export in XML (Kodierung nach TEI)

Funktionen bezogen auf Phasen des Data-Lifecycle:**Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

Client:

- ab Windows 2000
- Bildschirmauflösung: 1024*768 oder höher
- Internetanbindung: LAN / DSL
- freier Speicherplatz zur Installation ca. 70 MB
- Java-Software für die Online-Hilfe

Server:

- Linux
- MySQL Datenbank
- Tcl/Tk

**Lizenzierungsmodell:**

N. N.(Fertigstellung des Tools Ende 2012)

Dokumentation und weitere Informationen: <http://fud.uni-trier.de/>

Empfehlung:

Dank dem Schwerpunkt auf Online-Kollaboration eignet sich das Framework besonders für Kurationsprozesse, deren einzelne Arbeitsschritte an räumlich verteilten Standorten ausgeführt werden. Weitere Stärken der Software liegen im sprachtechnologischen Bereich, was einen Einsatz zu Textanalysezwecken nahelegt.

eSciDoc

Plattform zur disziplinübergreifenden Haltung, Visualisierung, Nutzung und Publikation von Forschungsdaten

Keine monolithische Applikation, sondern generische Infrastruktur („service-oriented architecture“), die für die unterschiedlichen Aufgaben so weit als möglich existierende Lösungen nutzt, z.B.: Fedora (Storage), SRW/U (Suche), Fedora (Indexierung), Shibboleth (Authentifizierung)

Anbieter:

Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. und FIZ Karlsruhe - Gesellschaft für wissenschaftlich-technische Information mbH

Existierende Softwarelösungen auf Basis von eSciDoc:

- PubMan (Publication Management): Speicherung von Publikationsdaten und Forscherprofilen; einfacher Publikations- und Kollaborationsworkflow; Export nach EndNote, BibTeX, XML (weitere Informationen: http://colab.mpg.de/mediawiki/PubMan_Functionalities)
- VIRR (Virtueller Raum Reichsrecht des Max-Planck-Instituts für europäische Rechtsgeschichte): Digitale Textsammlung zu juristischen Themen des heiligen römischen Reichs; kollaborative Annotierung per Online-Interface (weitere Informationen: <https://www.escidoc.org/JSPWiki/en/VirtuellerRaumReichsrecht>)
- Scholarly Workbench: Kollaboratives Autorentool für die Geisteswissenschaften (FIZ Karlsruhe), weitere Informationen: https://www.escidoc.org/media/docs/usc_sw_overview_v07.pdf

**Funktionsumfang (Services):**

- Object Manager
 - Item Service
 - Container Service
 - Context Service
 - Content Model Service
 - Organizational Unit Service
- Security (AA)
 - Policy Decision Point
 - User Account Service
 - User Group Service
 - Role Service
- Statistics
- Search
- Admin Service
- Set Service

Funktionen bezogen auf Phasen des Data-Lifecycle:**Full Lifecycle Actions**

- * Description and Representation Information
- * Preservation Planning
- * Community Watch and Participation
- * Curate and Preserve

Sequential Actions

- * Conceptualise
- * Create or Receive
- * Appraise and Select
- * Ingest
- * Preservation Action
- * Store
- * Access, Use and Reuse
- * Transform

Systemvoraussetzungen:

- Programmiersprache: J2EE
- Datenbank: PostgreSQL
- Application Server: JBoss
- Servlet Container: Tomcat
- Content Repository: Fedora

Lizenzierungsmodell:

Common Development and Distribution License (CDDL) 1.0

Dokumentation und weitere Informationen:

<https://www.esdoc.org/>

**Empfehlung:**

Der service-orientierte Ansatz sowie die hohe Kompatibilität zu gängigen Software-Applikationen macht das Framework besonders geeignet für Anwender, die eine umfassende Datenmanagementlösung in eine bestehende Infrastruktur aus Such- und Indexierungsmechanismen integrieren wollen.