FH;P
Fachhochschule Potsdam
University of
Applied Sciences

Fachbereich Informationswissenschaften
Studiengang Bibliotheksmanagement

Persistent Identifier für wissenschaftliche Einrichtungen

Bachelorarbeit

zur Erlangung des akademischen Grades
Bachelor of Arts (B.A.)

eingereicht von: Laura Rothfritz
Matrikelnummer: 13817

Gutachter/innen: Prof. Dr. rer. nat. Heike Neuroth
Dipl. Bib. Heinz Pampel

eingereicht am: 27. Januar 2018 verteidigt am: 7. Februar 2018

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und noch nicht für andere Prüfungen eingereicht habe. Sämtliche Quellen einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, sind als solche kenntlich gemacht. Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung eingeleitet wird.

Berlin, den 26. Januar 2018 .....

Abstract

In der vorliegenden Arbeit werden Probleme bei der eindeutigen Identifizierung wissenschaftlicher Einrichtungen aufgezeigt, Persistent Identifier als Lösung dieser Probleme vorgeschlagen und bereits bestehende Lösungsansätze in Form von Identifikatorensystemen für wissenschaftliche Einrichtungen ausgewertet. Außerdem werden Anwendungsszenarien für die eindeutige Identifikation wissenschaftlicher Einrichtungen und bisherige Lösungen vorgestellt. Die betrachteten Identifikatorensysteme werden bewertet und aus den Anwendungsfällen Kriterien für Persistent Identifier für wissenschaftliche Einrichtungen herausgearbeitet. Offene Forschungsinfrastrukturen als Organisationsform für ein Identifikatorensystem bieten hierfür die Grundlage. Bisher erfüllen keine der betrachteten Identifikatoren die Kriterien für Persistent Identifier und können nicht als diese bewertet werden. Es bestehen jedoch Anwendungsfälle, für die Persistent Identifier zur eindeutigen, anhaltenden Identifizierung von Einrichtung sowie dem automatisierten Austausch von Metadaten sehr gut geeignet sind.

Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 2.0 Deutschland zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie
`http://creativecommons.org/licenses/by/2.0/de/` oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

Inhaltsverzeichnis

1 Einleitung **1**

2 Forschungsstand **4**
2.1 Persistent Identifier 4
2.2 Bestehende Identifikatorensysteme 7
2.3 Projekte 8

3 Methode **12**

4 Identifikatoren für wissenschaftliche Einrichtungen **14**
4.1 Die Gemeinsame Normdatei (GND) 14
4.2 Der International Standard Name Identifier (ISNI) 15
4.3 Ringgold Identifier 17
4.4 Die Global Research Identifier Datenbank (GRID) 19
4.5 Identifikatoren der Deutschen Forschungsgemeinschaft (DFG ID) 20
4.6 Participant Identification Code (PIC) 21
4.7 Scopus Affiliation ID 21
4.8 CrossRef Funder ID 22

5 Nutzung von Identifikatoren **25**
5.1 Drittmittelanträge (Forschende) 25
5.2 Aggregation und Reporting (Forschungsinfastrukturen) 25
5.3 Automatisierung von Datenflüssen (Forschungsinfrastrukturen) 27
5.4 Tracking (Forschungsförderung) 28
5.5 Gutachtenverfahren für Anträge (Forschungsförderung) 30
5.6 Open-Access-Monitoring und APC-Management 31

6 Diskussion **33**
6.1 Einordnung der Identifikatorensysteme 33
6.2 Anforderungen an Metadaten 38
6.3 Bewertung der Anwendungsfälle 40
6.4 Offene Forschungsinfrastrukturen 41

7 Fazit und Ausblick **44**

A Anhang **54**
A.1 Anwendungsfälle für OrgIDs 55
A.2 IDs für das GeoForschungsZentrum Potsdam in Scopus 57
A.3 Antworten zum RFI der Organization Identifier Working Group 60

B Abbildungen **86**

Tabellenverzeichnis

1 Bereits durchgeführte Projekte, Untersuchte Identifier und Ergebnisse 11
2 Korrelierende Anwendungsfälle in jeweils mind. 2 Studien 13
3 Unterschiede zwischen einfachen, eindeutigen und Persistent Identifiern . . . 33
4 Bewertung der untersuchten Identifikatoren 35

Abbildungsverzeichnis

1 ISNI+ als hybrider Organizational Identifier. Aus: Ferguson, Moore und Schmoller, 2014, S. 19 86
2 Affiliation Profile der *Universität zu Berlin* in Scopus 86

1 Einleitung

Kommunikationsprozesse in der Wissenschaft bilden sich in Informationsflüssen zwischen den unterschiedlichen Akteursgruppen ab und münden in messbaren Produkten, dem Output der Wissenschaft. Die Identifizierung, Benennung und Verknüpfung von Forschenden, Forschungsinfrastrukturen, Forschungsförderern und Forschungsoutputs kann dabei zu einem nachvollziehbaren Netzwerk führen, welches den wissenschaftlichen Fortschritt trägt und vorantreibt. Dennoch ist gerade im digitalen Austausch von Informationen über Beteiligte am Forschungsprozess und die entstandenen Ergebnisse diese Identifizierung nicht immer eindeutig und leicht durchführbar.
Fehlerhafte Namen von wissenschaftlichen Einrichtung in Zitationsdatenbanken führen bei bibliometrischen Auswertungen von Zitationsdaten unter Umständen zu falschen Ergebnissen (vgl. De Bruin und Moed, 1990). Nicht standardisierte Benennungen von wissenschaftlichen Einrichtungen beruhen auf Namensvarianten, Schreibfehlern, unterschiedlichen Benennungen in verschiedenen Sprachen, Abkürzungen, semantischen Fehlern oder unklaren Hierarchieebenen (vgl. Huang u. a., 2014). Zudem verändern Institutionen mit der Zeit häufig ihre Struktur durch Zusammenlegungen, Aufsplittung in distinkte Einrichtungen, Übernahme durch größere Einrichtungen usw. (vgl. Hood und Wilson, 2003). Sind Adress- oder Namensfelder unzureichend formatiert, entstehen auch durch die automatische Indexierung von Einrichtungsnamen und -adressen Fehler in der eindeutigen und korrekten Bezeichnung der Körperschaft (vgl. Taşkın und Al, 2013).
Der Prozess der Auflösung doppeldeutiger Zeichenketten zur Benennung von Entitäten wird als *Disambiguierung* benannt. Im Mittelpunkt stehen dabei zwei Formen von mehrdeutigen Beziehungen zwischen Zeichenketten und Entitäten: *Synonyme* bezeichnen zwei unterschiedliche Zeichenketten, die die selbe Entität identifizieren, zum Beispiel wenn eine wissenschaftliche Einrichtung unterschiedliche Benennungen hat. *Homonyme* sind zwei oder mehr gleiche Zeichenketten, die verschiedene Bedeutungen haben. Beispielsweise können zwei unterschiedliche Einrichtungen unter dem selben Akronym bekannt sein. Wenn wissenschaftliche Einrichtungen eindeutig identifiziert werden sollen, müssen auch sogenannte „*Sektorenhybride*" (Einrichtungen, die aus Einrichtungen unterschiedlicher Sektoren entstehen)[1] oder „An-Institute", die an Universitäten angegliedert sind, aber rechtlich eigenständig agieren[2], korrekt zugeordnet und benannt werden (vgl. Winterhager, Schwechheimer und Rimmert, 2014). In der Bibliometrie wurden dafür einige Lösungen entwickelt, die auf der automatischen Indexierung mit Hilfe von *Named-Entity Recognition* Verfahren beruhen.[3]
Um der Problematik der uneinheitlichen Bezeichnung von Konzepten in der außersprachli-

[1] Beispiel: Das KIT Karlsruhe, entstanden aus einer Ausgliederung der Universität Karlsruhe und des Forschungszentrum Karlsruhe.

[2] Beispiel: Das Leibniz-Institut für Arteriosklerosеforschung (LIFA) an der Universität Münster - aber nicht Teil der Universität Münster.

[3] Unter Named-Entity Recognition wird die automatisierte Erkennung und Klassifikation von Eigennamen bei Informationsextraktion aus Texten verstanden. Dieses Verfahren wird häufig in der Computerlinguistik verwendet (vgl. hierzu Jiang, 2012). Beispiele für die Anwendung zur automatischen Erkennung von Namen wissenschaftlicher Einrichtungen sind *Enhanced Finite-State Transducer* in Galvez und Moya-Anegón, 2007 oder *Nooj Transducer* in Taşkın und Al, 2013.

chen Realität entgegenzuwirken entwickelten sich im bibliothekarischen Bereich sogenannte *Verzeichnisse für Normdaten*. In diesen werden Personen, Körperschaften, Geografika und Sachverhalte durch normierte Benennungen identifiziert und Publikationen zugeordnet. In Deutschland wird die Gemeinsame Normdatei (GND) als Werkzeug für die eindeutige Zuordnung von Namen verwendet, ein weiteres Beispiel ist die Name Authority Files der Library of Congress in den USA (vgl. Wiechmann, 2014).
Zunehmend wird versucht, das Problem durch die Vergabe von eindeutigen Kennungen zu lösen: *„[N]ame forms are being replaced by universal registered identifiers for individuals in these records, and internation registration agencies for such identifiers are being formed"* (Zhao und Strotmann, 2015, S.101-102).
Im Kontext der eindeutigen Identifikation von Forschenden trotz Namensgleichheiten werden Identifikatoren für Personen als eine gute Lösungsmöglichkeit angesehen. Um wissenschaftliche Publikationen eindeutig mit ihren Autor*inen zu verknüpfen, existieren eine Reihe von Diensten, wie die Researcher ID von Thomas Reuters/Web of Science, den Scopus Author Identifier und Author Profile von Elsevier oder das Virtual International Authority File (VIAF) (vgl. Walker und Armstrong, 2014). In den letzten Jahren hat sich die Open Research Contributer iD (ORCID) als globaler non-profit Anbieter für eindeutige, gleichbleibende und durch Internetbrowser auflösbare Identifikatoren (*Persistent Identifier*) für Personennamen etabliert. ORCID verzeichnet mittlerweile über 4 Millionen IDs[4], welche jedoch nur für individuelle Personen und nicht für wissenschaftliche Einrichtungen vergeben werden. Persistent Identifier (PID) haben zwei grundlegende Eigenschaften: Sie identifizieren ein Informationsobjekt eineindeutig zur Unterscheidung von anderen Objekten und sie ermöglichen die langfristige Auffindbarkeit von Objekten im digitalen Raum, unabhängig davon, an welchem Ort sie gespeichert sind (vgl. Lynch, 1998). Diese Arbeit wird sich vorrangig mit der eindeutigem Bezeichnung von wissenschaftlichen Einrichtungen befassen, die mit Hilfe von Persistent Identifiern in unterschiedlichen Nutzungsszenarien erfolgen kann. Die Nutzung von Persistent Identifiern führt dazu, dass der Informationsfluss zwischen unterschiedlichen Stakeholdern und Entitäten des Forschungsprozesses vereinfacht wird: „*Tracking information flows between researcher, facilities and publishers requires a common understanding and use of specific data fields and associated persistent identifiers*" (Haak, 2017). Für die Verbreitung von und das Vertrauen in Persistent Identifier werden offene Forschungsinfrastrukturen nach Bilder, Lin und Neylon als zuträglich angesehen.
Ziel dieser Arbeit ist es, einen Überblick über bisherige Lösungen für die eindeutige Identifizierung von Institutionen mit Hilfe von Identifikatoren zu schaffen, Nutzungsszenarien zu diskutieren und Vorschläge für die Entwicklung von Persistent Identifiern für wissenschaftliche Einrichtungen zu entwickeln. Dafür wird in Kapitel 2 zunächst der Forschungsstand zu Persistent Identifiern und bereits durchgeführten Projekten zu Identifikatoren für Organisationen umrissen. Nach der Klärung der Methode in Kapitel 3 erfolgt in Kapitel 4 eine Analyse bestehender Identifier sowie in Kapitel 5 ein exemplarische Analyse von Anwendungsfällen. In Kapitel 6 werden die Ergebnisse diskutiert. Es werden folgende Aspekte genauer betrachtet:

[4]Siehe: `https://support.orcid.org/knowledgebase/articles/150557-number-of-orcid-ids` (Abgerufen am 26. Januar 2018).

1. Welche Möglichkeiten zur eindeutigen Identifizierung von wissenschaftlichen Einrichtungen bestehen bereits?
2. Wie lassen sich bestehende Lösungen einordnen und bewerten?
3. Welche Anforderungen an Persistent Identifier für wissenschaftliche Einrichtungen gibt es?
4. Welche konkreten Anwendungsfälle ergeben sich für diese Persistent Identifier im Forschungsprozess?

2 Forschungsstand

2.1 Persistent Identifier

Im Zuge der Zugänglichkeit zu wissenschaftlicher Literatur über das Internet stellt sich die Herausforderung, wie digitale Informationsobjekte eindeutig und langfristig identifizierbar bleiben können, um das Auffinden und die zuverlässige Referenzierbarkeit von wissenschaftlichen Ergebnissen zu garantieren. Seit Anfang der 2000er Jahre erfolgte eine Reihe von Studien, die die mangelnde Persistenz von Belegen und Referenzen in wissenschaftlichen Publikationen aufzeigen (vgl. Lawrence u. a., 2001; Dellavalle, 2003; Hennessey und Ge, 2013). Das Ergebnis dieser Studien zeigt, dass die Referenzierung digitaler Informationsobjekte über einen Uniform Resource Locator (URL) keine langfristig verlässlichen Belege liefern kann, da URLs keine eineindeutige Bezeichnung für ein Objekt darstellen, sondern nur einen „Weg" hin zu einem Objekt beschreiben (vgl. Lynch, 1998). Wenn das referenzierte Objekt verschoben oder gelöscht wurde, der Service des Anbieters (Domain oder Host) nicht mehr existiert oder das Objekt umbenannt wurde, können URLs nicht mehr verarbeitet und das Objekt nicht mehr erreicht werden. Das „verschwinden" von digitalen Objekten aufgrund nicht funktionaler Links wird als *Link Rot* bezeichnet; im Kontext des Verweisens auf Quellen in wissenschaftlichen Publikationen spricht man von *Reference Rot* (vgl. Klein u. a., 2014). Ein Hauptproblem besteht darin, dass URLs sowohl ein Objekt identifizieren, als auch seinen Speicherort im Internet beschreiben (vgl. Dellavalle, 2003).

Um die Zugänglichkeit und Auffindbarkeit sowie Zitierbarkeit von wissenschaftlicher Literatur im Internet sicherzustellen, etablierten sich seit Mitte der 1990er Jahre unterschiedliche Verweis- und Identifikatorensysteme, die unter dem Oberbegriff *Persistent Identifier* zusammengefasst werden.[5] Die Research Data Alliance (RDA) definiert den Terminus Persistent Identifier als:

> „*[A] long-lasting ID represented by a string that uniquely identifies a DO*[6] *and that is intended to be persistently resolved to meaningful state information about the identified DO. An identifier should have an unlimited lifetime, even if the existence of identified entity ceases. This aspect of an identifier is called persistency*" (Berg-Cross, Ritz und Wittenburg, 2015).

Ein Persistent Identifier bezeichnet ein bestimmtes Informationsobjekt, beispielsweise ein Dokument. Informationsobjekte (oder Teile von Objekten) können mit unterschiedlichen PIDs beschrieben werden, jedoch ist einem Identifier immer nur ein Objekt zugeordnet. Es besteht somit eine 1:n-Beziehung zwischen Informationsobjekten und Persistent Identifiern. Ändert sich das Objekt oder verschwindet es, bleibt der Identifier bestehen und wird nicht wieder verwendet. Persistent Identifier haben demnach zwei grundlegende Eigenschaften:

[5] Siehe hierzu Sollins, Masinter (1994): *Functional Requirements for Uniform Resource Names* (RFC 1737). `https://www.ietf.org/rfc/rfc1737.txt` (Abgerufen am 26. Januar 2018).

[6] *Digital Object*, also ein Objekt im digitalen Raum/Internet.

1. Sie sind eineindeutig und einzigartig und
2. sie sind beständig unabhängig von dem Objekt, auf das sie verweisen.

Identifiziert werden die Objekte an sich, Metadaten zu den Objekten und Zugriffswege auf das Objekt. Die Metadaten zu den Objekten werden in sogenannten *PID Records* gespeichert:

> „*A PID record contains a set of attributes stored with a PID describing DO properties*" (Berg-Cross, Ritz und Wittenburg, 2015).

Der Unterschied zwischen PIDs und URLs liegt darin, dass PIDs auf ein identifiziertes Objekt unabhängig davon, wo es zu einem Zeitpunkt gespeichert ist, verweisen. Persistent Identifier an sich beinhalten nur Informationen über das Objekt, unter anderem seinen Speicherort. PIDs werden über ein *Resolver System* zu dem digitalen Objekt geleitet (vgl. Tonkin, 2008). Die RDA definiert dies als Resolution System:

> „*[A resolution system] is a globally available infrastructure system that has the capability to resolve a PID into useful, currentstate information describing the properties of a DO*" (Berg-Cross, Ritz und Wittenburg, 2015).

Überlegungen zur Trennung von Identifikatoren und tatsächlichen Informationsobjekten stammen aus Konzeptionen zur Architektur von digitalen Bibliotheken. Objekte in einer digitalen Bibliothek müssen durch „Namen" und Identifikatoren eindeutig identifizierbar und auffindbar sein, ähnlich wie Buchbestände in einer analogen Bibliothek über eine Signatur gekennzeichnet sind. Diese Namen müssen durch ein administratives System verwaltet werden, welches auf Änderungen der Objekte reagiert und ihren aktuellen Ort langfristig dokumentiert (vgl. Arms, 1995). Persistent Identifier als Identifikatorensysteme im Internet geben über den *Resolving Service* die Möglichkeit, auf eine Repräsentation des Informationsobjektes im digitalen Raum, welche Informationen über das Objekt bereithält, zuzugreifen (vgl. Car, Golodoniuc und Klump, 2017). Der Resolving Service sorgt auch dafür, dass der Identifier eindeutig mit dem Informationsobjekt verbunden bleibt (vgl. Askitas, 2010). Der Resolver identifiziert Informationen über das Informationsobjekt (Metadaten) und zeigt dabei auf seinen Speicherort. Bei der Auflösung des PID werden Informationen (Metadaten, Location URL) zum bezeichneten Objekt zurückgegeben und Anfragende zum Objekt selbst oder einer sogenannten *Landing Page* weitergeleitet. Dieser zweischichtige Aufbau führt dazu, dass, wenn sich der Speicherort des Informationsobjektes ändert, nur die Verknüpfung zwischen dem Identifikator und dem Informationsobjekt geändert wird. Für Nutzer*innen des PID bleibt der Identifikator gleich. PIDs ermöglichen, dass Informationsobjekte und ihre unterschiedlichen Versionen ohne die Gefahr der Verwechslung identifiziert werden können. Sie helfen beim Auffinden und Referenzieren von Informationsobjekten im Internet und sorgen dafür, dass wissenschaftlicher Output (Publikationen, Daten...) verlässlich zitierbar ist. Viele Identifikatorensysteme ermöglichen zudem, über die ihnen zugrundeliegenden Metadaten, die Interoperabilität mit anderen Identifiern und somit die Darstellung von Relationen zwischen Objekten (vgl. Davidson, 2006).

Im wissenschaftlichen Bereich haben sich im Laufe der letzten 20 Jahre unterschiedliche Systeme für Identifier etabliert.[7] Dabei haben sich vor allem der Digital Object Identifier (DOI), Handle PIDs, Persistent URL (PURL), Uniform Ressource Name (URN) und Archival Resource Key (ARK) durchgesetzt.
Klump und Huber untersuchten das Repository-Verzeichnis re3data auf die Nutzung von PID Typen. DOIs wurden in den meisten Repositories als Identifier benutzt (vgl. Klump und Huber, 2017). Sie können mittlerweile als ein quasi-Standard für Identifier für wissenschaftliche Publikationen angesehen werden. DOIs werden zudem nicht nur für traditionelle Publikationen in Schriftform vergeben, sondern auch für andere Formen von wissenschaftlichem Output, wie zum Beispiel Forschungsdaten. Für die Zitierung quantitativer Daten wird als Minimalanforderung das Vorhandensein eines Persistent Identifiers genannt (vgl. Altman und King, 2007).
Die Verlinkung und Interoperabilität zwischen den unterschiedlichen Identifier Systemen, die für die Entitäten bereits bestehen, fördern einen Informations- und Datenfluss über den gesamten Forschungszyklus und für alle Stakeholder. Die Nutzung von interoperablen PID Systemen ermöglicht die Auffindbarkeit von wissenschaftlichem Output, die richtige Zuordnung von Ergebnissen zu ihren Urheber*innen, nachvollziehbare Provenienz von nachgenutzten Quellen oder Daten und verbesserte Metadatenqualität über die Verknüpfung unterschiedlicher Metadatenquellen (vgl. Dappert u. a., 2017). Neuere Entwicklungen zielen darauf ab, PIDs für alle Entitäten im wissenschaftlichen Forschungskontext zu vergeben, um Interoperabilität zwischen ihnen zu schaffen. Folgende Entitäten werden von Dappert u. a. identifiziert:

- *Agents*: Individuelle Forschende, Organisationen und Konsortien
- *Resources*: Publikationen, Forschungsdaten, Software, Forschungsinstrumente, Laborbücher...
- *Rights statements*: Lizenzen, Förderungsmaßnahmen, Patente
- *Events*: Vorgänge, die die Provenienz von wissenschaftlichem Output beschreiben (Erstellung, Kuration, Updates, Zugriff, Claiming, Zitation, Review)
- *Derived entities*: Entitäten wie Projekte, die aus Verbindungen von allen oben genannten Entitäten entstehen oder diese erschaffen

Eindeutige, meist numerische Identifikatoren existieren auch außerhalb des Internets um Objekte (beispielsweise Bücher durch eine ISBN) oder Personen (beispielsweise Bürger eines Landes durch die Passnummer) zu identifizieren. Für die eindeutige Identifizierung von Personen findet seit ihrer Gründung 2012 die ORCID iD immer weitere Verbreitung. Diese Art der Selbst-Identifizierung von Personen in der Form eines offenen Infrastrukturservices wird in Deutschland durch das von der Deutschen Forschungsgemeinschaft (DFG) geförderte Projekt ORCID DE[8] unterstützt. Das Ergebnis einer 2016 durch ORCID DE durchgeführten Umfrage ergab für eine Mehrzahl der Befragten einen sehr hohen bis hohen Bedarf an persistenten Identifikatorensystemen für wissenschaftliche Institutionen, Forschungsinfrastrukturen und Förderorganisationen (vgl. Fuchs, Pampel und Vierkant, 2017, S. 53). Die Nutzung von

[7] Einen guten Überblick geben Hilse und Kothe, 2006: *Implementing persistent identifiers: overview of concepts, guidelines and recommendations*.
[8] Siehe: `http://www.orcid-de.org/` (Abgerufen am 26. Januar 2018).

Identifikatoren für Organisationen ist bisher in Deutschland jedoch nicht sehr weit verbreitet (vgl. Bryant, Dortmund und Malpas, 2017, S. 42).

2.2 Bestehende Identifikatorensysteme

Für die eindeutige Identifikation von wissenschaftlichen Einrichtungen bestehen bereits eine Reihe Identifikatorensysteme, die im Kapitel 4 näher analysiert werden. Im bibliothekarischen Kontext werden zur eindeutigen Identifizierung von Entitäten Normdaten, die in Normdateien gespeichert sind, verwendet. Sie werden aufgrund einer Ontologie gebildet und dienen zur Unterscheidung von Konzepten und aus der bibliothekarischen Tradition heraus zur Erleichterung des Sucheinstiegs, indem Namensvarianten zusammengeführt werden (vgl. Wiechmann, 2014). Normdateien werden landesintern durch nationale Informationsinfrastruktureinrichtungen, meist Nationalbibliotheken, gepflegt. In Deutschland wird die Gemeinsame Normdatei (GND) verwendet. Das Virtual International Authority File (VIAF) führt unterschiedliche nationale Normdateien zusammen.[9] Ein ISO-Standard (ISO-Standard 27729) für die Identifizierung von Personen und Organisationen ist der International Standard Name Identifier (ISNI).[10] Es werden Personen, Körperschaften oder fiktionale Charaktere identifiziert. Ringgold ist ein Unternehmen, welches seit 2003 Identifikatoren speziell für Organisationen vergibt. Das *Identify Database*[11] verzeichnet über 480.000 Organisationen (Stand: 26. Januar 2018), die meisten davon aus Nordamerika (ca. 47 %) und Westeuropa (ca. 25%)[12]. Die *Global Research Identifier Database*[13] (GRID) ist ein von der Firma Digital Science entwickeltes Identifier-System, welches wissenschaftliche Einrichtungen auf der ganzen Welt erfassen will. Entwickelt wurde das System 2015 von Digital Science, um die Firma bei ihren Reporting Aktivitäten zu unterstützen. Der GRID Identifier wird u.a. von Altmetrics und Figshare verwendet. Forschungsförderungseinrichtungen vergeben Forschungsgelder an Projekte, die an wissenschaftlichen Einrichtungen angesiedelt sind. Die DFG verzeichnet in ihrer Datenbank GEPRIS[14] (Geförderte Projekte Informationssystem) Förderprojekte und die durchführenden Personen und Institutionen. Institutionen erhalten innerhalb von GEPRIS eine *DFG ID*. Der Wissenschaftsrat empfiehlt in seinen Empfehlungen zum Kerndatensatz Forschung, die Verwendung der Instituts-ID der DFG und des DAAD, welche über den Research-Explorer[15], dem Verzeichnis der Forschungsstandorte in Deutschland, abgefragt werden kann (vgl. Wissenschaftsrat, 2016, S. 60). Für die Antragstellung und spätere Förderung werden von Institutionen Antragsdokumente ausgefüllt in die verantwortliche Institutionen eingetragen werden müssen. Für Projekte im Zuge des EU-Rahmenprogramm Horizon 2020 ist eine eindeutige Identifikationsnummer

[9] Siehe: `https://viaf.org/` (Abgerufen am 26. Januar 2018).
[10] Siehe: `http://www.isni.org/` (Abgerufen am 26. Januar 2018).
[11] Siehe: `https://www.ringgold.com/ringgold-identifier` (Abgerufen am 26. Januar 2018).
[12] Siehe: `https://support.ringgold.com/wp-content/uploads/2018/01/Ringgold_Infographics_2018_Jan.pdf` (Abgerufen am 26. Januar 2018).
[13] Siehe: `https://grid.ac/` (Abgerufen am 26. Januar 2018).
[14] Siehe: `http://gepris.dfg.de/gepris/OCTOPUS` (Abgerufen am 26. Januar 2018).
[15] Siehe: `http://www.research-explorer.de/research_explorer.de.html` (Abgerufen am 26. Januar 2018).

für wissenschaftliche Einrichtungen für die Antragstellung gefordert. Diesem Teilnehmercode (*Participant Identification Code (PIC)*)[16] hinterliegen alle Angaben zu den Institutionen. Um Forschungsförderer eindeutig zu identifizieren und das Tracking der wissenschaftlichen Ergebnisse zu erleichtern, wurde 2013 durch die DOI-Registrierungsagentur Crossref das Open Funder Registry[17] (ursprünglich FundRef genannt) gegründet. Es handelt sich hierbei um eine normalisierte Liste von Forschungsförderern aus aller Welt, die durch eine DOI identifiziert werden. In Literatur- und Zitationsdatenbanken werden Publikationsdaten sowohl Personen als auch wissenschaftlichen Einrichtungen zugeordnet. Die Datenbank Scopus der Firma Elsevier vergibt jeder Affiliation in ihrer Datenbank eine eindeutige Nummer, die sie *Affiliation-ID* nennen, um zwischen Organisationen zu unterscheiden und Publikationen zusammenzufassen, die zu einer Organisation gehören (vgl. Elsevier, 2018a).

2.3 Projekte

Identifier für Organisationen werden idealerweise zur Vereinfachung von Workflows innerhalb der digitalen Forschungsinfrastruktur verwendet:

> „*Identifiers are useful for disambiguating, consolidating and establishing a hierarchy view. They enforce uniqueness as they disambiguate institutional records, eradicate duplication of data, ensure correct delivery, entitlement and access rights, improve trust in data and map institutions into their hierarchy*" (Amante u. a., 2017, S. 139).

Ein Anwendungsgebiet ergibt sich im Bereich der Zugehörigkeit (Affiliation) von Forschenden zu ihren Einrichtungen. Bereits seit 2008 wird für die vom Verlag Elsevier geführte bibliografische Datenbank *Scopus* ein Affiliation Identifier angeboten, der die Angabe von Zugehörigkeiten normiert und es ermöglicht, alle Publikationen einer Einrichtung aggregiert anzuzeigen (vgl. Haerter, 2008).
Der *Journal Supply Chain Efficiency Improvement Pilot* der National Information Standards Organisation (NISO) identifiziert die Erleichterung von Transaktionsprozessen und Informationsflüssen zwischen Verlagen und Bibliotheken als ein weiteres Anwendungsgebiet für eindeutige Identifier für Organisationen (vgl. Chvatal, 2008). Eine Bewertung von bereits bestehenden Identifikatorensystemen ergab, dass ISNI als einziges System weitestgehend diese Anforderungen erfüllt, jedoch das verwendete Metadatenschema noch nicht vollständig ausgereift ist. Die Working Group gab dementsprechende Empfehlungen an ISNI weiter.

In Großbritannien wurde zwischen 2013 und 2014 eine umfangreiche Studie zur Organizational Identifiern in Großbritannien durch das Joint Information Systems Committee (Jisc) und das

[16] Siehe: `http://www.horizont2020.de/projekt-teilnehmercode.htm` (Abgerufen am 26. Januar 2018).
[17] Siehe: `https://www.crossref.org/services/funder-registry/` (Abgerufen am 26. Januar 2018).

Consortia Advancing Standards in Research Administration Information (CASRAI) durchgeführt (vgl. Hammond und Curtis, 2013). Die Studie ergab, dass keiner der 23 identifizierten und geprüften Identifikatoren als ausreichend „autoritativ" angesehen wird, die Identifikatorensysteme ISNI und UKPRN[18] jedoch am positivsten beurteilt werden. Ein hybrider Ansatz mit ISNI als „*bridging Identifier*" (Brückenidentifier) wird von der Working Group als optimal angesehen (vgl. Brown, 2015).

Aufbauend auf den Ergebnissen des Jisc/CASRAI Projektes evaluierte ein von OCLC geleitetes Projekt die Vergabe von ISNI Nummern anhand unterschiedlicher Anwendungsfälle. Diese bezogen sich auf folgende Akteure innerhalb der wissenschaftlichen Community: Administrative Mitarbeiter*innen, Forschungsförderer, Verlage, Verwaltungen, Forschende und Gruppen von Forschenden, Informationsspezialist*innen Softwareentwickler*innen und Rechtsabteilungen (vgl. Smith-Yoshimura u. a., 2016). Das Projekt zeigt anhand von Szenarien, wie innerhalb ISNI mit Herausforderungen bei der Identifizierung von wissenschaftlichen Einrichtungen umgegangen wird. Diese beinhalten Namensänderungen, genaue Angabe eines bevorzugten Namens, die Abbildung von Hierarchien in den Metadaten mit Hilfe von Beziehungsbeschreibungen, die Restrukturierung von wissenschaftlichen Einrichtungen, Fälle von Übernahmen oder Zusammenlegungen und der Umgang mit Forschungsgruppen und Konsortien. Hieraus ergeben sich Empfehlungen, die das Projekt an ISNI abgeben konnte.
Das Konzept von ISNI als Brückenidentifier ist in Portugal für die Entwicklung eines Identifikators für Organisationen innerhalb des nationalen Forschungsinformationssystems PTCRIS angewendet worden (vgl. Amante u. a., 2017). Der Identifier ist mit ISNI, bzw. auf ISNI gemappte Ringgold-Identifier interoperabel und dient zur Disambiguierung von Organisationen innerhalb von PTCRIS.

Das Projekt THOR befasste sich von 2016 bis 2017 mit der Interoperabilität und Verlinkung von Persistent Identifiern für die wissenschaftliche Infrastruktur (vgl. Fenner u. a., 2016), wobei es sich vor allem auf die Relationen zwischen Identifiern für Forschungsschaffende und Forschungsoutputs konzentrierte. Persistent Identifier für Organisationen werden bislang wenig genutzt. Obwohl es bereits eine Reihe von Identifikatorensystemen gibt, hat sich bisher keines in der wissenschaftlichen Community durchgesetzt:

> „*The use of persistent identifiers for organisations lags behind the use of persistent identifiers for research outputs and people. Despite the work by ISNI, FundRef and others, community uptake is still low. In addition, for some of these organizational identifiers (e.g. FundRef) there is no openly available central service that systematically collects links to other identifiers*" (Fenner u. a., 2015, S. 19).

Das Organization Identifier Projekt verfolgt das Ziel, ein Identifikatorensystem für Organisationen zu schaffen, welches den „*Principles for open scholarly infrastructures*" (vgl. Bilder, Lin

[18] UK Provider Reference Number, beispielsweise verwendet von der Higher Education Statistics Agency (HESA) in der UK. Siehe: `https://www.hesa.ac.uk/support/providers` (Abgerufen am 26. Januar 2018).

und Neylon, 2015) entspricht. Dabei geht es vor allem um das Vertrauen in die Governance, Nachhaltigkeit und die Ansiedlung der Infrastruktur in der Community. Bestehende Organizational Identifier, so die Arbeitsgruppe, seien diesen Prinzipen bisher nicht gewachsen: „*Among the gaps that have been identified during our consultations this year are transparent, non-profit governance and the ability for organizations to manage their own records*" (Cruse, Haak und Pentz, 2016, S. 1). Die Arbeitsgruppe stellte eine Liste von Anforderungen an ideale Organizational Identifier zusammen, wobei sie sich auf die vorangegangenen Studien bezog. Anders als bei diesen Studien liegt hier der Fokus weniger auf dem Verbreitungsgrad der bisherigen Lösungen sondern auf Offenheit (öffentliche API-Schnittstellen), Transparenz (getragen durch gemeinnützige Organisationen) und Nachnutzbarkeit (Daten sind nicht proprietär und liegen unter einer Lizenz vor, die die Nachnutzbarkeit ermöglicht). Die Arbeitsgruppe identifizierte folgende Anforderungen an Organizational Identifier:

- *Unique*: Die Identifier sind global eineindeutig, unabhängig davon wie groß ihre Anzahl ist.
- *Stable*: Die Identifier garantieren Funktionalität.
- *Discoverable*: Die Identifier können leicht gefunden werden.
- *Resolvable*: Die Identifier können im Browser aufgelöst werden.
- *Not recycled*: Im Falle der Obsoleszenz eine identifizierten Objektes wird der Identifier nicht wieder verwendet.
- *Documented*: Die Funktionsweise des Identifiersystems ist offen dokumentiert.
- *Have appropriate metadata*: Die Identifier sind mit für die Identifizierung und Beschreibung wissenschaftlicher Einrichtungen angemessenen Metadaten versehen.
- *Interoperable*: Die Identifier sind mit anderen Identifiern über Metadaten, die ihre Relationen beschreiben verknüpfbar.
- *Can be merged/split*: Die Identifier können im Falle struktureller Veränderungen von Einrichtungen zusammengeführt oder aufgeteilt werden.
- *Expressed as HTTP(S) URIs*: Die Identifier werden als URI, die über das HTTP(S)-Protokoll im Internet ausgeführt werden kann, ausgedrückt.
- *Support content negotiation for machine representations*: Die Identifier unterstützen das System der Inhaltsausgabe[19] um die bestmögliche Darstellung der Informationen zu ermöglichen.
- *Support discovery APIs*: Das Identifiersystem kann über eine Programmierschnittstelle (API) abgefragt werden.
- *Have transparent, non-profit governance*: Die Identifier werden von einer non-profit Organisation mit einer transparenten Verwaltungsstruktur angeboten.
- *Offer the ability for organizations to manage their own records*: Die Identifier und die Metadaten können von den identifizierten Einrichtungen selber verwaltet werden.

Bisher erfüllt kein Identifier diese Anforderungen. Kritikpunkte an der ISNI beziehen sich auf das Business-Model (Kosten für die Benutzung/Registrierung), fehlende offene Lizenzen, fehlender

[19] Siehe hierzu: Holdman und Mutz (1998): RFC 2295. Transparent Content Negotiation in HTTP. `https://tools.ietf.org/html/rfc2295` (Abgerufen am 26. Januar 2018).

Fokus auf Organisationen, fehlende Transparenz im Government und fehlende Nutzbarkeit der Daten in maschinenlesbarer Form für die Öffentlichkeit. Außerdem sieht die Arbeitsgruppe aufgrund von Vertrauen und Nutzbarkeit einen Mehrwert in offenen Identifikatorensystemen (vgl. Bilder, Brown und Demeranville, 2016).

Überblick

Projekt	***Untersuchte Identifier***	**Ergebnis**
I2 Working Group	ISNI, MARC, SAN, DUNS	ISNI mit angepasstem Metadatenschema
Landscape Study	23 Identifikatoren	ISNI
Jisc CASRAI	ISNI, Digital Science (jetzt GRID), Ringgold, UKPRN	ISNI+ als „Brücke" mit angepasstem Metadatschema
OCLC	ISNI	ISNI+ möglich, aber Anpassungen im Metadatenformat notwendig, Nutzung muss verbreitet werden
ORCID	Open Funder Registry, ISNI, Ringgold, Publisher Solutions International, GRID, LEI, OrgRef	Kein offenes Identifikatorensystem vorhanden

Tabelle 1: Bereits durchgeführte Projekte, Untersuchte Identifier und Ergebnisse

3 Methode

Das Hauptziel dieser Arbeit besteht darin, einen Überblick über bisherige Lösungen für die eindeutige Identifizierung von wissenschaftlichen Einrichtung mit Hilfe von Identifikatoren zu schaffen. Dabei werden im ersten Schritt bereits bestehende Identifikatorensysteme betrachtet. Die Auswahl der Identifikatoren erfolgt anhand einer Recherche nach Identifikatoren für wissenschaftliche Einrichtungen und den Beispielen aus den genannten Studien. Der Begriff Identifikator wird dabei als „eindeutige Kennung, die eine auf eine Entität verweist" verwendet. Die Auswahl der Identifikatoren ist zudem auf ihre mögliche Anwendung im deutschsprachigen Raum beschränkt. Dies spiegelt sich unter anderem darin, dass im Zuge der Betrachtung von Normdaten nur die GND ausgewählt wurde. Andere Normdatendateien, wie das *Name Authority File* der Library of Congress in den USA oder das international übergreifende VIAF wurden nicht näher betrachtet. Die GND ID und die ISNI ID sind zwei Identifikatorensysteme in der Auswahl, die nicht ausschließlich wissenschaftliche Einrichtungen beschrieben, sondern deren Anwendungsgebiete auch auf Personen (ISNI, GND) und Orte, sowie Sachverhalte (GND) bezogen werden. Da die GND als Normdatensatz die Grundlage der bibliothekarischen Disambiguierung legt und die ISNI in bereits durchgeführten Studien sehr gut bewertet wurde, werden die beiden Identifier trotzdem betrachtet und nur auf ihre Anwendung für wissenschaftliche Einrichtungen bzw. Körperschaften untersucht.
Zuerst werden bereits verwendeten Identifikatoren kurz beschrieben und darauf hin untersucht, welche Informationen zu den Einrichtungen in welcher Form abgebildet werden, welche Relationen zwischen Einrichtungen untereinander und ggf. zu Personen bestehen können, ob die Identifikatoren im Internet anwendbar sind und in wie fern sie zugänglich und nachnutzbar sind. Dabei werden induktiv Kriterien für Persistent Identifier für wissenschaftliche Einrichtungen abgeleitet.
Innerhalb der bisherigen Studien zu Identifikatoren für wissenschaftliche Einrichtungen wurden Anwendungsszenarien entwickelt. Dafür wurden die besprochenen Anwendungsfälle zunächst Oberthemen zugeordnet, wie im Anhang dargestellt. Die Anwendungsszenarien, die in jeweils mindestens einer Studie besprochen wurden, sind in Tabelle 2 noch einmal dargestellt. Dabei werden die Anwendungsszenarien aus der NISO-Studie nicht berücksichtigt, da die Anwendungsfälle hier sehr spezifisch auf ein fiktives Konsortium zugeschnitten sind. Für den Anwendungsfall der konsistenten und zeitunabhängigen Zuweisung von Affiliationen wird keine Lösung recherchiert, da dieses Anwendungsgebiet im Zuge der Untersuchung der Metadaten aufgegriffen wird. Der Anwendungsfall für das Open-Accesss-Monitoring und APC-Management konnte nur in einem Projekt identifiziert werden. Aufgrund seiner Aktualität und angenommenen Nützlichkeit wird dieser mögliche Einsatz von Identifikatoren für wissenschaftliche Einrichtungen dennoch behandelt. Die Anwendungsfälle werden daraufhin untersucht, welche Lösungen (im deutschsprachigen Raum) im Sinne der eindeutigen Identifizierung von wissenschaftlichen Einrichtungen bisher erfolgt sind. Aufgrund des eingeschränkten Umfangs der Arbeit, kann keine vollständige Erhebung zu Lösungen durchgeführt werden. Stattdessen wird für jeden Anwendungsfall beispielhaft nach Lösungen bzw. bestehenden Workflows recherchiert. Diese werden in Kapitel 5 vorgestellt.

Auswahl der Anwendungsfälle

Kategorie	Landscape Study für Jisc/CASRAI	Jisc/CASRAI	OCLC
Forschende	Drittmittelförderung (Antrag)	Drittmittelförderung (Antrag)	Drittmittelförderung (Gruppen)
Forschungsinfrastruktur	Reporting	Reporting (Forschungsergebnisse/Drittmittel)	Aggregation und Reporting
		Automatisierung von Datenflüssen	Automatisierung von Datenflüssen
Forschungsförderung	Tracking	Tracking	Tracking
		Gutachtenverfahren (Anträge)	Gutachtenverfahren (Anträge)
	Open Access Monitoring und APC-Management		

Tabelle 2: Korrelierende Anwendungsfälle in jeweils mind. 2 Studien

Die Diskussion der Ergebnisse erfolgt anhand von vier Kriterien:

1. Definition von Kriterien für Persistent Identifier für wissenschaftliche Einrichtungen und Kategorisierung Kategorisierung der untersuchten Systeme im Bezug auf diese Kriterien.
2. Anforderungen an Metadaten für PIDs für wissenschaftliche Einrichtungen im Bezug auf die untersuchten Anwendungsfälle.
3. Bewertung der Anwendungsfälle im Bezug auf den Nutzen von Persistent Identifiern.
4. Einschätzungen zu Organisationsansätzen für ein Identifiersystem für wissenschaftliche Einrichtungen im Bezug auf offene Forschungsinfrastrukturen.

Durch die Diskussion werden die Ergebnisse der Analyse der Identifikatoren und die Recherche nach Lösungen zusammengeführt und in einen gemeinsamen Kontext, der Entwicklung eines Persistent Identifiers für wissenschaftliche Einrichtungen basierend auf offenen Forschungsinfrastrukturen gebracht.

4 Identifikatoren für wissenschaftliche Einrichtungen

4.1 Die Gemeinsame Normdatei (GND)

Für Körperschaften entstand in den 1970er Jahren die GKD (Gemeinsame Körperschaftdatei), die aus Körperschaftsdaten der Zeitschriftendatenbank hervorging und 2012 in die Gemeinsame Nordatei (GND) aufging. Körperschaften werden definiert als „*Eine Organisation oder Gruppe von Personen und /oder Organisationen, die mit einer bestimmten Bezeichnung identifiziert wird, und die als Einheit handelt*" (Patton u. a., 2010, S. 27f.). Dazu gehören auch Tagungen, Kongresse, Konferenzen, Messen und weitere kurzzeitige vorhandene Organisationen, die noch bestehen oder in der Vergangenheit bestanden haben.
Innerhalb des Modells der Funktionalen Anforderungen an bibliografische Datensätze (FRBR) werden Körperschaften als separate bibliografische Entitäten behandelt, deren Bezeichnungen mit Normdaten kontrolliert und deren Datensätze mit anderen Datensätzen verlinkt werden können (Ebd., S. 17f.). Beziehungen zwischen Körperschaften sind in FRBR als hierarchische Beziehungen (untergeordnet, übergeordnet) oder (temporale) Folge-Beziehungen (Namensänderungen, gesplittete Körperschaft, fusionierte Körperschaft) modelliert (Ebd., S. 77f.). Kernelemente eines Normdatensatzes für Körperschaften sind nach dem *Resource Description and Access* (RDA) Standards, der seit 2015 für die GND gilt: *Name, bevorzugter Name, Ort, der mit der Körperschaft in Verbindung steht (Ort einer Konferenz, Ort des Hauptsitzes), Datum, das mit der Körperschaft in Verbindung steht (Datum der Konferenz, Gründungsdatum, Auflösungsdatum), in Verbindung stehende Institution (bei Konferenzen), sonstige zur Körperschaft gehörende Kennzeichnung (Art, sonstige Kennzeichnung)* und *Identifikator für die Körperschaft* (*RDA*, S. 545ff.). Mit Identifikatoren sind externe Identifikatoren gemeint, die mit der vergebenden Stelle in den Datensatz übernommen werden. Jeder Datensatz in der GND erhält eine Nummer (Identifikator), die diesen Datensatz eindeutig bezeichnet. Die Fachhochschule Potsdam hat beispielsweise die GND-Nummer 5105272-6.[20] GND-Datensätze können auch im Marc21-XML[21] und RDF (Resource Description Framework)[22] Format heruntergeladen werden. Der RDF-Datensatz zeigt die Properties, die von der GND verzeichnet werden.

```
<http://d-nb.info/gnd/5105272-6> a gndo:CorporateBody ;
foaf:page <https://de.wikipedia.org/wiki/Fachhochschule_Potsdam> ;
gndo:gndIdentifier "5105272-6" ;
gndo:oldAuthorityNumber "(DE-588)4492533-5" ;
owl:sameAs <http://d-nb.info/gnd/4492533-5> ;
dnbt:deprecatedUri "http://d-nb.info/gnd/4492533-5" ;
gndo:oldAuthorityNumber "(DE-588c)4492533-5" , "(DE-588b)5105272-6" ;
gndo:variantNameForTheCorporateBody "FHP","University of Applied Sciences
    Potsdam","FHP (Fachhochschule, Potsdam)" ;
gndo:preferredNameForTheCorporateBody "Fachhochschule Potsdam" ;
```

[20] Siehe: `http://d-nb.info/gnd/5105272-6`.
[21] Machine-Readable Cataloging, ein Datenformat für den Austausch bibliografischer Daten. Siehe: `https://www.loc.gov/marc/bibliographic/` (Abgerufen am 26. Januar 2018).
[22] Ein Datenformat zur Modellierung von semantischen Aussagen (Subjekt-Prädikat-Objekt) im Internet.

```
gndo:gndSubjectCategory <http://d-nb.info/standards/vocab/gnd/gnd-sc#6.6> ;
gndo:geographicAreaCode <http://d-nb.info/standards/vocab/gnd/geographic-area-
    code#XA-DE-BB> ;
gndo:homepage <http://www.fh-potsdam.de> ;
gndo:dateOfEstablishment "1991" ;
gndo:placeOfBusiness <http://d-nb.info/gnd/4046948-7>
```

Im Datensatz sind keine Relationen zu verwandten Organisationen verzeichnet, obwohl diese als „untergeordnet" erfasst sind. Für die Fachhochschule Potsdam sind dies insgesamt 13 untergeordnete Einrichtungen: fünf Fachbereiche, das Usbekisch-Deutsche Zentrum für Architektur und Bauwesen[23], die Gleichstellungsbeauftragte[24], die Potsdam School of Architecture[25], das Informationszentrum für Informationswissenschaft und -praxis (Potsdam)[26] und das Institut für Information und Dokumentation (Potsdam)[27]. Zudem bestehen Datensätze für die vorherige Benennung des Fachbereichs Informationswissenschaften (Fachbereich Archiv-Bibliothek-Dokumentation[28]) und dem Fachbereich Sozialwesen[29]. Der Datensatz des Fachbereichs Archiv-Bibliothek-Dokumentation verweist auf den neuen Datensatz des neu benannten Fachbereichs, und ist im RDF-Datensatz über

```
gndo:succeedingCorporateBody <http://d-nb.info/gnd/10358471-7>
```

definiert. Der Datensatz für den Fachbereich Sozialwesen beinhaltet keine Relationen zu anderen GND-Datensätzen.
Für Organisationen sind in der GND folgende Rollen vorgesehen: „Urheber von", „Beteiligt an" und „Thema in". Körperschaften sind sowohl Thema von Publikationen als auch geistige Schöpfer von Werken.[30] In der GND gibt es keine Relationen zwischen Personennormdaten und Körperschaftsnormdaten, also keine Möglichkeit, Affiliationen auszudrücken. Die Datensätze sind einheitlich über eine URL nach dem Schema <http://dnb.info/gnd/Indentifikationsnummer> aufrufbar. Es gibt kein Resolving System. Der Datensatz ist in Linked Data Anwendungen über die Einbindung der GND-Ontologie vernetzbar aber in den öffentlichen Datensätzen sind keine weiteren externen Identifikatoren eingebunden.

4.2 Der International Standard Name Identifier (ISNI)

Der International Standard Name Identifier (ISNI) identifiziert Personen oder Organisationen, die an der Herstellung und/oder Distribution von medialen Inhalten beteiligt sind (vgl. DIN ISO, 2012). Der Identifikator soll als Brückenidentifikator fungieren und die Nutzung anderer Identifier erleichtern. Daher enthalten die Datensets nur Kerndaten zur Disambiguierung von Identitäten und verlinken auf andere Systeme, die mehr Informationen bereit halten. ISNI hat

[23] Siehe: http://d-nb.info/gnd/1069427608.
[24] Siehe: http://d-nb.info/gnd/10026349-5.
[25] Siehe: http://d-nb.info/gnd/16086296-6.
[26] Siehe: http://d-nb.info/gnd/5105274-X.
[27] Siehe: http://d-nb.info/gnd/5129722-X.
[28] Siehe: http://d-nb.info/gnd/5253999-4.
[29] Siehe: http://d-nb.info/gnd/2160393-5.
[30] Lt. RDA 19.2.1.1.1 (*RDA*, S. 660f.).

zum Ziel, die Auffindbarkeit von Informationsressourcen zu erleichtern, die Disambiguierung von Namen zu unterstützen und eine Infrastruktur zu stellen, die die Informationsdistribution für Organisationen erleichtert (vgl. ISNI, 2018a). ISNI führt die Daten zu Identitäten aus unterschiedlichen Datenbanken aus der ganzen Welt zusammen (vgl. ISNI, 2018b).
Die Kernelemente der Datensätze sind über die Suche in der ISNI Datenbank einsehbar. Angegeben werden *Name, Location/Nationality, Creation Class, Creation Role, Related Names* und *Titles* (Titel von Publikationen). Zudem können *Notes* vergeben werden. Unter *Sources* wird angezeigt aus welchen Quellen die Informationen stammen. Es kann eine Liste der Datenelemente auf der ISNI Website[31] heruntergeladen werden (vgl. ISNI, 2014). In der neusten Version von 2014 sind bereits einige der Vorschläge aus durchgeführten Studien zur Identifikatoren für wissenschaftliche Einrichtungen integriert. Für Organisationen können im MARC-Feld 710$q Angaben zu *organisationType* gemacht werden. Eine Liste von Typen wurde von ISNI definiert. Aus dieser Liste können auch Konsortien, Forschungsgruppen oder Subscription Services ausgewählt werden. Namen werden in Feld 710 $3 mit folgenden Attributen aufgenommen: *legalName, acronym, nickname, assignedName, transliteratedName, difusedName, commonForm (default)*. Zwischen Personen und Organisationen können folgende Relationen definiert werden:

- isRelatedTo
- isMemeberof / hasMember
- isAffiliatedWith / hasEmployee
- contact, no relation oder undefined

Organisationen untereinander können eine Reihe von Relationen haben:

- isMemberOf / hasMember
- isUnitOf / hasUnit
- isSupersededBy / supersedes
- isAffiliatedWith
- isRelatedTo
- formerName
- laterName

In der ISO-Norm für die ISNI ist festgelegt, dass, wenn Namensänderungen von Körperschaften durch Aufteilung oder Zusammenführung geschehen, diese Körperschaften eine neue ISNI-Nummer erhalten (DIN ISO, 2012, S. 5). Jedoch können Namensänderungen über die Relationsattribute *formerName* und *laterName* nachvollzogen werden.
Aus dem Metadatenschema der ISNI wird nicht ersichtlich, welche Rollen eine Körperschaft einnehmen kann. Die Suche in der Datenbank ergibt eine Reihe von Titeln, die mit der Körperschaft assoziiert sind. Da die ISNI vor allem dazu dient, Namen zu disambiguieren, scheint kein großer Wert auf die Auszeichnung von Rollen für Körperschaften gelegt zu werden, sie sind meistens *creator*. Relationen können in der ISNI sowohl zwischen Organisationen untereinander als auch zwischen Personen und Organisationen angegeben werden. Dadurch

[31] Siehe: `http://www.isni.org/content/documents-related-data-submission` (Abgerufen am 26. Januar 2018).

lassen sich Affiliationen darstellen. Auch Namensänderungen für Organisationen lassen sich über Relationen nachvollziehen. ISNIs sind einfache numerische Kennzahlen, bestehend aus 16 Zahlen. Sie enthalten keine semantischen Merkmale.[32] ISNIs werden als persistente URIs ausgedrückt, nach dem Schema <http://isni.org/isni/Identifikationsnummer>.[33] Für den Zugriff ist kein Resolver Service vorgesehen. ISNI-Daten können als MARC21 Code und Linked Data heruntergeladen werden. Die Daten sind unter einer „Open License" lizensiert.[34]

4.3 Ringgold Identifier

Ringgold ist ein Unternehmen, welches Identifikatoren für Einrichtungen aus einem breiten Spektrum von Sektoren, beispielsweise auch Krankenhäuser (Typ *hospital*) oder Verwaltungseinrichtungen (Typ *govt*) anbietet (Delpeuch, 2017). Das Identifiersystem wurde begründet, um eine Best-Practice-Lösung für die Disambiguierung von Organisationen durch Verlage zu schaffen (vgl. Ringgold, 2018d). Dabei ging es vor allem um die korrekte Zuordnung von Transaktionsdaten im Subskriptionsprozess. Das Unternehmen ist eine Registrierungseinrichtung für ISNI. Ringgold IDs werden auf ISNI-IDs gemapped und wenn neue Ringgold IDs erstellt werden, werden diese mit der ISNI Datenbank abgeglichen und ggf. eine neue ISNI hinzugefügt (vgl. Ringgold, 2018b). ORCID verwendet Ringgold IDs für die Angabe von Affiliationen. Die *Identify* Datenbank kann gegen Bezahlung von Organisationen lizensiert werden. Sie bietet Zugang zu den IDs, zu Zusatzinformationen und zu Datenhierarchien. Für einen limitierten Gast-Zugang kann sich jeder registrieren[35], über die Kosten für den vollen Zugriff gibt es auf der Webseite keine genaueren Angaben.
Ringgold IDs sind speziell dafür ausgelegt, Organisationen zu identifizieren. Eine öffentlich zugängliche Version des Schemas der Datensätze gibt es nicht, jedoch werden Elemente der Identify Datenbank dokumentiert.[36] Standard Kernelemente beinhalten *Name, Alternativer Name, Adresse (Ort, Postleitzahl, Region, Land)*. Ringgold kategorisiert Arten von Organisationen nach einem eigenen Standard (vgl. Ringgold, 2018f). Alternative Namen von Organisationen werden mit Hilfe eines Kürzels in verschiedene Arten geordnet, Beispielsweise in *Akronyme, vorherige Namen, Transliterierte Namen, Zusammengelegte Namen*. Organisationen werden nach Sektoren und Typen geordnet. Sektoren sind *academic, corporate, government (central government), hospital, other (non-profit institutions), public (local government), school* oder *consortium*. Diese werden wiederum in Typen bzw. Schwerpunkte aufgeteilt. Somit ist die Typisierung sehr granulär. Ein zweites von Ringgold entwickeltes Klassifizierungsschema ist die Einordnung von Organisationen in Tiers (Ebenen/Stufen) (vgl. Ringgold, 2018e). Auch hier werden Sektoren angegeben und mit dem Scope (Größe/Reichweite) der Organisation gepaart. Diese bleiben äquivalent zu den Ringgold Types, es wird jedoch nicht nach Schwerpunkten sondern nach Größe (anhand der Anzahl von Mitarbeitenden für Wirtschaftliche Einrichtungen) oder Reichweite (anhand der Menge an Forschungsoutput für akademische Einrichtungen) ge-

[32] Zum Beispiel ist die ISNI für die Fachhochschule Potsdam: 0000 0001 0680 6484.
[33] Beispiel für die Fachhochschule Potsdam: `http://isni.org/isni/0000000106806484`.
[34] Siehe: Open License, Version 1.0 (2011): `http://ddata.over-blog.com/xxxyyy/4/37/99/26/licence/Licence-Ouverte-Open-Licence-ENG.pdf` (Abgerufen am 26. Januar 2018).
[35] Siehe: `https://www.ringgold.com/identify-online-guests` (Abgerufen am 26. Januar 2018).
[36] Siehe: `https://support.ringgold.com/` (Abgerufen am 26. Januar 2018).

wichtet. Weitere Tiers, die, wenn Angaben vorhanden sind, in den Metadaten auftauchen, sind Jisc Banding Tier[37] und Carnegie Classification Daten.[38] Die Ringgold Academic Classification (vgl. Ringgold, 2018c) beinhaltet neun Kernelemente für wissenschaftliche Einrichtungen mit zugehörigen festgelegten Attributen:

- *Economic model*: for-profit oder non-profit
- *Governance*: academia, corporate, government, membership (bis zu zwei Attribute sind auswählbar)
- *Reach*: (geografische Reichweite): local, regional, national, internationl
- *Mission*: management, service provision, production or manufacture, advocacy and standards, research, teaching, healthcare (die Attribute beinhalten jeweils noch Unterklassen)
- *Subject interests*: Über 800 Themenkomplexe sind aus einer hierarchisch Strukturierten Subject-Datenbank mit fünf Hierarchieebenen auswählbar. Die erste Ebene besteht aus den Themen Research, Pure Sciences, Mathematics, Applied Sciences, Health, Political Science, Economics, Business and Managament, Law, Social Sciences, Education, Language, Arts and Humanities
- *Description*: Eine Auswahl an 51 Schlagworten, die die Einrichtung verwenden kann, um sich selbst zu beschreiben
- *Level*: 9 hierarchisch geordnete Klassifikationsebenen um Organisationen in hierarchische Beziehungen zueinander einzuordnen. Level 4 als höchste Ebene und Level -4 als unterste Ebene. Nicht für alle Sektoren gelten alle Hierarchieebenen. Im akademischen Bereich befinden sich Universitäten auf dem Level 0, Universitätssysteme können das Level 1 erreichen und Fakultäten, Labore/Research Center und Kurse oder Programme bis auf das Level -4 kategorisiert werden.
- *Activity status*: Aktivitätslevel der Einrichtung, active dormant (nicht operativ aber nicht geschlossen), ceased (geschlossen)
- *Number of Sites*: Anzahl der Standorte, multiple oder single

Konsortien werden in Ringgold als hierarchische Oberelemente behandelt und mit den Mitgliedseinrichtungen verknüpft. Organisationshierarchien werden über Parent- und Child-Relationen angezeigt (vgl. Ringgold, 2018a). Es besteht die Möglichkeit, dass eine Entität (eine Teilorganisation) mehrere Parent-Organisationen haben kann, auch dies wird in den Hierarchiebäumen dargestellt. Zudem lassen sich Kunden wirtschaftlicher Einrichtungen in den Metadaten verzeichnen. Externe Identifier, die in Ringgold eingebunden werden können, sind der OpenAthens Organisation Identifier[39], die CrossRef Funder ID und ISNI. Ringgold ist eine proprietäre Datenbank. Mitglieder können Datensätze als .csv-Dateien herunterladen. Nicht-Mitglieder können über einen Gastzugang nur einen sehr eingeschränkten Einblick bekommen. Auch eine erweiterte Suche in der Identify Datenbank ist nur für zahlende Mitglieder möglich.[40]

[37] Siehe hierzu die Jisc Banding Tiers: `https://www.jisc-collections.ac.uk/Support/Jisc-Banding/HE-bands-by-band1/` (Abgerufen am 26. Januar 2018).

[38] Siehe hierzu die Carnegie Classification of Institutions of Higher Education. `http://carnegieclassifications.iu.edu/` (Abgerufen am 26. Januar 2018).

[39] Siehe hierzu: Organisation Identifiers der OpenAthens Federation `https://login.openathens.net/org-list` (Angerufen am 26. Januar 2018).

[40] Interface Demo: `https://idproduction.ringgold.com/webservicedemo25.html` (Abgerufen am 26. Januar 2018).

Die API-Schnittstelle ist nicht öffentlich. Der Identifier hat keinen Resolving Service. Es gibt keine Lizenzierung der Daten unter offenen Lizenzen. Die Datensätze sind sehr granulär und Disambiguierung kann aufgrund der Menge der Metadaten zuverlässig statt finden.

4.4 Die Global Research Identifier Datenbank (GRID)

Der GRID Identifier soll Akteure in der wissenschaftlichen Community auf drei Arten unterstützen: Genauigkeit und Konsistenz bei der Erfassung von Daten über Forschungseinrichtungen (vgl. Digital Science, Szomszor und Adams, 2017), Datenintegration aus anderen mit GRID verlinkten Identifikationssystemen (vgl. GRID, 2018b) und Disambiguierung von Affiliationsdaten (vgl. GRID, 2018a). GRID verzeichnet über 80.000 Organisationen aus 221 Ländern (vgl. GRID, 2018c). Die GRID Website bietet einen *Disambiguate* Service, der frei zugänglich ist und die Landing Page des Identifiers zeigt. Der gesamte Datensatz ist als .csv-Datei, JSON[41]- und RDF-Datei unter der Lizenz CC0 herunterladbar und nachnutzbar. Der GRID Identifier soll durch seine Persistenz und Eineindeutigkeit diese Systeme unterstützen. GRID IDs werden mit einer Reihe von externen Identifiern verknüpft. Dazu gehören der Wikidata Identifier[42], ISNI, CrossRef Funder ID, UPKN und der OrgRef Identifier[43]. Zudem wird auf den Wikipedia-Eintrag der Institution sowie auf Geonames[44] für die Lokalisierung der Institution verlinkt. GRID soll dazu dienen, mehrere Identifikationssysteme einfach zu verlinken. Der Disambiguate Service bietet automatisierte Disambiguierung über einen Algorithmus, der geografische Verortungen in textuellen Affiliationsdaten erfassen kann, multilinguale Angaben unterstützt und dem ein Mapping von alternativen Namensvariationen zu Grunde liegt. Durch die Einbindung von GeoName-Daten können beispielsweise globale Kollaborationsnetzwerke einfach erfasst werden (vgl. Digital Science, Szomszor und Adams, 2017).
Kernelemente des GRID Identifiers sind *Name, Status (aktiv, inaktiv), Adresse(n) mit Stadt, Region und Ländercode* und *Labels* (der Name in einer anderen Sprache und der dazugehörige Ländercode nach ISO 639-1). Zudem können *Aliases* und *Akronyme* für den Namen, *IP-Adressen*, *URLs*, *Wikipedia-URLs*, *Kontakt E-Mail Adressen* und *Gründungsdaten* hinzugefügt werden. Organisationen werden auch in GRID in Typen kategorisiert. Diese sind: Education, Healthcare, Company, Archive, Non-profit, Government, Facility und Other. Die Angabe des Types ist kein Kernelement, jedoch beinhalten laut Angaben von GRID 96% der Datensätze dieses Element (GRID, 2018c). Relationen von Organisationen untereinander können auf zwei Arten ausgedrückt werden: Hierarchisch, in einer Parent-Child Relation oder als einfache *Related Organisation*. Hierbei werden jeweils der Typ der Relation und der zugehörige Identifier für die verwandte Organisation spezifiziert. Die Angabe des Typs der Organisation hat Einfluss auf das Mapping der Relationen. Für Universitäten (Typ Organisationen) werden Fakultäten, Departments, Campusse oder Lehrstühle immer auf die den Datensätzen der übergeordneten Universität gemapped. Bei Universitäten, die Teil eines Universitätssystems sind, erhält jede

[41] JavaScript Object Notation, ein Datenaustauschformat. Siehe: `https://www.json.org/` (Abgerufen am 26. Januar 2018).
[42] Siehe hierzu: Wikidata. `https://www.wikidata.org/wiki/Wikidata:Main_Page` (Abgerufen am 26. Januar 2018).
[43] Siehe hierzu: OrgRef. `http://www.orgref.org/web/index.htm` (Abgerufen am 26. Januar 2018).
[44] Siehe hierzu: Geonames. `http://www.geonames.org/` (Abgerufen am 26. Januar 2018).

Einrichtung und das übergeordnete System einen eigenen Identifier. Wirtschaftseinrichtungen mit mehreren Standorten werden auf den Identifier für ihren Hauptstandort gemapped. Lehrkrankenhäuser erhalten einen eigenen Identifier unabhängig von der wissenschaftlichen Einrichtung mit der sie assoziiert sind. Externe Identifier werden über Typ und Angabe des Identifiers eingebunden. Adressen werden mit Geoname Identifiern auf den Ebenen Stadt, Region und Land aufgenommen. Zudem werden NUTS (Nomenclature of territorial units for statistics der EU[45] eingebunden. Die GRID Datenbank verfügt noch nicht über eine API-Schnittstelle. Diese ist als kostenpflichtiger Service in Zukunft geplant. Der Identifier bietet keinen Resolver Service. Affiliationen zwischen Personen und Organisationen lassen sich in den Metadaten nicht angeben. Temporale Daten und Namensänderungen sind ebenfalls nicht aufgeführt.

4.5 Identifikatoren der Deutschen Forschungsgemeinschaft (DFG ID)

In der Datenbank GEPRIS der Deutschen Forschungsgemeinschaft kann nach geförderten Projekten, zugehörigen Personen und Institutionen gesucht werden. Institutionen werden in GEPRIS nach folgendem Schema kategorisiert.

- Bundesforschungseinrichtungen
- Frauenhofer-Gesellschaft
- Helmholtz-Gemeinschaft
- Hochschulen ohne Promotionsrecht/Fachhochschulen
- Landesforschungseinrichtungen
- Leibniz-Gemeinschaft
- Max-Planck-Gesellschaft
- Musik- und Kunsthochschulen
- Universitäten

Angaben zu Institutionen beinhalten *Adresse, Telefon, Telefax, E-Mail* und *Homepage*. Zudem werden geförderte Projekte (abgeschlossen und laufend) sowie untergeordnete Institutionen, falls vorhanden, verlinkt. Übergeordnete Institutionen werden nicht angezeigt. Da es keine Dokumentation oder ein öffentlich einsehbares Metadatenschema gibt, können keine weiteren Angaben zu Relationen oder Rollen von Institutionen gemacht werden. Die eigentliche ID ergibt sich aus der URL des Datensatzes. Der Datensatz für die Fachhochschule Potsdam hat die URL `http://gepris.dfg.de/gepris/institution/10372` und somit die ID 10372. Die ID ist zwar eindeutig, jedoch ohne Resolving Service oder transparente Metadaten. Auch im Research Explorer sind keine genauen IDs verzeichnet, sie sind nur über die URL ablesbar. Einrichtungstypen werden zwischen universitären Forschungseinrichtungen, Fachhochschulen sowie außeruniversitären Forschungseinrichtungen des Bundes und der Länder unterschieden (vgl. Research Explorer, 2018). Zudem werden Zentraleinrichtungen und Forschungseinrichtungen als Einheiten innerhalb der Zentraleinrichtungen unterschieden. Fächer, Forschungsbereiche

[45] Siehe hierzu: Nomenclature of territorial units for statistics der EU. `http://ec.europa.eu/eurostat/web/nuts/overview` (Abgerufen am 26. Januar 2018).

und Fachgebiete werden nach dem Fächerschlüssel des statistischen Bundesamts[46] klassifiziert. Es werden sowohl unter- als auch übergeordnete Einrichtungen angezeigt, jedoch ist hier die Systematisierung anders als in GEPRIS. Für die Fachhochschule Potsdam werden als untergeordnete Einrichtungen die jeweiligen Fachbereiche sowie das Institut für Angewandte Forschung Urbane Zukunft angegeben. Dem Fachbereich Architektur und Städtebau sind der Studiengang Kulturarbeit und der Studiengang Studiengang [sic!] Konservierung und Restaurierung untergeordnet. Dem Fachbereich Informationswissenschaften sind keine Fachbereiche, sondern die Professur für Bibliothekswissenschaft untergeordnet.

4.6 Participant Identification Code (PIC)

Die PIC Datenbank kann nicht öffentlich eingesehen werden. Anleitungen zu PIC-Registration (vgl. Léargas, 2018) beschreiben folgende Angaben, die bei der Registrierung gemacht werden müssen: Offizieller Name der Einrichtung (Legal Name), VAT-Nummer, Business Registration Nummer, Website URL, Art der Organisation (non-profit, öffentliche Einrichtung, NG), Geschäftsname, Geschäftsnummer, sowie Datum der Registrierung, Name und Land der Registrierungsinstitution, Land des Sitzes, Amtssprache, Adresse mit Straße, Hausnummer, Postleitzahl, Stadt und Land, Name, Position, Kontaktadressen und Geschlecht einer Kontaktperson, Organisationstyp (Auswahl aus einer vorgegebenen Liste oder „other" und Angaben zur Größe der Organisation (nur für small/medium-sized enterprises (SME) mit weniger als 250 Angestellten). Zur Verifikation müssen beglaubigte Dokumente (z.B. Gründungsurkunde, Steuernummer) und ein ausgefülltes Privacy Statement (FEL) mit Grunddaten (Name(n), Adresse, Umsatzsteuernummer...) auf das Portal hochgeladen werden. Die hinterlegten Daten können während des Validierungsprozesses von der verantwortlichen Person (LEAR) und auch nach der Validierung angepasst werden (vgl. European Commission, 2017a, FAQ ID 245). Teileinrichtungen (beispielsweise Institute an einer Universität) können keine eigene PIC beantragen (vgl. European Commission, 2017a, FAQ ID 250). Sollten Dubletten entstehen, können diese vom Validation Service zu einem „master PIC" zusammen geführt werden (vgl. European Commission, 2017a, FAQ ID 244). Daraus lässt sich folgern, dass keine Daten zu Relationen zwischen Teilen einer Organisation und einer übergeordneten oder verwandten Einrichtung erfasst werden. PIC IDs werden für die gesamte Einrichtung als legale Entitäten vergeben.[47] Der Identifikator verfügt über keine öffentlich zugänglichen Abfragemöglichkeiten außerhalb der Suchfunktion oder einen Resolving Service. Sie sind nicht herunterladbar oder über eine technische Schnittstelle nachnutzbar.

4.7 Scopus Affiliation ID

Scopus vergibt jeder Affiliation in ihrer Datenbank eine eindeutige Nummer, die sie Affiliation-ID nennen, um zwischen Organisationen zu unterscheiden und Publikationen zusammen zu fassen,

[46] Siehe: Statistischen Bundesamt (2015): *Personal an Hochschulen - Fächersystematik*. `http://www.research-explorer.de/includes/PersonalStellenstatistik.pdf` (Abgerufen am 26. Januar 2018).

[47] Die PIC Nummer der Fachhochschule Potsdam ist 994876039.

die zu einer Organisation gehören (vgl. Elsevier, 2018b). Der Identifikator ist eine achtstellige Nummer, nach der in Scopus auch gesucht werden kann. Zudem wird zu jeder ID eine eigene Profilseite, das *Affiliation Profile* angeboten, welche den bevorzugten Namen, die Adresse, die ID und andere Namensschreibweisen angibt. Es werden Daten zu Menge der Dokumente nach Fachgebiet, Kollaborationen mit anderen Einrichtungen und Journals, in denen die meisten Publikationen erschienen sind angezeigt. Außerdem besteht die Möglichkeit, alle Dokumente, alle Autor*innen und alle Patente, die mit dem Identifier verbunden sind anzuzeigen. Es können potentiell passende Affiliationsangaben (*potential affiliation matches*) angezeigt werden. Die Suche nach Affiliationen kann nach Land eingeschränkt werden. Ergebnisse können als .cvs oder zip-Datei exportiert, als E-Mail versendet oder direkt ausgedruckt werden. Bei der Prozessierung der bibliografischen Daten, wird jede Affiliationsangabe sowie angegebene Förderungseinrichtungen aus den Artikeldaten über ein Parser-System extrahiert, gegen OrgDB, einer Datenbank mit Organisationsnamen, geprüft und bei einer Übereinstimmung entsprechend formatiert. Die OrgDB verzeichnet über 70.000 Organisationen weltweit (el Asati, 2016). Wenn Affiliationsangaben nicht auf einen OrgDB Eintrag abgeglichen werden können, werden sie automatisch zu einem Cluster zusammengefügt, um einen potentiellen neuen OrgDB-Eintrag zu veranlassen. Diese Einträge werden manuell durchgeführt. Scopus ist ein proprietäres System, die Profilseiten von Einrichtungen sind nicht offen zugänglich. Es besteht keine Möglichkeit, die Affiliation-IDs herunterzuladen oder ohne eine Scopus Zugangslizenz einzusehen. Es besteht dementsprechend kein Resolver Service und keine Angabe der IDs als eine URI.

4.8 CrossRef Funder ID

Funder Registry der Firma CrossRef wird von Elsevier gesponsert und erhält monatliche Updates. Version 1.14 vom Dezember 2017 beinhaltete 15.918 Forschungsförderungseinrichtungen, von denen 15.551 aktiv waren und 268 nicht mehr operabel oder ersetzt (vgl. Crossref, 2017). Die Daten werden von Verlagen gesammelt. Dies kann entweder über das Einreichsystem für Publikationen geschehen oder über die automatische Extraktion der Namen aus den Manuskripten. Danach werden die Namen der Forschungsförderer mit der Funder ID im Open Funder Registry abgeglichen. Bei der Registrierung von Metadaten für Publikationen in Crossref werden die Namen der Forschungsförderer, die ID und die Bewilligungsnummer (*Grant Number*) mit angegeben. Forschungsförderer können auch unabhängig von Publikationen in der Datenbank registriert werden (vgl. Hendricks, 2015). Die DOI wird aufgelöst zu einer Seite mit Angaben zu der Einrichtung im JSON-Format.
Die Metadaten zu den Fundref IDs werden als RTF-Dateien ausgezeichnet. Die Organisation wird als <skos:Concept>[48] begriffen. Folgende Elemente können angegeben werden:

[48] Simple Knowledge Organisation System (SKOS) ist ein RDF-Schema für kontrollierte Vokabulare. Die W3C Empfehlung ist unter `https://www.w3.org/2009/08/skos-reference/skos.html` einsehbar (Abgerufen am 26. Januar 2018).

rdf:about	DOI der Einrichtung
skosxl:prefLabel	Bevorzugter Name (nsetzungsname) als skosxl:label (URI) und skosxl:literalForm (literaler String) mit Sprachbezeichnung
skosxl:altLabel	Alternative Namensformen als skosxl:Label und skosxl:literalForm
dct:created	Datum der Anlegung des Datensatzes
dct:modified	Datum der Änderung des Datensatzes
dct:isReplacedby	DOI verweist auf neuen Datensatz
fref:status	URI zu 'Deprecated', Einrichtung ist veraltet
svf:fundingBodyType	Typ der Einrichtung, 'pri' (private) oder 'gov' (Government)
svf:fundingBodySubType	Sub-Typus der Einrichtung: academic, corporate, foundation, international, other non-profit, professional associations and societies für private Types, federal/national government und government non-federal für Government Types
svf:country	Herkunftsland als Geonames URI
svf:affilWith	DOI einer zugehörigen Einrichtung
svf:continuationOf	DOI einer vorangegangenen Einrichtung
svf:incorporatedInto	DOI der Einrichtung, in der die benannte Einrichtung eingetragen ist
svf: incorporates	DOI einer Einrichtung, die bei der benannten Einrichtung eingetragen ist
svf:mergedWith	DOI der Einrichtung, mit der die benannte Einrichtung zusammen geschlossen wurde
svf:mergerOf	DOI der Einrichtungen aus deren Zusammenschluss die benannte Einrichtung entstanden ist
svf:renamedAS	neue DOI bei Namensänderung

svf:splitInto	DOIs der Einrichtungen, die aus der Teilung benannten Einrichtung hervorgingen
svf:splitFrom	DOI der vorangegangenen Einrichtung, durch deren Aufteilung die benannte Einrichtung entstanden ist

Die Helmholtz-Gemeinschaft hat folgenden Eintrag im Funder Registry:

```
<skos:Concept rdf:about="http://dx.doi.org/10.13039/501100001656">
     <skos:inScheme rdf:resource="http://data.crossref.org/fundingdata/vocabulary" />
     <skosxl:prefLabel>
        <skosxl:Label rdf:about="http://data.crossref.org/fundingdata/vocabulary/Label-35186">
           <skosxl:literalForm xml:lang="de">Helmholtz-Gemeinschaft</skosxl:literalForm>
        </skosxl:Label>
     </skosxl:prefLabel>
     <skosxl:altLabel>
        <skosxl:Label rdf:about="http://data.crossref.org/fundingdata/vocabulary/Label-6505719">
           <skosxl:literalForm xml:lang="en">Helmholtz Gemeinschaft</skosxl:literalForm>
        </skosxl:Label>
     </skosxl:altLabel>
     <skosxl:altLabel>
        <skosxl:Label rdf:about="http://data.crossref.org/fundingdata/vocabulary/Label-6505720">
           <skosxl:literalForm xml:lang="en">Helmholtz Association</skosxl:literalForm>
        </skosxl:Label>
     </skosxl:altLabel>
     <skosxl:altLabel>
        <skosxl:Label rdf:about="http://data.crossref.org/fundingdata/vocabulary/Label-14525448">
           <skosxl:literalForm xml:lang="en">HGF</skosxl:literalForm>
           <fref:usageFlag rdf:resource="http://data.crossref.org/fundingdata/vocabulary/abbrevName" />
        </skosxl:Label>
     </skosxl:altLabel>
     <dct:modified>2017-08-17T14:55:24.000000</dct:modified>
     <dct:created>2011-09-13T10:11:41.000000</dct:created>
     <svf:fundingBodyType>pri</svf:fundingBodyType>
     <svf:country rdf:resource="http://sws.geonames.org/2921044/" />
     <svf:fundingBodySubType>professional associations and societies</svf:fundingBodySubType>
  </skos:Concept>
```

Innerhalb des Crossref Deposit Schemas werden mindestens der *funder_name* und der *funder_identifier* angegeben. Idealerweise wird noch die *award_number* für das Förderprogramm aus der die Publikation hervorging, angegeben. Unter *funder_name* wird der Name der Einrichtung, der in den Metadaten der Funder ID als <prefLabel><literalForm> angegeben ist, verstanden. Für Verlage, die Funder in ihren Metadaten identifizieren wollen ist es daher von Vorteil, auf die Richtige Ansetzung des Namens zu achten. Das Funder Registry verfügt über eine frei zugängliche API-Schnittstelle und kann als RDF- oder CSV-Datei unter der Lizenz CC0 heruntergeladen werden. Die Funder ID ist als DOI kann über den Browser resolved werden. Die Metadaten sind frei zugänglich und das Schema ist offen dokumentiert.

5 Nutzung von Identifikatoren

5.1 Drittmittelanträge (Forschende)

Für Projektanträge bei der DFG müssen diejenigen Institutionen mit Namen und Anschrift benannt werden, an denen das Projekt durchgeführt wird, sowie Institutionen, die am Projekt beteiligt sind.[49] Die Felder sind freie Textfelder, es bestehen keine Richtlinien dazu, in welcher Form die Institutionen angesetzt werden sollen. Online werden Formulare über das Portal e-lan[50] ausgefüllt. Das System ist nicht offen zugänglich.
Um einen Antrag im Rahmen des Förderprogramm Horizon2020 stellen zu können, müssen sich Organisationen im *beneficiary register* für einen Participant Identification Code registrieren (vgl. Bundesministerium für Bildung und Forschung, 2016). Diese neunstellige Nummer gilt als eindeutiger Identifikator für legale Entitäten (beispielsweise Forschungseinrichtungen, Universitäten u.a.). In der URF-PDM (Unique Registration Facility - Participant Identificaton Code) werden institutionsspezifische Daten zu allen Antragsteller*innen bearbeitet. Über den PIC werden für Antragsteller*innen automatisiert Informationen zu den zugehörigen Organisationen in den Antrag oder Bericht hinzugefügt. Über das PIC Search Portal kann nach Institutionen gesucht werden, um einen eventuell schon vorhandenen PIC herauszufinden. Durchsuchbar sind die Felder Name (auch Teile von Namen) und Land. In der erweiterten Suche können zudem Suchen nach Städten, Registrationsnummern, Umsatzsteuernummern (VAT Number), Erasmus Codes oder PICs eingefügt werden.

5.2 Aggregation und Reporting (Forschungsinfastrukturen)

Um den Forschungsoutput einer Einrichtung sichtbar zu machen, das Berichtswesen (Reporting) zu verbessern, sowie Anfragen von Drittmittelgebern auf Grundlage von gesammelten Daten beantworten zu können, streben wissenschaftliche Einrichtungen an, eine möglichst vollständige Liste aller Publikationen der mit ihr assoziierten Wissenschaftler*innen zu sammeln (vgl. Josenhans, 2011). Dies geschieht traditionell in sogenannten Hochschulbibliografien und vermehrt mit Hilfe von institutionellen Repositorien, sowie in manchen Fällen Forschungsinformationssystemen (FIS). Die Aggregation der bibliografischen Daten zu allen Publikationen ist besonders für größere Einrichtungen eine Herausforderung, da sie häufig noch auf die Meldung der Publikationen durch das wissenschaftliche Personal angewiesen sind. Bibliotheken, die Hochschulbibliographien pflegen, stehen zudem bibliografische Datenbanken wie Google Scholar, Scopus oder Web of Science zur Verfügung. Diese Datenbanken können durchsucht und ausgewertet werden, idealerweise über die Suche nach passenden Affiliationen der Autor*innen.
Google Scholar Citations führte im August 2015 die Möglichkeit ein, nach Autor*innen anhand ihrer Affiliationen zu suchen, sowie anhand von Affiliationen Autor*innen zu finden (vgl. Orduña-Malea u. a., 2015). Das Tool kann zwischen unterschiedlichen Institutionen mit

[49] Bezogen auf den schriftlichen Antrag von 2014. Siehe Dokument 54.011 unter `http://www.dfg.de/foerderung/programme/einzelfoerderung/sachbeihilfe/formulare_merkblaetter/index.jsp` (Abgerufen am 26. Januar 2018).
[50] Siehe: `https://elan.dfg.de/dana-na/auth/url_2/welcome.cgi` (Abgerufen am 26. Januar 2018).

den selben Akronymen unterscheiden und Benennungen von Institutionen in unterschiedlichen Sprachen zusammen führen. Es sind jedoch nicht alle wissenschaftlichen Einrichtungen in Google Scholar erfasst und nicht alle Autor*innen einer korrekten Affiliation zugeordnet, da dies über ein Match zwischen der angegeben Institution und der verifizierten E-mail-Adresse (name@institution.edu) geschieht. Auch die Normalisierung von Namen, die Disambiguierung zwischen Institutionen mit dem selben Namen oder die Zuordnung zu mehreren Institutionen ist oft fehlerhaft (vgl. Orduña-Malea u. a., 2017).
Über die *Organization-Enhanced* Suche kann im Web of Science nach Publikationen einer Einrichtung gesucht werden. Hierbei kann jedoch nur mit Vollständigkeit gerechnet werden, wenn entweder eine normierte Form des Namens oder eine normierte Variante des Namens der Institution in den Metadaten der Publikation angegeben wurde oder wenn alle relevanten Namensvarianten bei der Suche ausgewählt werden. Die Suche nach Publikationen mit Hilfe von Wildcards (Platzhaltern) oder Boolschen Operatoren bietet sich ebenfalls an.
Scopus bietet sowohl die Suche nach Affiliationen als auch eigene Affiliation Profile Pages, auf denen Publikationsdaten, die mit der jeweiligen Affiliation-ID versehen sind, ausgewertet werden. Es werden die Anzahl der Dokumente nach Fachgebiet, Kollaborationen mit anderen Einrichtungen nach Menge der Dokumente und die Menge der Publikationen nach Publikationsquelle (Journal) angezeigt. Zudem können mit der Affiliation verknüpfte Autor*innen abgerufen werden. Es wird eine Liste von potentiell verwandten Einrichtungen angezeigt und diese können mit vorliegender ID gruppiert werden. Bei den Vorschlägen zu verwandten Einrichtungen handelt es sich häufig um die gleiche Einrichtung unter einem anderen Namen (Abkürzungen, andere Sprache, Tippfehler usw.).

Es besteht für registrierte Nutzer*innen die Möglichkeit, einen RSS-Feed für einzelne Affiliation-IDs einzurichten und ihnen zu „folgen". Jedoch sollten Restriktionen einerseits im Scope der bibliografischen Datenbanken[51] und anderseits in der Genauigkeit der Daten beachtet werden. Beispielsweise verzeichnet Scopus allein 27 unterschiedliche Identifier für das GeoForschungs-Zentrum Potsdam.[52] Für eine vollständige Liste von Publikationen sollten diese Identifier unbedingt gruppiert werden.
Für die Nutzung von Daten bibliografischer Datenbanken bietet es sich bisher für Einrichtungen an, auf ihrer Webseite Empfehlungen für die korrekte Angabe des Namens der Institution bei der Einreichung von Publikationen zu machen.[53] Die Hochschulbibliographie der Universität Duisburg-Essen ist mit der Personendatenbank der Universität verlinkt und gleicht Daten zu Affiliationen sowohl über eine interne Datenbank mit internen Identifikatoren als auch über Daten aus Scopus und externen Identifikatoren für Personen (ORCID und Scopus Author ID) ab (vgl. Spielberg, 2017).

[51] Sowohl Scopus als auch das Web of Science decken Journals aus dem STEM-Bereich und aus englischsprachigen und europäischen Ländern breiter ab als geisteswissenschaftliche Publikationen (vgl. Mongeon und Paul-Hus, 2016).

[52] Siehe Anhang.

[53] Wie zum Beispiel für die Universität Potsdam: `http://www.ub.uni-potsdam.de/de/publizieren/universitaetsbibliographie-und-affiliation/affiliation.html` oder Friedrich-Alexander-Universität Erlangen-Nürnberg: `https://www.fau.de/intranet/corporate-identity/leitfaden-affiliation-deutsch-und-englisch/` (Beides abgerufen am 26. Januar 2018).

Forschungsinformationssysteme erfassen Forschungsinformationen, also Metadaten über Forschungsaktivitäten, d.h. Publikationen und publizierte Datensätze, Infrastrukturen und Personen oder Forschungsteams. Diese Informationen werden aus Verwaltungseinrichtungen und Wissenschaft zusammengeführt und führen zu einer Übersicht über Leistungen und Ausstattungen einer wissenschaftlichen Einrichtung. Der Prozess der Erfassung von Forschungsinformationen wird als Forschungsdokumentation bezeichnet (vgl. Ebert u. a., 2016). Die Informationen können als Grundlage für die Berichterstattung (Reporting) einer wissenschaftlichen Einrichtung verwendet werden. Integrierte Forschungsinformationssysteme fassen Forschungsinformationen aus unterschiedlichen Quellen zusammen. Dies sind zum Beispielen Personendaten, Publikationsmetadaten oder Daten zu Fördermitteln. Identifier spielen eine verbindende Rolle zwischen verteilten Informationsquellen und helfen bei der eindeutigen Zuordnung von Informationen sowie bei der Gestaltung von interoperablen Schnittstellen für die Integration von externen Datenquellen (vgl. Jörg, Höllrigl und Sicilia, 2012). Es gibt eine Reihe von proprietären Anbietern für Forschungsinformationssysteme[54] sowie unterschiedliche interne Lösungen[55], zum Beispiel VIVO oder DSpace CRIS als Open Source Software-Lösungen (vgl. Herwig und Schlattmann, 2016). Ein gemeinsamer Datenstandard um die Interoperabilität der unterschiedlichen Systeme zu ermöglichen wird angestrebt. Hierbei handelt es sich um das Common European Research Information Format (CERIF).[56] Ein Standard für die Nutzung von (nicht internen) Identifikatoren besteht bisher nicht. Das Converis System der Friedrich-Alexander-Universität Erlangen integriert Affiliationsdaten aus der Datenbank des Web of Science (vgl. Walther und Melsheimer, 2016) und es gibt Vorschläge für eine Integration von ISNI in VIVO (vgl. Wang, Smith-Yoshimura und Gatenby, 2015).

5.3 Automatisierung von Datenflüssen (Forschungsinfrastrukturen)

Datenflüsse, bzw. Metadatenflüsse zu wissenschaftlichem Output von Forschenden sind nicht nur bei der quantitativen Aggregation, sondern auch bei der Population von Repositorien, die zur Auffindbarkeit, Verbreitung und Wiederverwendung des wissenschaftlichen Outputs beitragen, von Bedeutung. Das Projekt *DeepGreen* hat zum Ziel, Publikationen, die in Allianz-Lizenz lizenzierten Zeitschriften erschienen sind, nach ihrer Embargofrist offen (im Sinne von Open Access) zugänglich zu machen. Dies soll automatisiert über Schnittstellen, einen designierten Workflow und in Kooperation mit Verlagen und (institutionellen) Repositorien geschehen (vgl. Putnings und Rusch, 2016). Es wurde ein Metadatenschema entwickelt, um Schnittstellen und Ablieferungsverfahren aus der Perspektive von Verlagen, Bibliotheken und Repositorienbetreiber sowie Repositoriensoftware zu harmonisieren (vgl. Becker u. a., 2016a). Das Metadatenschema enthält Angaben zur Affiliation für die Elemente <*Creator*> und <*Contributer*> sowie Angaben zum <*Funder*>. Namensangaben (als Freitext) sind vorgeschrieben, zusätzliche Elemente sind alternative Namen, Initialen, Adress- und Kontaktdaten, Affiliation History und Persistent Identifier. Für <Funder> können zusätzliche

[54] Pure (Elsevier), Converis (Thomas Reuters), Sympletic Elements (Digital Science).
[55] Beispielsweise PUB der Universität Bielefeld.
[56] Siehe: `https://www.eurocris.org/cerif/main-features-cerif` (Abgerufen am 26. Januar 2018).

Angaben zu Funder und Projektrelationen gemacht werden (funding area, funding programm, funding programm identifier, aknowledgement set by funder, grant number). Für Freitextfelder werden kontrollierte Vokabulare (Normdaten) angegeben: LCNAF (Library of Congress Name Authority File), ULAN (Union List of Artist Names), GND und VIAF. Zudem wird eine Liste von Identifiern vorgegeben:

- ISNI
- Ringgold Identifier
- Scops Affiliation ID
- GND ID
- LCNAF ID
- ULAN ID
- VIAF ID
- Fundref ID (für Funder)

Die Angaben sind auf unterschiedliche Metadatenformate gemappt (Dublin Core, MARC, ONIX, JATS und OPUS Import-XML). Angaben des Identifiers sind in keinem Fall vorgeschriebene Elemente.
Zur Entwicklung des Metadatenschemas wurde eine Umfrage unter Verlagen zu Praktiken und Prozessen des Metadaten- und Dokumentenaustauschs durchgeführt (vgl. Becker u. a., 2016b). Dabei wurde erfragt, ob Metadatenfelder für Affiliationen unterstützt werden[57], zum Typ des benutzten Identifiers (Name, Ringgold oder andere)[58], zur Verlässlichkeit und Scope der Affiliationsangabe[59] und zu Methoden, wie Affiliationen gesucht werden.[60]

5.4 Tracking (Forschungsförderung)

Die DFG verfügt über ein eigenes Forschungsinformationssystem, welches seit 1980 Daten zu Forschungsanträgen sammelt. Seit 2005 werden Daten zu Anträgen über ein elektronisches Antragsbearbeitungssystem (ElektrA) prozessiert (vgl. Hahnen und Güdler, 2008). Das System wird vorrangig genutzt, um den Schriftverkehr mit Antragsteller*innen und Gutachter*innen zu führen und interne Dokumente für die entscheidenden Gremien zu produzieren. Ein zweites Anwendungsgebiet ist, das Förderhandeln der DFG auf eine empirische Basis zu grundieren. Statistische Informationen, Evaluationsstudien, Online-Informationssysteme und Monitoring-Programme wie der „Förderatlas" sind auf valide Datengrundlagen angewiesen (vgl. Güdler, 2016). In der ElektrA Datenbank werden über 230.000 Personendaten, mehr als 620.000 Antragsdaten und 23.000 universitäre, 5.400 außeruniversitäre und 22.000 ausländische Institutionsdaten gespeichert. Das Institutional Repository der DFG (IRD) verwaltet Daten zu Institutionen. Es werden Adressdaten gesammelt, sowie Angaben zu

[57] E 20, Frage 4.4: Besides common metadata fields (article title, author, etc.), do you support the following data fields?

[58] E 25: Please indicate your communly used affiliation identifier(s).

[59] E26: Is there a reliable allocation between all authors and their affiliations? Please tell us if there are any limitations.

[60] E27: How do you record or query the affiliation name (e. g. standardised string; if not standardised: what possibilities are available to get a comprehensive coverage of the institutional name(s) and its abbreviations)?

Fachschwerpunkten. Die Institute werden zudem in die Hierarchie ihrer Dachorganisationen eingeordnet (vgl. Hahnen und Güdler, 2008). Berichte wie der DFG-Hochschulreport aber auch GEPPRIS und die Research Explorer Datenbank beruhen auf diesen Daten.
Auf dem Portal GEPRIS, welches als Förderdatenbank DFG-geförderte Projekte verzeichnet, kann nach Institutionen gesucht werden. GEPRIS erfasst Projekte seit Anfang der 2000er Jahre (vgl. Deutsche Forschungsgemeinschaft, 2018). Es informiert über inhaltliche Angaben zu Projekten, fachliche Zuordnungen und über beteiligte Personen und Forschungseinrichtungen. Zudem wird die eindeutige Kennungen zu Projekten angegeben. Bei der Suche nach IDs für das Deutsche Luft- und Raumfahrzentrum und seinen Instituten fielen Dubletten und Inkonsistenzen in den Datensätzen auf. So existiert sowohl ein Datensatz für DLR - Institut für Strömungsmechanik, Göttingen (ID 20685784)[61] als auch für DLR Göttingen Institut für Strömungsmechanik (ID 20728648)[62] mit identischen Adressangaben aber unterschiedlichen verknüpften Projekten.

Wie in GEPRIS lassen sich H2020-Projekte über eine Förderdatenbank recherchieren. Das *Commision database of EU-funded research and innovation projects* (CORDIS) verzeichnet alle geförderten Projekte. Für Deutschland werden 66.302 Projekte gelistet. Die H2020 Projekte können über das EU Open Data Portal heruntergeladen werden.[63] Sie sind aufgeteilt in Projekte und Organisationsdaten. Das Datenset für Organisationen beinhaltet die Projektnummer und -ID, das Projekt-Akronym, die Rolle der Institution (participant, coordinator, hostInstitution, benificiary), die PIC der Organisation, Namen und Abkürzung der Organisation, activityType, Status der Partizipation (endOfParticipation true oder false), ecContribution, Land, Straße, Stadt, Postleitzahl, URL, contactType und Kontaktangaben, falls es eine Ansprechperson gibt (vgl. Publications Office, 2017). Trotz der eindeutigen Identifizierung der Organisationen über die PIC ist der Datensatz nicht „sauber": Adressdaten sind den falschen Feldern zugeordnet (Straßenname und Name der Stadt sind vertauscht oder Städtenamen nicht normalisiert.[64] Die Daten stehen unter dem Copyright der Europäischen Union, können aber unter Nennung des*der Urhebers*in nach genutzt werden (vgl. European Commission, 2017b).
Crossref bietet eine Suche nach Publikationen, die durch eine bestimmte Forschungsförderungseinrichtung gefördert wurden.[65] Der Name der Förderorganisation wird in das Suchfeld in natürlicher Sprache eingegeben. Es werden Vorschläge gemacht, die zum eingegeben Suchbegriff passen. Es muss einer der Vorschläge ausgewählt werden. Die Liste der gefundenen Publikationen kann nach Typ, Jahr, Journal, Kategorie (Fachgebiet) und Verlag gefiltert werden. Die Suchergebnisse zeigen den Titel der Publikation, den Typ, das Datum, den Namen des Journals sowie Jahrgang/Band und Seitenzahl, den Namen der Forschungsförderungseinrichtung, ggf. mit Grant Nummer, sowie Autor*innen und die DOI der Publikation als Link. Die eigentliche Funder-ID wird nicht angezeigt. Um sie herauszufinden, muss entweder das

[61] `http://gepris.dfg.de/gepris/institution/20685784`.
[62] `http://gepris.dfg.de/gepris/institution/20728648`.
[63] Datensatz für die Projekte 2014-2020: `https://data.europa.eu/euodp/en/data/dataset/cordisH2020projects` (Abgerufen am 26. Januar 2018).
[64] Beispiele: MUNCHEN/MUENCHEN, KOELN/KOLN, FRANKFURT AM MAIN/FRANKFURT (MAIN)
[65] Siehe: `https://search.crossref.org/funding` (Abgerufen am 26. Januar 2018).

Registry heruntergeladen werden oder ein Service wie der CrossRef Funder Registry Browser[66] verwendet werden. Zudem kann über die API Schnittstelle nach der Funder ID gesucht werden. Über die offene API-Schnittstelle können Abfragen nach Publikationen, die mit der Funder-ID verknüpft sind gestellt werden.[67]

5.5 Gutachtenverfahren für Anträge (Forschungsförderung)

Anträge an die DFG werden in einem zweistufigen Verfahren begutachtet: Die Anträge werden in einem ersten Schritt von fachlich qualifizierten freiwilligen Gutachter*innen auf inhaltliche Relevanz geprüft. Danach entscheiden die sogenannten Fachkollegiaten, die aus gewählten Vertreter*innen der Fachgebiete bestehen, über die Qualität der Begutachtungen und formulieren die Förderungsempfehlungen (vgl. Hornbostel und Olbrecht, 2007). Die Fachkollegiaten werden für eine Amtszeit von vier Jahren aus einer Kanditat*innen liste gewählt. Diese wird beruhend auf Vorschlägen aus der wissenschaftlichen Community vom Senat der DFG zusammen gestellt. Die Erstgutachter*innen werden von der DFG aufgrund ihrer fachlichen Expertise ausgewählt. Um eine mögliche Befangenheit der Gutachter*innen zu vermeiden, bestehen eine Reihe von Kriterien, die ggf. zum Ausschluss der Gutachter*innentätigkeit für einen Antrag führen (vgl. Deutsche Forschungsgemeinschaft, 2014). Dabei stehen neben verwandtschaftlichen Verhältnissen (Punkt 1), vorangegangene wissenschaftlichen Kooperationen (Punkt 3), persönlichen wirtschaftlichen Interessen (Punkt 2) und Beteiligungen an Berufungsverfahren in den letzten 12 Monaten (Punkt 14) auch dienstliche Abhängigkeiten oder Betreuungsverhältnisse in den letzten sechs Jahren (Punkt 5) und Zugehörigkeiten zu derselben oder einer beteiligten Einrichtung (Punkt 6) zur Debatte. Ob und inwieweit Identifikatoren zur Bestimmung von Zugehörigkeiten verwendet werden, ist nicht öffentlich verzeichnet.
Auch die Begutachtung von Anträgen in H2020 erfolgt über ein Peer Review Verfahren. Hierfür wurde eine Datenbank eingerichtet[68], in der sich Personen registrieren können. Befangenheitskriterien sind ähnlich wie für Gutachter*innen der DFG. Zugehörigkeiten zu der selben Einrichtung sowie persönliche Verhältnisse und wissenschaftliche Kollaborationen sowie Betreuungsverhältnisse führen zu einem Ausschluss aus dem Gutachterverfahren (vgl. European Commission, 2017c, S. 43). Bei Zugehörigkeiten können jedoch Ausnahmen entstehen, wenn der*die Expert*in in einem anderen Department oder Institut an der selben Einrichtung arbeitet, wenn die Departments oder Institute innerhalb der Einrichtung sehr unabhängig arbeiten oder wenn es keine bessere*n Kandidat*innen aufgrund einer kleinen Auswahl gibt. Expert*innennamen werden mit Abschluss eines Jahres öffentlich gemacht.[69] Aus der Liste ist ersichtlich, dass die Felder „Most Recent Current Employer Name", „Employer City" und „Employer Country" belegt sind. Hinweise auf die Nutzung eines Identifiers (beispielsweise die PIC) existieren nicht.

[66] Siehe: `https://www.e-sciencecentral.org/funder/` (Abgerufen am 26. Januar 2018).
[67] Beispielabfrage für die Helmholtz-Gemeinschaft: `https://api.crossref.org/funders/http://dx.doi.org/10.13039/501100001656/works`.
[68] Siehe: `http://ec.europa.eu/research/participants/portal/desktop/en/experts/index.html` (Abgerufen am 26. Januar 2018).
[69] Siehe: `http://ec.europa.eu/research/participants/portal/desktop/en/funding/reference_docs.html` (Abgerufen am 26. Januar 2018).

5.6 Open-Access-Monitoring und APC-Management

Unter Open-Access-Monitoring wird die Messung des Anteils der in Open Access (OA) erschienen Publikationen verstanden. Open-Access-Monitoring Methoden bleiben nicht ohne Fehlerquellen, hierzu zählen unter anderem Probleme bei der genauen Bestimmung des Begriffs „Publikation", die Unterscheidung von Gold, Green und Hybrid Open Access und Probleme bei der Indexierung in den Quellen zur Datengrundlage (vgl. Putnings, 2017). Erhebungen zum Stand von Open Access werden häufig auf Landesebene durchgeführt.[70] In Deutschland wurde für die Bundesländer Berlin und Brandenburg der Anteil an OA-Publikationen in den Jahren 2013-2015 erhoben (vgl. Voigt und Winterhalter, 2016). Neben dem Anteil an OA-Publikationen werden im Open-Access-Monitoring auch Kosten für Article Proccessing Charges (APC) gemessen. In Deutschland arbeitet das DFG-geförderte Projekt INTACT[71] an der Offenlegung von APC-Kosten und der Analyse der Verbreitung von OA-Publikationen in Deutschland. Außerdem werden innerhalb der Open APC Initiative[72] offene Daten zu Publikationskosten gesammelt und zur Verfügung gestellt. In der 2016 erschienen Open-Access-Strategie des Bundesministeriums für Bildung und Forschung kündigte das Ministerium an, einen nationalen Open-Access-Monitor etablieren zu wollen (vgl. Bundesministerium für Bildung und Forschung, 2016, S. 12). Persistent Identifier spielen für das Monitoring von Open-Access-Publikationen bei der Datenerhebung eine herausgehobene Rolle. Eine Methode der Datenerhebung für die Auswertung des OA-Anteils beruht auf der Abfrage von bibliografischen Datenbanken, der Normalisierung und des Abgleichs von OA-Zeitschriften mit dem Directory of Open Access Journals (DOAJ) mittels ISSN oder eISSN. Hybride Publikationen[73] werden Mittels ihrer DOI über die Crossref API-Schnittstelle auf ihren Access Status geprüft (vgl. Voigt und Winterhalter, 2016). Wenn für Institutionen das Aufkommen der OA-Publikationen anhand der oben beschriebenen Vorgehensweise untersucht werden soll, muss in der Abfrage der Daten, wie in diesem verwendeten Python-Skript, die Einrichtung so angesetzt werden, dass alle möglicherweise verwendeten Schreibweisen in den bibliografischen Datenbanken berücksichtigt werden können. Für die Technische Universität Berlin sind folgende Namensvarianten vorgeschlagen (Bunge und Voigt, 2016, S. 8):

```
TU.nameVar1 = [['Technische Universitat Berlin'],
               ['Technische Universitaet Berlin'],
               ['Technische Universität Berlin'],
               ['Berlin Institute of Techn'],
               ['Tech Univ Berlin'],
               ['Berlin Univ Technol'],
               ['Univ Technol Berlin'],
               ['TU Berlin'],
               ['Tech. Univ. Berlin'],
```

[70] Beispiele: Open-Access-Indikator in Dänemark: http://www.forskningsdatabasen.dk/en, Meassuring Open Access in Finnland: http://septentrio.uit.no/index.php/SCS/article/viewFile/4276/3960 (Beide Quellen abgerufen am 26. Januar 2018).

[71] Siehe: https://www.intact-project.org/ (Abgerufen am 26. Januar 2018).

[72] Siehe: https://github.com/OpenAPC/openapc-de (Abgerufen am 26. Januar 2018).

[73] Open Access publizierte Artikel in Journals, die nicht ausschließlich Open Access publizieren.

```
['Berlin Inst Technol'],
['Technical University Berlin'],
['Technische Universitaet de Berlin'],
['Technical University of Berlin'],
['Berlin University of Technology']]
```

Durch das Monitoring von OA-Publikationen kann nicht nur das Fortschreiten der „Open-Access-Transformation", also die Abkehr des wissenschaftlichen Publikationsmarktes von subskriptionsbasierten Zeitschriften hin zu Open-Access-Zeitschriften, sondern auch institutionsintern der wahrscheinliche Bedarf an Mitteln für die Finanzierung von Publikationsgebühren ermittelt werden (vgl. Bruch u. a., 2016). Um das Management der APC-Kosten von Seiten der Verlage und der Publikationsfonds, die viele Bibliotheken an wissenschaftlichen Einrichtungen zur Verfügung stellen, zu optimieren, müssen stabile Workflows und Geschäftsprozesse entwickelt werden.[74] Dafür ist bei der Einreichung von wissenschaftlichen Publikationen die korrekte Angabe der Affiliation der Autor*innen besonders wichtig, um die korrekte Kostenstelle zu identifizieren und die Qualifizierung der Publikation für die Kostenübernahme zu prüfen. Diese Angaben sind häufig in einem Freitextfeld einzutragen und bieten somit eine Fehlerquelle. Wird die Zugehörigkeit der Autor*innen nicht richtig angegeben, kann eine zentrale Kostenübernahme der APC-Gebühren nicht ohne weiteres durchgeführt werden (vgl. Geschuhn und Sikora, 2015). Die Erstattung von Publikationskosten durch einen Publikationsfond ist bisher mit einem großen Verwaltungsaufwand verbunden. Die Normierung der Daten zu Autor*innen und Affiliationen, die zentralisierte Speicherung der Daten, beispielsweise in einer Veröffentlichungsdatenbank, und eine Verbesserung der Rechnungsdaten könnten diesen Aufwand schmälern (vgl. Frick, 2017). Die Angabe der ORCID iD bei der Einreichung von Manuskripten ist eine Lösung für die Normierung von Angaben zu Autor*innen. Verlage wie PLOS, eLife, Springer Nature oder Wiley haben sich bereits dafür ausgesprochen, ORCID iDs innerhalb ihrer Einreichungssysteme zu verlangen (vgl. Haak, 2015). Eine Nutzung von PID für wissenschaftliche Einrichtungen für die Transaktionsprozesse und die Abrechnungen zwischen Verlagen und wissenschaftlichen Einrichtungen ist bisher nicht bekannt.

[74] Die Vision eines idealen Publikationsprozesses von OA-Artikeln ist beschrieben in Vierkant u. a., 2017.

6 Diskussion

6.1 Einordnung der Identifikatorensysteme

Identifikatoren können in unterschiedlichen Formen bestehen. Einfache Identifier identifizieren eine bestimmte Entität im Kontext ihrer Nutzung. Sie fungieren als Benennung des Objektes, meistens in Form eines Zahlenstrings. Eindeutige Identifier (Unique Identifier) stellen sicher, dass eine Disambiguierung zwischen unterschiedlichen Entitäten stattfindet. Jede Entität erhält einen eindeutigen Identifier, ohne dass Dopplungen von Identifiern oder Dopplungen in der Vergabe vorkommen. Persistent Identifier besitzen die selben Funktionen wie eindeutige Identifier, haben jedoch zusätzliche Eigenschaften: Sie sind für die Nutzung im Internet geschaffen und daher darstellbar als HTTP(S) URI und ausführbar durch das Resolving im Browser. Außerdem sind sie angereichert mit Metadaten und tragen somit zusätzliche Informationen über das identifizierte Objekt bei. Dabei verweisen sie eindeutig auf eine oder mehrere Repräsentation des Objektes und seiner Metadaten, der Landing Page. Sie sind sowohl für Maschinen ausführbar als auch für Menschen verständlich und persistent in dem Sinne, dass sie weiter bestehen, sobald ein Objekt nicht mehr existiert. Die folgende Tabelle verdeutlicht die Unterschiede zwischen den verschiedenen Erscheinungsformen von Identifikatoren.

Einfache Identifier	Eindeutige Identifier	Persistent Identifier
Identifizierung	Identifizierung	Identifizierung
	Disambiguierung	Disambiguierung
		HTTP(S) URI
		Resolving Service
		Metadaten und Services

Tabelle 3: Unterschiede zwischen einfachen, eindeutigen und Persistent Identifiern

Eindeutige Identifier können auch im Sinne ihrer Langlebigkeit persistent sein. Für die Funktion von internen Datenbanken ist es von Vorteil, sowohl eindeutige als auch gleichbleibende Identifier zu vergeben. Jedoch fehlt ihnen das Element der sogenannten „*actionability*" im Internet, also des automatisierten Abrufen von Informationen über das identifizierte Objekt und der Auflösbarkeit im Browser. In relationalen Datenbanken fungieren IDs als Primärschlüssel, die auch nur in der Datenbank das Objekt referenzieren. Außerhalb der Datenbanken kann der String nicht für den Abruf von Informationen verwendet werden. Beispielsweise hat die DFG-ID 10372 nur innerhalb der Datenbank der DFG die Bedeutung „Fachhochschule Potsdam". Ein außenstehendes System könnte die Nummer nur verwenden,

wenn das System die selben Relationstypen wie die DFG benutzen würde oder wenn die ID über ein Mapping auf das außenstehende System verfügen würde. Für die Bewertung der Identifier wird diese Funktion Scope genannt. Damit ist die Nutzungsmöglichkeit und Ausführbarkeit des Identifiers außerhalb der Datenbanken, in denen er vergeben wird, gemeint. Alle in dieser Arbeit betrachteten Identifier sind eindeutige IDs, sonst wären sie für die Disambiguierung von Namen nicht einsetzbar. Jedoch erfüllt keiner der Identifier, außer der Crossref Funder ID, die Kriterien zu Resolving Systemen und der Möglichkeit zur automatisierten Abrufung von Metadaten. Um Persistent Identifier global im Internet ausführbar zu machen, sollten sie in ihrer Syntax auf einen Namespace oder einen Resolving Service verweisen. Namespaces sorgen zudem dafür, dass direkt erkennbar ist, in welchem Kontext der Identifier auf welche Weise aufzulösen ist. Die Auflösung erfolgt dann über den referenzierten Resolving Service, der die Standorte des eigentlichen Objektes bereit hält.

Besonders wichtig ist dieser Resolving Service bei genuin digitalen Objekten, die sich im Internet befinden. Bei der Identifikation von wissenschaftlichen Einrichtungen besteht dieses Problem nicht unbedingt, da die Einrichtung selber an einem analogen Ort auf der Welt existiert. Jedoch löst die Anwendung von Resolving Funktionen das Problem der verminderten Reichweite eines Identifiers. Für ein Identifier Registry, ähnlich des ORCID Registry, würde für wissenschaftliche Einrichtungen eine Landing Page mit Informationen über die Einrichtung über den PID erreicht werden, die als digitale Repräsentation der Einrichtung angesehen wird. Interoperabilität zwischen Identifiern führt dazu, dass Akteur*innen und Produkte innerhalb der Wissenschaft auf eine eindeutige und maschinenlesbare Weise miteinander verbunden sind. Informationsflüsse innerhalb der Wissenschaft können so vereinfacht, automatisiert und beschleunigt werden. Dies geschieht über die Verlinkung unterschiedlicher Identifier Systeme, wie es im Projekt THOR untersucht wurde.

Kriterien

Wenn wissenschaftliche Einrichtungen mit Persistent Identifiern versehen werden, ergeben sich unterschiedliche Anwendungsfälle und Grundeigenschaften, die ein PID System für wissenschaftliche Einrichtungen haben sollte. Angelehnt an die Anforderungen für Persistent Identifier der Deutschen Nationalbibliothek (vgl. Schroeder, 2009) werden folgende Grundkriterien für Persistent Identifier definiert:

- **Eindeutigkeit**: Jeder Identifier bezeichnet nur eine Entität, in diesem Falle nur eine Einrichtung oder einen Teil der Einrichtung. Es bestehen keine Dopplungen, Identifier werden nur ein Mal vergeben und nicht wieder verwendet. Es besteht eine eindeutige Beziehung zwischen der bezeichneten Entität und dem Identifier.
- **Persistenz**: Der Identifier bezeichnet die Einrichtung auf unbestimmte Zeit. Wenn sich die Einrichtung ändert oder nicht mehr existiert, besteht der Identifier weiter. Verändert sich nur die Bezeichnung der Einrichtung bleibt der Identifier gleich. Bei Änderungen innerhalb der Struktur von Einrichtungen wird ein neuer Identifikator vergeben, auf den der ursprüngliche Identifikator verweist.
- **Resolving Funktion**: Der Identifier wird als permanenter Link ausgegeben, der auf

eine Landing Page verweist, die die Repräsentation der Einrichtung im digitalen Raum darstellt.
- **Interoperabel**: Der Identifier kann in seinen Metadaten mit anderen Identifiern verknüpft werden. Es können Services aufbauend auf diesen Verknüpfungen angeboten werden.
- **Scope**: Der Identifier ist auch außerhalb der namensgebenden Datenbanken nutzbar und ausführbar.

Auswertung

	Eindeutigkeit	Persistenz	Resolving	Interoperabel	Scope
GND ID	✓	✓	nur URL	X	intern
DFG ID	✓	✓	X	X	intern
PIC	✓	✓	X	X	intern
ISNI	✓	✓	nur URL	✓	auch extern
Ringgold	✓	✓	X	teils	intern
GRID	✓	✓	X	✓	auch extern
Crossref Funder ID	✓	✓	✓	✓	extern
Scopus Affiliation ID	✓	✓	X	X	intern

Tabelle 4: Bewertung der untersuchten Identifikatoren

Von allen Identifikatoren erfüllt nur die CrossRef Funder ID alle Kriterien. Wie bereits oben erwähnt, geht die hier verwendete Definition eines PID Systems über das der einfachen Vergabe einer eindeutigen Kennziffer hinaus. Aufbauend auf dieser Definition und oben genannten Kriterien für Persistent Identifier konnte für wissenschaftliche Einrichtungen außerhalb der Crossref Funder ID kein PID System identifiziert werden.

Die GND ID verfügt zwar über die Funktion der Disambiguierung und der angenommenen Persistenz, jedoch findet kein Resolving statt, der Datensatz ist über eine URL erreichbar. Normdatensätze bieten für die Nutzung von Bibliothekskatalogen den Vorteil, dass

unterschiedliche Versionen eines Namens so zusammengefasst werden, dass alle zugehörigen Dokumente gefunden werden können. Es geht dabei weniger um die Nachvollziehbarkeit von Entwicklungen des Namens oder der Körperschaft, auf die der Name verweist, sondern um eine einfache Disambiguierung sowie Zusammenfassung unterschiedlicher Benennungen unter einer Hauptbenennung. Jedoch ist dies außerhalb von bibliothekarischen Kontexten der Erfassung und die Zugangsschaffung zu Literatur für die Beschreibung von wissenschaftlichen Einrichtungen nicht ausreichend.

Die DFG ID distinguiert unterschiedliche wissenschaftliche Einrichtungen innerhalb der internen DFG Datenbank und für Angebote wie GEPRIS oder den Research Explorer. Die ID existiert nicht als HTTP URI und ist dementsprechend nicht außerhalb der Datenbanken im Internet anwendbar. Interoperabilität mit anderen Identifiern ist nicht nachzuweisen. Dies liegt allerdings auch daran, dass es für die DFG ID kein (offen zugängliches) Metadatenschema gibt. Wenn in den Empfehlungen zum Kerndatensatz Forschung die DFG ID als Identifier für wissenschaftliche Einrichtungen empfohlen wird, sollten zumindest auch die Kriterien der Interoperabilität erfüllt sein. Die reine Zuordnung der DFG ID in einem Forschungsinformationssystem erscheint sinnvoll für die Disambiguierung außerhalb des Systems der DFG, jedoch wenig sinnvoll für weitere Entwicklungen wie die automatischen Aggregation von Publikationsdaten oder der Messung des Forschungsoutputs.

Einige dieser Funktionen scheint die PIC innerhalb des H2020 Systems zu erfüllen. Sie ist für die Disambiguierung und die automatische Generierung von Daten zu einer Einrichtung konzipiert. Allerdings bestehen keine Möglichkeiten des Resolving oder des direkten Zugriffs auf Metadaten, die die Einrichtung beschreiben. Über Formen der Interoperabilität können keine Angaben gemacht werden, es wird jedoch davon ausgegangen, dass die PIC Nummer nur für die Nutzung innerhalb des Research Participant Portals gemacht ist und somit keine Verweise auf andere Identifier beinhaltet.

Im Gegenteil dazu liegen die Vorteile der ISNI Kennung in ihrer Konzeption als Brückenidentifier. Das System ist so aufgebaut, dass nur die notwendigsten Daten zur Disambiguierung vorgehalten werden. Durch die Interoperabilität mit anderen Identifiern können weitere Informationen hinzu gefügt werden. Das Konzept von ISNI+ ist in Abbildung 1: **??** dargestellt. Ein Beispiel für die Verknüpfung ist die Open ISNI Datenbank, die Ringgold Identifier mit ISNI verknüpft und über eine API-Schnittstelle abfragbar macht.[75] ISNI hat keine Resolving Funktion, bietet jedoch über die normierte Syntax eine Möglichkeit, auf den Datensatz einer Einrichtung zuzugreifen, ähnlich der GND ID. Über diesen Datensatz werden auch gleichzeitig verknüpfte Identifier angezeigt, zum Beispiel Normdatensätze aus den großen Nationalbibliotheken (British National Library, Library of Congress, Deutsche Nationalbibliothek). Wie bereits in den vorangegangen Studien zu Identifiern für wissenschaftliche Einrichtungen festgestellt wurde, eignet sich ISNI (mit einem erweiterten Metadatenschema) durch seine Brückenfunktion als interoperables Glied für die Verknüpfung unterschiedlicher Identifier.
Der Ringgold Identifier liefert wahrscheinlich die größte Menge an Informationen zu einer

[75] Siehe: `http://isni.ringgold.com/` (Abgerufen am 26. Januar 2018).

Einrichtung. Von den betrachteten unterschiedlichen Identifier Systemen ist Ringgold sowohl das spezialisierteste als auch das am kommerziellsten ausgerichtete System. Grundprinzipien der Eindeutigkeit und, so lange das Unternehmen existiert, der Persistenz des Identifiers sind gewährleistet. Jedoch ist der Zugriff auf die Daten sehr stark eingeschränkt und kaum frei zugänglich. Durch das Mapping der Ringgold IDs auf die ISNI besteht ein gewisses Maß an Interoperabilität.

Die GRID ID erfüllt alle Kriterien bis auf eine Resolving Funktion. Die Datensätze sind mit einer befriedigenden Anzahl an Metadaten ausgestattet. Da auch obsolete Datensätze vorhanden sind, kann davon ausgegangen werden, dass die Identifier persistent sind (solange Digital Science den GRID Service betreibt). Die Verknüpfung mit anderen Identifiern und lokale Verortungen schaffen ähnlich wie bei der ISNI einen inhaltlichen Mehrwert. Relationen zwischen einzelnen Einrichtungen sind zum Mindestmaß erfüllt. Gäbe es eine Resolving Funktion und eine offene technische Schnittstelle (beispielsweise eine API) um automatisiert auf die in GRID gespeicherten Daten zugreifen zu können, würde sich GRID als PID für wissenschaftliche Einrichtungen besser eignen als ISNI, da der Fokus auf der Identifizierung für Einrichtungen liegt und zudem auch ein hohes Maß an Interoperabilität besteht.

Die Scopus Affiliation-ID ist bereits bei der Bewertung ihrer Eindeutigkeit problematisch. Zwar stimmt die eindeutige Beziehung zwischen dem Identifier und der Entität, jedoch fehlt die Prüfung auf die Validität der Entität. Beispielsweise wird unter der Affiliation-ID 100361864 die „Universität zu Berlin" am Spandauer Damm 130 in Berlin, Schweiz aufgeführt, die mit der 17 Dokumente, 7 Autor*innen und 194 Patenten verknüpft ist. (Siehe hierzu Abbildung 2: Affiliation Profile der *Universität zu Berlin* in Scopus.) Diese wissenschaftliche Einrichtung existiert jedoch nicht, hat aber eine (eindeutige) ID.[76] Für die Identifizierung von wissenschaftlichen Einrichtungen hat es sicherlich Sinn, wenn diese auch tatsächlich in der realen Welt existieren.[77] Die Scopus Affiliation-ID ist außerhalb des proprietären Systems Scopus nicht anwendbar, hat folglich auch keinen Resolving Service oder irgendeine Form von Interoperabilität mit anderen Identifiern.

Die Crossref Funder ID wird in dieser Arbeit als einziger Persistent Identifier für wissenschaftliche Einrichtungen identifiziert. Sie erfüllt alle oben aufgeführten Kriterien, inklusive der Resolving Funktion. Die ID ist ein gutes Beispiel dafür, wie die Nutzung von Persistent Identifiern die Verknüpfung von Informationen im Forschungskreislauf ermöglichen und als Grundlage für auf ihnen aufbauenden Services, wie das Funder Registry dienen können. Jedoch bezieht sich die ID nur auf Forschungsförderungsinstitutionen. Funder-IDs werden innerhalb des Crossref Schemas ausschließlich für Informationen zu Funding verwendet. Affiliations oder Organisationen als Contributer werden als Freitext <xsd:string> angegeben. Affiliations lassen sich über die Crossref API Schnittstelle anfragen.[78]

[76] Die Humboldt-Universität zu Berlin wurde als Universität zu Berlin gegründet, jedoch nicht in der Schweiz.

[77] Dies gilt auch für die Auswertung bibliometrischer Daten für Netzwerkanalysen und Rankings von wissenschaftlichen Einrichtungen, die sicherlich von validen Daten profitieren könnten.

[78] Beispielabfrage für Affiliations, die „Helmholtz-Gemeinschaft" beinhalten: `https://api.crossref.org/works?query.affiliation=Helmholtz-Gemeinschaft`.

6.2 Anforderungen an Metadaten

Um für die digitale Antragstellung von Nutzen zu sein, sollte der Identifier nach Möglichkeit alle wichtigen Metadaten zu einer Einrichtung beinhalten. Dabei handelt es sich vor allem Adressdaten und Typisierung der Einrichtung. Wenn diese Daten in den Metadaten vorkommen, könnte durch die einmalige Angabe des Identifiers die wiederholte Eingabe der Daten der Institution verhindert werden. Zudem würden alle Daten auf eine normierte Form bereitgestellt werden. Für Kooperationsprojekte wird über die Verknüpfung der unterschiedlichen Identifier die Antragstellung einfacher. Der Identifikator PIC ist eine Lösung, die sich in diese Richtung orientiert. Er ist jedoch kein offen zugänglicher Identifier. Die Generierung der Daten läuft über eine Datenbank und die Daten können nicht außerhalb von H2020 verwendet werden. Ein PID für wissenschaftliche Einrichtungen, der mit elektronischen Systemen zur Antragstellung interoperabel ist oder der als *Singel Sign On System*[79] für die Software fungiert, wäre hierbei eine befriedigende Lösung.

Forschungsinformationssysteme sollen das Reporting innerhalb einer wissenschaftlichen Einrichtung erleichtern. Dabei ist gerade für einrichtungsinterne FIS die Granularität, mit der Einrichtungssektionen identifiziert werden, von großer Bedeutung. Beispielsweise identifiziert CRIS der Friedrich-Alexander-Universität Erlangen Einrichtungseinheiten anhand der der internen FAU Org-ID, die meist der Nummer der Kostenstelle der Einheit entspricht (vgl. Friedrich-Alexander Universität Erlangen, 2018). Wie Sektionen innerhalb einer Einrichtung im Einrichtungsorganigramm verteilt sind, ist jedoch für jede wissenschaftliche Einrichtung anders. Ein globaler PID für wissenschaftliche Einrichtungen sollte in den zugehörigen Metadaten daher die Möglichkeit bieten, eine tiefstmögliche gewünschte Granularität zu erreichen und die einzelnen Teilorganisationen durch korrekte Beschreibung der Relationen mit den übergeordneten Einheiten zu verknüpfen. Dies kann durch einfache Parent-Child-Relationen erfolgen, jedoch ist mit steigender Granularität sicherlich auch eine tiefere Erschließung der Relationen von Vorteil. Dies setzt jedoch wiederum ein kontrolliertes Vokabular zur Beschreibung von Relationen zwischen Einrichtungseinheiten voraus, welches den Metadaten des Identifiers zu Grunde liegen sollte. Wenn für die Aggregation von wissenschaftlichem Output für ein Forschungsinformationssystem und Universitätsbibliographien weiterhin interne IDs proprietärer bibliografischer Datenbanken benutzt werden, sollten diese IDs innerhalb der Metadaten eines PIDs für wissenschaftliche Einrichtungen beispielsweise mit einer <sameAs> oder <realtedIdentifier> Relation verknüpft werden, um Interoperabilität zu ermöglichen. Je nach Design des Identifiers wäre auch denkbar, eine Relation zu Personen herzustellen, um Affiliationen mit Personendaten zu verknüpfen. Dies geschieht innerhalb von ORCID bereits über die Ringgold ID bzw. über den GRID Identifier.

Für die Automatisierung von Datenflüssen und Integrationsmöglichkeiten des PID Systems in bereits bestehende Systeme ist die Übertragungsmöglichkeit der Metadaten über eine standardisierte Schnittstelle (zum Beispiel eine API-Schnittstelle) von großer Bedeutung. Wenn Services, wie die automatisierte Übernahme von bibliografischen Daten in Repository

[79] Die Möglichkeit, nach einmaligem Anmelden, zum Beispiel über eine verifizierte IP-Adresse auf alle im Netzwerk befindlichen Services zuzugreifen.

Systeme einer Einrichtung angeboten werden sollen, sind Persistent Identifier zur eindeutigen Identifizierung der Einrichtung und ihrer Rechte unabdinglich. In einer idealen Welt würden wissenschaftliche Publikationen und andere Arten von Forschungsoutputs von Repositorysystemen automatisch aufgrund ihrer Metadaten identifiziert und in das System übernommen. Persistent Identifier könnten hier durch ihre Funktion der Maschinenlesbarkeit die Kommunikationen zwischen Publikationssystemen und Repositorysystemen vereinfachen.

Für das Tracking von Forschungsförderungsmaßnahmen von Seiten der Forschungsförderungseinrichtungen ist mit der Crossref Funder ID bereits ein gutes Beispiel für den Nutzen von PID zur vernetzten Darstellung von Forschungsinformationen entstanden. Die Datenbank der DFG GEPRIS gibt bisher nur Angaben zu durchgeführten Projekten, ohne diese jedoch mit den Produkten der Fördermaßnahmen (Publikationen) zu verknüpfen. Ein Identifiersystem für wissenschaftliche Einrichtungen müsste in diesem Sinne so aufgebaut sein, dass es sowohl frei einsehbar, als auch mit anderen Metadatenschemata, beispielsweise für die Vergabe von DOIs für Publikationen interoperabel ist.

Persistent Identifier sollten innerhalb der Metadaten die Benennung von Einrichtungen in eine temporale Relation setzen. Dies könnte über zeitliche Angaben zu einer bestimmten Form der Benennung einer Einrichtung geschehen. Es wäre denkbar, jede Form des Namens zusätzlich mit einem <NameTime> oder <NameDate> Attribut zu versehen. Um Kollegialitäten innerhalb einer Datenbank von potentiellen Gutachter*innen für Förderanträge zu vermeiden, würde über den Persistent Identifier für die Zuordnung von Zugehörigkeiten so sichergestellt, dass trotz Änderungen von Namen der Einrichtung die Angaben zu Kollegialitäten und eventuellen Befangenheiten valide bleiben. Auch hierarchische Beziehungen zwischen Teilen von Einrichtungen und übergeordneten Einrichtungen können dabei helfen, eine eindeutige Zuordnung von Personen zu wissenschaftlichen Einrichtungen ohne weitere Kenntnisse über ihre Struktur zu erreichen.

Für das Monitoring von Open Access Publikationen einer Einrichtung gelten grundsätzlich die selben Kriterien wie für die Aggregation von Publikationen. Persistent Identifier in den Affiliationsangaben von Publikationen würden eine klare Zuordnung zu Einrichtungen schaffen und somit eine weitestgehend vollständige Angabe zu allen aus einer Einrichtung erschienenen Publikationen ermöglichen. Auch hier sind Relationen zwischen untergeordneten und verwandten Einrichtungen innerhalb der Metadaten sowie Angaben zu alternativen Benennungen (zum Beispiel über <alternativeName> sowie <alternativeNameType>) wichtig. In der Zukunft könnten automatisierte Monitoringverfahren denkbar sein, die für jede Einrichtung über Schnittstellen zwischen Publikations- und Monitoringdiensten mit Hilfe des Persistent Identifiers den Anteil an Open Access erschienen Veröffentlichungen misst.
Für die Verwaltung von Article Processing Charges, sowie automatisierte Transaktionsflüsse zwischen Verlagen und Einrichtungen ist es notwendig, dass jede beteiligte Kostenstelle eindeutig identifiziert werden kann und, wieder über Relationsangaben in den Metadaten, mit den Identifikatoren der Affiliation verbunden werden kann. Erfüllt werden kann diese Anforderung durch die Selbstverwaltung und selbstständige Registrierung von Identifikatoren durch die Einrichtung. Wenn die Zuordnung von Identifiern nur aus Angaben aus Publikationen oder

durch externe Unternehmen geschieht, wird es kaum möglich sein, die richtigen Kostenstellen zu identifizieren, da diese meistens weder selber publizieren noch durch externe Einrichtungen erkennbar sind. Die Verknüpfung von Identifikatoren für Affiliationsangaben und Kostenstellen würde dazu führen, dass für alle Beteiligten am Publikationsprozess (Autor*innen, Verlage und Verantwortliche für Publikationsfonds) der Transaktionsprozess für die APC-Kosten klar vorgegeben und automatisch ausführbar ist und somit der Verwaltungsaufwand reduziert wird.

Es lässt sich festhalten, dass die Anforderungen an Persistent Identifier für wissenschaftliche Einrichtungen abhängig davon sind, für welche Anwendungen sie bestehen sollen. Aus den behandelten Anwendungsfällen ergeben sich allerdings neben Grundelementen für Metadaten für Persistent Identifier (Name, alternativer Name, Adresse, Typ) folgende Anforderungen:

- **Granularität sowie Skalierbarkeit**: Es sollte die Möglichkeit geben, Teile von Einrichtungen, wie Institute oder Sektionen, so genau wie möglich und nötig zu identifizieren. Dafür müssen Angaben innerhalb der Metadaten eingrenzbar oder erweiterbar sein, je nachdem auf welcher Ebene die Einrichtung identifiziert wird.
- **Angaben von Relationen**: Innerhalb der Metadaten sollten sowohl hierarchische Relationen zwischen Teilen von Einrichtungen als auch temporäre Informationen (zum Beispiel Zeitraum der Benennung einer Einrichtung) sowie Mappings auf externe Identifier über <sameAs> Relationen um Interoperabilität zu fördern.
- **Angaben zu Rollen von Einrichtungen in Relation zu anderen Entitäten**: Hierbei wäre vorstellbar, innerhalb der Metadaten Einrichtungen und ihre Rolle außerhalb von einfacher Affiliation zu identifizieren, wie zum Beispiel oben genannte Kostenstellen für APC Kosten. Demeranville u. a. schlugen in der Diskussion zu Open Organisation Identifier vor, solche Relationen ggf. als Linked Data Tripel (Subjekt-Prädikat-Objekt) in einem parallelen, externen System abzulegen, welche die Darstellung von Relationen zwischen unterschiedlichen Identifiern ermöglicht (vgl. Demeranville u. a., 2016).

6.3 Bewertung der Anwendungsfälle

Die Nutzung von Persistent Identifiern für wissenschaftliche Einrichtungen findet in den betrachteten Anwendungsfällen bislang nicht oder nicht offen nachvollziehbar statt. Von denen in dieser Arbeit betrachteten Nutzungsszenarien werden folgende als Kernanwendungsbereiche identifiziert, für die in der Recherche bereits Ansätze für die Verwendung von verschiedenen Arten von Identifikatoren auftauchten bzw. die Vorteile der Nutzung eines PIDs aufgrund der Maschinenlesbarkeit und inkludierten Metadaten hier besonders gut greifen würden.

- **Aggregation von wissenschaftlichem Output und Reporting**
  Bisher werden für die Erstellung von Hochschulbibliografien oder innerhalb von Forschungsinformationssystemen entweder Identifikatoren für Forschende einer Einrichtung oder proprietäre Systeme, wie die Scopus Affiliation ID oder die Suche im Web of Science verwendet. Diese Identifikatoren sind einerseits weder Persistent Identifier im hier definierten Sinne und andererseits nicht offen zugänglich und von der Einrichtung kontrollierbar. Durch die Nutzung eines PIDs könnten die Aggregation und die Auswertung von wissenschaftlichem Output und Forschungsinformation durch die eindeutige

Bezeichnung der Einrichtung an Transparenz und Qualität gewinnen. Die Nutzung eines offenen Persistent Identifiers würde sich zudem in der Zukunft gut in das CERIF Modell für einheitliche Forschungsinformationen im europäischen Raum einbinden lassen. *OrganisationUnit* ist hier eines der Kernelemente.[80]

- **Open Access Monitoring und APC-Management**
  An diesem Anwendungsfall lassen sich die zwei Eigenschaften von Persistent Identifiern und somit die Vorteile für die Verwendung gut zeigen. Für das Open Access Monitoring für einzelne wissenschaftliche Einrichtungen ist die eindeutige Bezeichnung dieser besonders wichtig. Das Retrieval der Daten von bibliometrischen Datenbanken kann so erleichtert werden. Die Suche nach unterschiedlichen Ansetzungsformen der Einrichtung und die Gefahr, keine vollständige Liste von Publikationen zu erhalten, weil Zugehörigkeiten nicht richtig indexiert wurden, könnte durch die Verwendung eindeutiger Bezeichnungen durch Identifier beseitigt werden. Die Verwaltung und richtige Zuordnung von APC-Gebühren sowie eine Automatisierung von Transaktionsflüssen kann durch die maschinelle Verarbeitungsmöglichkeit sowie die Anreicherung mit Metadaten, beispielsweise zu Kostenstellen, erheblich erleichtert werden.

Es bleibt festzustellen, dass die Anwendungsbereiche für Persistent Identifier für wissenschaftliche Einrichtungen sehr breit sind. Als besonders geeignet werden sie in den Fällen angesehen, wo sie in einer Doppelfunktion zur Disambiguierung und eindeutigen Benennung und zur Automatisierung von Datenflüssen und Interoperabilitäten zwischen unterschiedlichen Systemen angewendet werden.

6.4 Offene Forschungsinfrastrukturen

Persistent Identifier sind ein Teil der Forschungsinfrastruktur. Sie sollen Forschungsinformationen, Forschungsförderungen, sowie Publikations- und Verwaltungsprozesse vereinfachen, Workflows unterstützen und Informationsflüsse durch Interoperabiltät mit anderen Identifiern ermöglichen. Nicht zuletzt können durch die Nutzung von PID für wissenschaftliche Einrichtungen Metadatenqualitäten verbessert, und durch die Öffnung des Identifikationssystems die Transparenz von Informationsflüssen und Forschungsinformationen vergrößert werden. Bilder, Lin und Neylon verdeutlichen, dass das Streben nach Öffnung von Forschungsergebnissen (Open Access, Open Data) sich in der Forderung nach offenen Forschungsinfrastrukturen widerspiegeln soll:

> „*We believe that data about the research process itself deserves exactly the same level of respect and care. The scholarly community does not own or control most of this information. For example, we could have built or taken on the infrastructure to collect bibliographic data and citations but that task was left to private enterprise*" (Bilder, Lin und Neylon, 2015, S. 1).

Offene Forschungsinfrastrukturen sind laut der Autoren stabil, nachhaltig und vertrauenswürdig. Vertrauen der wissenschaftlichen Community steht dabei an wichtigster Stelle. Wie bereits

[80] Siehe: `https://www.eurocris.org/Uploads/Web%20pages/CERIF-1.6/documentation/MImage.html` (Abgerufen am 26. Januar 2018).

oben erwähnt bezieht sich dieses Vertraue auf die Bereiche Governance (Verwaltung), Sustainability (Nachhaltigkeit im finanziellen Sinn) und Insurance (Ansiedlung in der Community). Das Vertrauen der Community kann insbesondere durch ihre Einbindung in wissenschaftliche Infrastrukturen, durch transparente Verwaltungsvorgänge und durch eine Ansiedlung der Infrastrukturservices innerhalb der Kontrolle der Nutzer*innen erreicht werden. Betrachtete man die in dieser Arbeit aufgeführten bisherigen Möglichkeiten zur eindeutigen Identifizierung von wissenschaftlichen Einrichtungen unter diesen Gesichtspunkten fällt schnell auf, dass keines der beschriebenen Identifikatorensysteme diese Anforderungen erfüllt.

Im Bereich der Governance lassen sich die untersuchten Identifier in zwei Kategorien einteilen: (1) Verantwortung für die Verwaltung der Identifier liegt bei traditionell gewachsenen Informationsdienstleistungseinrichtungen (Nationalbibliotheken, Forschungsförderungseinrichtungen, Verlage)[81] oder (2) kommerzielle Dienstleistungsfirmen bieten proprietäre Lösungen für wissenschaftliche Einrichtungen an.[82] Governance Entscheidungen sind, außer für Crossref, für keines der Systeme einsehbar oder offen dokumentiert. Die operationalen Verantwortungen für ein Identifikatorensystem sollten jedoch offen dargelegt sein und nach Möglichkeit alle Stakeholder aus der wissenschaftlichen Community berücksichtigen. Dabei sollten innerhalb der Verantwortlichkeitsbereiche keine bevorzugten Stellungen von einer bestimmten Gruppe (beispielsweise nur Verlage, nur Informationsdiensleister) liegen. Die Organisation Identifier Working Group empfiehlt, für das in Planung stehende Produkt auf bereits bestehende non-profit Organisationen als Gastgebereinrichtungen (Hosts) für die Entwicklung des Identifiersystems zurückzugreifen (vgl. Pentz u. a., 2017). Damit können Verantwortungsstrukturen noch wachsen und bestehende Infrastrukturen nach genutzt werden. Der Request for Information für dieses Vorhaben wurde im September 2017 publik gemacht (vgl. Haak u. a., 2017).

Die Nachhaltigkeit von Identifikatorensystemen bestimmt maßgeblich auch ihre Persistenz. Bilder, Lin und Neylon legen den Fokus von Sustainability auf die finanzielle Nachhaltigkeit. Vertrauen in Infrastruktureinrichtungen kann sicherlich dadurch erhöht werden, dass die finanzielle Zukunft und somit die Existenz des Services gesichert ist. Für Persistent Identifier Systeme ist gerade dieser Blick in die Zukunft essentiell. Zwar ist im Bereich digitaler Infrastrukturen „Persistenz" ein Terminus, der zeitlich noch nicht genau eingegrenzt ist, jedoch bedingen sich das Vertrauen in die garantierte Nachhaltigkeit eines Systems und die finanzielle Absicherung gegenseitig. Finanzierungsmodelle, wie das von Bilder, Lin und Neylon favorisierte non-profit, demokratisch geführte Mitgliedschaftsorganisation, bei der das Produkt von Mitgliedschaftsbeiträgen getragen wird, haben sich beispielsweise für ORCID als bisher finanziell nachhaltig und dabei so „offen" wie möglich erwiesen.

Die Ansiedlung von Forschungsinfrastrukturen in der Community und eine höchstmögliche Transparenz sowohl für die verwendeten Produkte (Software, Hardware) als auch für die verwendeten und produzierten Daten führt dazu, dass der Nutzercommunity die Kontrolle

[81] GND ID, ISNI, DFG ID, PIC.
[82] Ringgold ID, GRID, Scopus Affiliation ID, Crossref Funder ID.

über das Produkt zurückgegeben wird. Im Bezug auf ein Identifiersystem spiegelt sich dieses Prinzip in der Möglichkeit einer Selbstverwaltung der erhobenen Daten und der Identifier einer Einrichtung. Der Übergang vom passiven identifiziert werden zum aktiven Selbstidentifizieren zeigt sich für den ORCID Identifier als ein gut funktionierendes und attraktives Konzept. Wenn wissenschaftliche Einrichtungen die Kontrolle über die Verwaltung der eigenen Daten erhalten, kann dies zu einem größeren Vertrauen in das Identifiersystem sowie potentiell zu einer höheren Qualität der Daten und somit des Produktes führen. Keiner der in dieser Arbeit analysierten Identifiersysteme bietet der wissenschaftlichen Community diese Möglichkeit. Die Absicherung der Validität der Daten, der gewünschten und benötigten Granularität bei der Identifizierung von einzelnen Organisationsteilen und die Kontrolle über die Nutzung der Daten ist durch keines der Systeme gesichert.

7 Fazit und Ausblick

In dieser Arbeit wurden Probleme bei der eindeutigen Identifizierung wissenschaftlicher Einrichtungen aufgezeigt, Persistent Identifier als Lösung dieser Probleme vorgeschlagen und bereits bestehende Lösungsansätze in Form von bestehenden Identifikatorensystemen für wissenschaftliche Einrichtungen ausgewertet. Außerdem wurden Anwendungsszenarien für die eindeutige Identifikation wissenschaftlicher Einrichtungen und bisherige Lösungen vorgestellt. Zuletzt wurden die betrachteten Identifikatorensysteme bewertet und aus den Anwendungsfällen Kriterien für Persistent Identifier, für den Einsatz in den Nutzungsszenarien, herausgearbeitet. Offene Forschungsinfrastrukturen als Organisationsform für ein Identifikatorensystem bieten dabei eine wichtige Grundlage. Fazit dieser Arbeit ist, dass bisher verwendete Identifikatoren, außer die als DOI ausgegebene Crossref Funder ID, die Anforderungen an PIDs nicht erfüllen und nicht als Persistent Identifier gewertet werden können. Es bestehen jedoch Anwendungsfälle, für die PIDs zur eindeutigen, anhaltenden Identifizierung von Einrichtung sowie dem automatisierten Austausch von Metadaten sehr gut geeignet sind.

Als problematisch bei der Bewertung der Identifikatoren stellte sich heraus, dass nicht alle Identifikatorensysteme offen zugänglich sind. Eine Möglichkeit herauszufinden, welche Informationen über die Einrichtung aufgenommen werden, besteht darin, die ausgegebenen Felder bei der Darstellung der Einrichtung zu betrachten. Daraus folgt, dass bei der Analyse die Vollständigkeit der Informationen nicht garantiert werden kann. Außerdem sind bei geschlossenen Systemen häufig keine Informationen vorhanden, auf welche Art und Weise die Daten gespeichert werden. Die Anwendungsszenarien sind teilweise sehr breit und bisherige Lösungen lassen sich schlecht oder nur beispielhaft recherchieren. Methodisch würden sich qualitative Erhebungen wie Interviews oder Umfragen für die Entwicklung von valide auswertbaren Anwendungsfällen als wichtige Ergänzung zu dieser Arbeit anbieten. Die Ergebnisse dieser Arbeit können nur im Rahmen der betrachteten Anwendungsfälle auf Identifier-Systeme übertragen werden.

Die eindeutige Identifizierung von Organisationen oder wissenschaftlichen Einrichtungen ist aufgrund von veränderlichen internen Strukturen, Veränderungen des Status einer Einrichtung, Zusammenschlüsse oder Aufteilungen sowie Umbenennungen oder Ausgründungen nicht einfach. Tatsächlich aber ist die korrekte Angabe des Namens einer Einrichtung in vielen Bereichen des Wissenschaftsbetriebs dringend notwendig und nicht normierte Angaben von Affiliationen führen zu unsauberen Metadaten oder falschen Informationen. Im Zuge dieser Arbeit konnten vor allem für die Aggregation von wissenschaftlichen Publikationen und die Erhebung von Forschungsinformationen, sowie im Bereich des Open Access Monitorings und der Verwaltung von APC-Kosten Anwendungsbereiche für Persistent Identifier für wissenschaftliche Einrichtungen identifiziert werden. Das Tracking von wissenschaftlichen Ergebnissen durch Forschungsförderungseinrichtungen kann bereits durch die Crossref Funder ID weitestgehend gut bewerkstelligt werden, jedoch greift der Identifier nicht weit genug, um für alle wissenschaftlichen Einrichtungen anwendbar zu sein.
Forschungsinformationen über die Tätigkeiten einzelner Einrichtungen sowie deren wis-

senschaftlichen Output gewinnen für bibliometrische Auswertungen oder Rankings von Forschungseinrichtungen an Bedeutung. Für eine transparentere Nutzung von Identifikatoren zur Datenerhebung und Verbesserung der Datenqualität für Forschungsinformationen würden sich Persistent Identifier sehr gut eignen. Neben Transparenz können Persistent Identifier aufgrund ihrer Maschinenlesbarkeit auch zur Vereinfachungen von Workflows und Verwaltungsaufgaben beitragen. Dies zeigt sich vor allem im Anwendungsfall Open Access Monitoring und APC Management.

Für Persistent Identifier sind die mitgelieferten Metadaten ein großer Mehrwert. Hierbei ist festzustellen, dass die Zugänglichkeit dieser Metadaten sowie die Kontrolle über die Granularität und Validität der Daten Einfluss auf die Anwendungsbereiche für PIDs für wissenschaftliche Einrichtungen nehmen. Im Sinne der offenen wissenschaftlichen Infrastrukturen kann das Vertrauen in Validität und ausreichende Granularität sicherlich davon profitieren, wenn die Verantwortung für die vorhandenen Daten bei den identifizierten Organisationen selbst liegt. Die steigende Nutzung des ORCID Registries, welches nach dem selben Prinzip funktioniert, bestätigt diese Vermutung. Dennoch ist dieser Ansatz nicht immer unproblematisch. Es bleibt unter anderem zu klären, auf welche Art und Weise Identifier von den Einrichtungen erstellt werden sollen, in welcher Form eine Qualitätssicherung der angegebenen Informationen geschieht, wer innerhalb der Einrichtung die Verantwortung für die Betreuung des Identifiers übernimmt, was mit Dubletten oder falschen Einträgen geschehen muss und wie der Missbrauch des Identifiers (beispielsweise durch fälschliche Verwendung bei der Angabe von Affiliationen für Publikationen) sowie datenschutzrechtliche Aspekte behandelt werden.
Metadaten haben vor allem dann Wert, wenn sie Verwendung in auf sie aufbauende Services finden können. Die Anreicherung von Identifikatoren mit Zusatzinformationen zur Disambiguierung von Namen oder Organisationen greift nur so weit, wie der Identifier wirken kann. Identifikatoren für interne (proprietäre) Datenbanksysteme sind, auch wenn sie wissenschaftliche Einrichtungen noch so genau beschreiben, für die Anwendung im Sinne offener Wissenschaft nicht geeignet. Die Möglichkeit, PID Systeme für Services, die auf ihnen beruhen zu verwenden, ist nur gegeben, wenn die Identifier und die Metadaten offen zugänglich und nach nutzbar sind. Beispielsweise würden Services wie das FundRef Registry oder der Crossref DOI Citation Formatter[83] niemals mit in sich verschlossenen, proprietären Datenbanksystemen möglich sein. Wissenschaftsprozesse und Wissenschaftskommunikation reicht über nationale Grenzen hinweg. Nationale Identifikationssysteme, wie Normdatensätze oder IDs von nationalen Forschungsförderungseinrichten sind ohne Interoperabilitätsmöglichkeiten daher für internationale Systeme nicht geeignet.

Offene Forschungsinfrastrukturen können das Vertrauen in Identifikatorensysteme durch transparentes Government, finanzielle Absicherung und Verankerung in der Community unterstützen. Dies ist gerade für den Aspekt der Langlebigkeit eines Identifiers von großer Bedeutung. Persistenz im Bezug auf digitale Objekte und Services ist bislang kein ganz klar

[83] Siehe: `https://citation.crosscite.org/` (Abgerufen am 26. Januar 2018).

definierter Terminus.[84] Dabei bedingen sich das Vertrauen in die Langlebigkeit eines Identifiers, seine Nutzung und seine tatsächliche Langlebigkeit gegenseitig. Wenn das Vertrauen der wissenschaftlichen Community in den Identifier groß genug ist, wird er verbreitete Nutzung finden, welche wiederum zu einer angenommenen erweiterten Langlebigkeit führt. Wenn ein Identifikatorensystem für wissenschaftliche Einrichtungen geschaffen wird, sollte dieses System alle Anforderungen an Persistent Identifier erfüllen, sowie global nutzbar und offen zugänglich sein. Neben einer verlässlichen technischen Infrastruktur betrifft dies auch die Übernahme der Verantwortung für den Service durch eine vertrauenswürdige Organisation. Kurz vor Abgabe dieser Arbeit wurden im Zuge des Organization ID Stakeholder Meetings am 22. Januar 2018[85] die Antworten auf den Request for Information als Google Doc veröffentlicht (Organization Identifier Working Group, 2018).[86] 24 Organisationen bekunden ihr Interesse an der gemeinsamen Arbeit an einem internationalen offenen Identifikatorensystem für wissenschaftliche Einrichtungen, darunter auch ISNI, Ringgold, Digital Science, CASRAI und die DFG.
Das Thema scheint durch das Organization Identifier Projekt und die Working Group in das Interesse der wissenschaftlichen Community zu rücken, aber die Entwicklung und Verbreitung eines Persistent Identifiers braucht mit Sicherheit aufgrund seines breiten Anwendungsbereiches und unterschiedlicher Stakeholdergruppen sowie den benötigten technischen Voraussetzungen noch einige Zeit. Unterstützend wirken könnten hier ein niederschwelliger Zugang zu dem Service, eine Sensibilisierung für Herausforderungen bei der eindeutigen Identifizierung wissenschaftlicher Einrichtungen und Konsequenzen, die sich aus Fehlern in Datenbanken u.ä. ergeben, sowie eine klare Kommunikation und Präsentation der Mehrwerte eines eindeutigen, offenen und langlebigen Identifikatorensystems für wissenschaftliche Einrichtungen.

[84] Siehe hierzu sehr ausführlich und anschaulich Kunze u. a., 2017: *Persistence Statements: Describing Digital Stickiness*.
[85] Siehe: `https://orcid.org/content/2018-org-id-meeting` (Abgerufen am 26. Januar 2018).
[86] Siehe auch Anhang.

Literatur

Altman, Micah und King, Gary (2007). „A Proposed Standard for the Scholarly Citation of Quantitative Data“. In: *D-Lib Magazine* 13 (3/4). DOI: https://doi.org/10.1045/march2007-altman.

Amante, Maria João u. a. (2017). „PTCRIS_OrgID - Portuguese Organisation Identifiers Authoritative System“. In: *Procedia Computer Science* 106, S. 138–145. DOI: https://doi.org/https://doi.org/10.1016/j.procs.2017.03.046.

Arms, William Y. (1995). „Key Concepts in the Architecture of the Digital Library“. In: *D-Lib Magazine* July. URL: http://www.dlib.org/dlib/July95/07arms.html.

Askitas, Nikos (2010). *What Makes Persistent Identifiers Persistent?* Working Paper Series des RatSWD 147. URL: https://ideas.repec.org/p/rsw/rswwps/rswwps147.html (Abgerufen am 26. 01. 2018).

Becker, Pascal-Nicolas u. a. (2016a). *DeepGreen – Metadata Schema for the exchange of publications between publishers and open access repositories*. DOI: https://doi.org/10.12752/3.dg.1.0.

— (2016b). *Questionnaire for effective exchange of bibliographic metadata – current status of publishing houses*. de. URL: https://opus4.kobv.de/opus4-zib/frontdoor/index/index/docId/6041 (Abgerufen am 26. 01. 2018).

Berg-Cross, Gary, Ritz, Raphael und Wittenburg, Peter (2015). *RDA Data Foundation and Terminology-DFT: Results RFC*. Version 1.5. URL: https://www.rd-alliance.org/system/files/DFT%20Core%20Terms-and%20model-v1-6.pdf (Abgerufen am 15. 12. 2017).

Bilder, Geoffrey, Brown, Josh und Demeranville, Tom (2016). *Organisation identifiers: current provider survey*. DOI: https://doi.org/10.5438/4716.

Bilder, Geoffrey, Lin, Jennifer und Neylon, Cameron (2015). *Principles for Open Scholarly Infrastructures-v1*. DOI: https://doi.org/10.6084/m9.figshare.1314859.v1.

Brown, Christopher (2015). *Organisational Identifiers Working Group – outputs and update*. URL: https://jisccasraipilot.jiscinvolve.org/wp/2015/03/06/organisational-identifiers-working-group-outputs-and-update/ (Abgerufen am 08. 12. 2017).

Bruch, Christoph u. a. (2016). *Empfehlungen zur Open-Access-Transformation: strategische und praktische Verankerung von Open Access in der Informationsversorgung wissenschaftlicher Einrichtungen. Ad-hoc-AG Open-Access-Gold im Rahmen der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen*. DOI: https://doi.org/10.3249/allianzoa.011.

Bryant, Rebecca, Dortmund, Annette und Malpas, Constance (2017). *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information Management*. Dublin, Ohio. DOI: https://doi.org/10.25333/C32K7M.

Bundesministerium für Bildung und Forschung, Hrsg. (2016). *Open Access in Deutschland. Die Strategie des Bundesministeriums für Bildung und Forschung*. URL: https://www.bmbf.de/pub/Open_Access_in_Deutschland.pdf (Abgerufen am 08. 12. 2017).

Bunge, Eva und Voigt, Michaela (2016). *Analyse an institution's publication output and determine the open access share*. URL: https://github.com/tuub/oa-eval (Abgerufen am 12. 01. 2018).

Car, Nicholas J., Golodoniuc, Pavel und Klump, Jens (2017). „The Challenge of Ensuring Persistency of Identifier Systems in the World of Ever-Changing Technology". In: *Data Science Journal* 16. DOI: https://doi.org/10.5334/dsj-2017-013.

Chvatal, Donald (2008). „The Journal Supply Chain Efficiency Improvement Pilot: What's Good, What's Bad, What's Missing?" In: *The Serials Librarian* 54 (1-2), S. 105–119. DOI: https://doi.org/10.1080/03615260801973901.

Crossref (2017). *Open Funder Registry*. URL: https://github.com/CrossRef/open-funder-registry (Abgerufen am 03. 01. 2018).

Cruse, Patricia, Haak, Laure und Pentz, Ed (2016). *Organization Identifier Project: A Way Forward*. DOI: https://doi.org/10.5438/2906.

Dappert, Angela u. a. (2017). „Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research". In: *Data Science Journal* 16. DOI: https://doi.org/10.5334/dsj-2017-028. (Abgerufen am 28. 10. 2017).

Davidson, Joyce (2006). *Persistent Identifiers*. DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. URL: http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/persistent-identifiers (Abgerufen am 18. 11. 2017).

De Bruin, R. E. und Moed, H. F. (1990). „The unification of addresses in scientific publications". In: *Informetrics 89/90*, S. 65–78. URL: https://uhdspace.uhasselt.be/dspace/handle/1942/850 (Abgerufen am 08. 12. 2017).

Dellavalle, R. P. (2003). „Going, Going, Gone: Lost Internet References". In: *Science* 302 (5646), S. 787–788. DOI: https://doi.org/10.1126/science.1088234.

Delpeuch, Antonin (2017). *Aligned Isni And Ringgold Identifiers For Institutions*. DOI: https://doi.org/10.5281/zenodo.844869.

Demeranville, Tom u. a. (2016). *Organisation Identifiers - Minimum viable product requirements*. DOI: https://doi.org/10.6084/m9.figshare.3479141.v1.

Deutsche Forschungsgemeinschaft (2014). *Hinweise zu Fragen der Befangenheit*. URL: http://www.dfg.de/formulare/10_201/10_201_de.pdf (Abgerufen am 24. 01. 2018).

— (2018). *Über GEPRIS*. URL: http://gepris.dfg.de/gepris/OCTOPUS?task=showAbout (Abgerufen am 24. 01. 2018).

Digital Science, Szomszor, Martin und Adams, Jonathan (2017). *Digital Research Report: Discovery and Analysis of Global Research Trends Using GRID*. DOI: https://doi.org/10.6084/m9.figshare.4714249.v1.

DIN ISO (2012). „Information and documentation - International standard name identifier (ISNI)". Pat. ISO 27729:2012.

Ebert, Barbara u. a. (2016). *Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen. Positionspapier. Version 1.1*. DOI: https://doi.org/10.5281/zenodo.45564.

el Asati, M'hamed (2016). *Scopus data and affiliation handling in Scopus*. URL: http://www.um.rnu.tn/uploaded/Actualites/Pr%C3%A9sentation%201elsev.pdf (Abgerufen am 19. 01. 2018).

Elsevier (2018a). *What is the Scopus Affiliation Identifier?* URL: https://service.elsevier.com/app/answers/detail/a_id/11215/supporthub/scopus/.

— (2018b). *What is the Scopus Affiliation Identifier? - Scopus: Access and use Support Center*. URL: https://service.elsevier.com/app/answers/detail/a_id/11215/supporthub/scopus/%7D/ (Abgerufen am 19. 01. 2018).

European Commission (2017a). *FAQ - Research Participant Portal*. URL: https://ec.europa.eu/research/participants/portal/desktop/en/support/faq.html (Abgerufen am 19. 01. 2018).

— (2017b). *Legal notice*. URL: https://ec.europa.eu/info/legal-notice_en (Abgerufen am 24. 01. 2018).

— (2017c). *Model contract for experts*. URL: http://ec.europa.eu/research/participants/data/ref/h2020/experts_manual/h2020-experts-mono-contract_en.pdf (Abgerufen am 24. 01. 2018).

Fenner, Martin u. a. (2015). *D2.1: Artefact, Contributor, and Organisation Relationship Data Schema*. DOI: https://doi.org/10.5281/zenodo.30799.

Fenner, Martin u. a. (2016). *THOR: Conceptual Model of Persistent Identifier Linking*. DOI: https://doi.org/10.5281/zenodo.48705.

Ferguson, Nicky, Moore, Richard und Schmoller, Seb (2014). *Review of selected organisational IDs and development of use cases for the Jisc CASRAI - UK Organisational Identifiers Working Group*. Hrsg. von Christopher Brown. URL: http://repository.jisc.ac.uk/id/eprint/5853 (Abgerufen am 08. 12. 2017).

Frick, Claudia (2017). „Empfehlungen für Workflows zur Übernahme von Publikationsgebühren“. In: *Praxishandbuch Open Access*. Hrsg. von Konstanze Söllner und Bernhard Mittermaier. Berlin, Boston: De Gruyter Saur. DOI: https://doi.org/10.1515/9783110494068-037.

Friedrich-Alexander Universität Erlangen (2018). *Woher erhalte ich die Organisationseinheitennummer (FAU.Org-Nr) die in Formularen angeben werden müssen?* URL: https://helpdesk.rrze.uni-erlangen.de/otrs/public.pl?Action=PublicFAQZoom;ItemID=1632 (Abgerufen am 24. 01. 2018).

Fuchs, Claudio, Pampel, Heinz und Vierkant, Paul (2017). „ORCID in Deutschland – Ergebnisse einer Bestandsaufnahme im Jahr 2016“. In: *o-bib. Das offene Bibliotheksjournal* 4 (2), S. 35–55. DOI: https://doi.org/10.5282/o-bib/2017H2S35-55.

Galvez, Carmen und Moya-Anegón, Félix (2007). „Standardizing formats of corporate source data“. In: *Scientometrics* 70 (1), S. 3–26. DOI: https://doi.org/10.1007/s11192-007-0101-0.

Geschuhn, Kai und Sikora, Adriana (2015). „Management von Article Processing Charges - Herausforderungen für Bibliotheken“. In: *o-bib. Das offene Bibliotheksjournal* 2 (1), S. 27–34. DOI: https://doi.org/https://doi.org/10.5282/o-bib/2015H1S27-34.

GRID (2018a). *Disambiguate*. URL: https://grid.ac/pages/disambiguate_your_data (Abgerufen am 17. 01. 2018).

— (2018b). *Slice and dice your data*. URL: https://grid.ac/pages/slice_and_dice_your_data (Abgerufen am 17. 01. 2018).

— (2018c). *Statistics*. URL: https://grid.ac/stats (Abgerufen am 17. 01. 2018).

Güdler, Jürgen (2016). *Daten@DFG. Nutzung des DFG-Datenpools für informatorische Zwecke*. Bonn. URL: http://www.wissenschaftsmanagement-online.de/system/files/G%C3%BCdler_im_MPA_Alumni_160414.pdf (Abgerufen am 21. 12. 2017).

Haak, Laurel (2015). *Requiring ORCID in Publication Workflows: Open Letter*. URL: https://orcid.org/content/requiring-orcid-publication-workflows-open-letter (Abgerufen am 23. 01. 2018).

— (2017). *Research Reproducibility: the Role of Persistent Identifiers*. DOI: https://doi.org/10.23640/07243.5435389.v1.

Haak, Laurel u. a. (2017). „Organization Identifier Project: Request for Information". In: DOI: https://doi.org/10.23640/07243.5458162.v1.

Haerter, Dagmar (2008). *SCOPUS mit Affiliation Identifier – medinfo*. URL: http://medinfo.netbib.de/archives/2008/04/28/2574 (Abgerufen am 24. 01. 2018).

Hahnen, Holger und Güdler, Jürgen (2008). „Quality is the Product is Quality – Information Management as a Closed-loop Process". In: *Get the Good CRIS Going - Ensuring Quality of Service for the User in the ERA*. Proceedings of the 9th International Conference on Current Research Information Systems (2008). (5.–7. Juni 2008). URL: hhttp://hdl.handle.net/11366/332.

Hammond, Max und Curtis, Goeff (2013). *Landscape Study for CASRAI-UK Organisational ID*. URL: http://repository.jisc.ac.uk/5381/1/CC549D001%2D1.0_org_ID_landscape_study.pdf.

Hendricks, Ginny (2015). *Best Practices for Depositing Funding Data*. URL: https://0-www-crossref-org.lib.rivier.edu/blog/best-practices-for-depositing-funding-data/ (Abgerufen am 03. 01. 2018).

Hennessey, Jason und Ge, Steven Xijin (2013). „A cross disciplinary study of link decay and the effectiveness of mitigation techniques". In: *BMC Bioinformatics* 14 (14). DOI: https://doi.org/10.1186/1471-2105-14-S14-S5.

Herwig, Sebastian und Schlattmann, Stefan (2016). „Eine wirtschaftsinformatische Standortbestimmung von Forschungsinformationssystemen". In: *Informatik 2016*. 46. Jahrestagung der Gesellschaft für Informatik. (Klagenfurt, 26.–30. Sep. 2016), S. 901–914. DOI: https://doi.org/10.5281/zenodo.291965.

Hilse, Hans-Werner und Kothe, Jochen (2006). *Implementing persistent identifiers: overview of concepts, guidelines and recommendations*. London: Consortium of European Research Libraries.

Hood, William W. und Wilson, Concepción S. (2003). „Informetric studies using databases: Opportunities and challenges". In: *Scientometrics* 58 (3), S. 587–608. DOI: https://doi.org/10.1023/B:SCIE.0000006882.47115.c6.

Hornbostel, Stefan und Olbrecht, Meike (2007). *Peer Review in der DFG: die Fachkollegiaten*. Bd. 2. iFQ-Working Paper. URL: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-7339 (Abgerufen am 27. 12. 2017).

Huang, Shuiqing u. a. (2014). „Institution name disambiguation for research assessment". In: *Scientometrics* 99 (3), S. 823–838. DOI: https://doi.org/10.1007/s11192-013-1214-2.

ISNI (2014). *ISNI Data Elements: Valid Values*.

ISNI (2018a). *About the ISNI International Agency*. URL: http://www.isni.org/about (Abgerufen am 17. 01. 2018).

— (2018b). *Data Contributers*. URL: http://www.isni.org/content/data-contributors (Abgerufen am 24. 01. 2018).

Jiang, Jing (2012). „Information Extraction from Text“. In: *Mining Text Data*. Boston: Springer, S. 11–41. DOI: https://doi.org/10.1007/978-1-4614-3223-4_2.

Joint Steering Committee for Development of RDA u. a., Hrsg. *RDA: resource description & access: deutsche Übersetzung*. 1. Auflage. Berlin, Boston: De Gruyter Saur.

Josenhans, Veronika (2011). *Hochschulbibliographie interoperabel*. Humboldt-Universität zu Berlin, Philosophische Fakultät I, Institut für Bibliotheks- und Informationswissenschaft. DOI: https://doi.org/10.18452/2040.

Jörg, Brigitte, Höllrigl, Thorsten und Sicilia, Miguel Angel (2012). „Entities and Identities in Research Information Systems“. In: *e-Infrastructures for Research and Innovation Linking Information Systems to Improve Scientific Knowledge Production*. CRIS2012: 11th International Conference on Current Research Information Systems. (Prag, 6.–9. Juni 2012). Hrsg. von Keith G. Jeffery und Jan Dvořák, S. 185–194. URL: http://hdl.handle.net/11366/102.

Klein, Martin u. a. (2014). „Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot“. en. In: *PLoS ONE* 9 (12). Hrsg. von Judit Bar-Ilan. DOI: https://doi.org/10.1371/journal.pone.0115253.

Klump, Jens und Huber, Robert (2017). „20 Years of Persistent Identifiers – Which Systems are Here to Stay?“ In: *Data Science Journal* 16. DOI: https://doi.org/10.5334/dsj-2017-009.

Kunze, John u. a. (2017). „Persistence Statements: Describing Digital Stickiness“. In: *Data Science Journal* 16. DOI: https://doi.org/10.5334/dsj-2017-039.

Lawrence, S. u. a. (2001). „Persistence of Web references in scientific research“. In: *Computer* 34 (3), S. 26–31. DOI: https://doi.org/10.1109/2.901164.

Lynch, Clifford (1998). „Identifiers and Their Role In Networked Information Applications“. en. In: *Bulletin of the American Society for Information Science and Technology* 24 (2), S. 17–20. DOI: https://doi.org/10.1002/bult.80.

Léargas (2018). *(Reasonably) quick PIC guide*. URL: http://www.gaelscoileanna.ie/files/Quick-PIC-presentation-KA1-030314.pdf (Abgerufen am 17. 01. 2018).

Mongeon, Philippe und Paul-Hus, Adèle (2016). „The journal coverage of Web of Science and Scopus: a comparative analysis“. In: *Scientometrics* 106 (1), S. 213–228. DOI: https://doi.org/10.1007/s11192-015-1765-5.

Orduña-Malea, Enrique u. a. (2015). „Improvements in Google Scholar Citations are for the summer: creating an institutional affiliation link feature“. In: *arXiv:1509.04515*. URL: http://arxiv.org/abs/1509.04515 (Abgerufen am 27. 12. 2017).

Orduña-Malea, Enrique u. a. (2017). „The lost academic home: institutional affiliation links in Google Scholar Citations“. en. In: *Online Information Review* 41 (6), S. 762–781. DOI: https://doi.org/10.1108/OIR-10-2016-0302.

Organization Identifier Working Group (2018). *ORG ID RFI Responses*. URL: https://docs.google.com/document/d/1g2L1dlxRps7_n10tVL0HQjCDsrIBJOhJ1kysYBdYcjM/edit#heading=h.gjdgxs (Abgerufen am 24. 01. 2018).

Patton, Glenn E. u. a., Hrsg. (2010). *Funktionale Anforderungen an Normdaten: ein konzeptionelles Modell ; Abschlussbericht Dezember 2008 genehmigt von der IFLA Cataloguing Section und der IFLA Classification and Indexing Section März 2009*. IFLA series on bibliographic control 41. Berlin, Boston: de Gruyter Saur.

Pentz, Edward u. a. (2017). „ORG ID WG Governance Principles and Recommendations". In: DOI: https://doi.org/10.23640/07243.5402002.v1.

Publications Office (2017). *CORDIS - EU research projects under Horizon 2020 (2014-2020)*. URL: https://data.europa.eu/euodp/en/data/dataset/cordisH2020projects (Abgerufen am 24. 01. 2018).

Putnings, Markus (2017). „Die Rolle der Metadaten – Indexierung und Sicherung der Auffindbarkeit". In: *Praxishandbuch Open Access*. Hrsg. von Konstanze Söllner und Bernhard Mittermaier. Berlin, Boston: De Gruyter Saur. DOI: https://doi.org/10.1515/9783110494068-036.

Putnings, Markus und Rusch, Beate (2016). „DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft". In: *o-bib. Das offene Bibliotheksjournal* 3 (4), S. 110–119. DOI: https://doi.org/10.5282/o-bib/2016H4S110-119.

Research Explorer (2018). *Tipps und Hinweise zur Suche nach Forschungseinrichtungen in Deutschland*. URL: http://www.research-explorer.de/tipps.de.html (Abgerufen am 24. 01. 2018).

Ringgold (2018a). *Hierarchies in Identify Online*. URL: https://support.ringgold.com/identify-online-hierarchy/ (Abgerufen am 17. 01. 2018).

— (2018b). *ISNI Registration Agency – Ringgold*. en-US. URL: https://www.ringgold.com/isni/ (Abgerufen am 17. 01. 2018).

— (2018c). *Ringgold Academic Classifications*. URL: https://support.ringgold.com/identify-classifications-explanation/ (Abgerufen am 17. 01. 2018).

— (2018d). *Ringgold Identifier*. en-US. URL: https://www.ringgold.com/ringgold-identifier/ (Abgerufen am 17. 01. 2018).

— (2018e). *Ringgold Tiers*. URL: https://support.ringgold.com/ringgold-tiers/ (Abgerufen am 17. 01. 2018).

— (2018f). *Ringgold Types*. URL: https://support.ringgold.com/ringgold-types/ (Abgerufen am 17. 01. 2018).

Schroeder, Kathrin (2009). „Persistent Identifier (PI) - ein Überblick". ger. In: *Nestor-Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Hrsg. von Heike Neuroth u. a. Version 2.0, [Stand:] Juni 2009. Boizenburg: Hülsbusch.

Smith-Yoshimura, Karen u. a. (2016). *Addressing the challenges with organizational identifiers and ISNI*. Dublion, Ohio: OCLC Research. URL: http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-organizational-identifiers-and-isni-2016.pdf (Abgerufen am 08. 12. 2017).

Spielberg, Eike T. (2017). „DuEPublicA: Automated bibliometric reports based on the University Bibliography and external citation data". In: *The Code4Lib Journal* 37. URL: http://journal.code4lib.org/articles/12549 (Abgerufen am 27. 12. 2017).

Taşkın, Zehra und Al, Umut (2013). „Institutional Name Confusion on Citation Indexes: The Example of the Names of Turkish Hospitals". In: *Procedia - Social and Behavioral*

Sciences 73 (Supplement C), S. 544 –550. DOI: https://doi.org/https://doi.org/10.1016/j.sbspro.2013.02.089.

Tonkin, Emma (2008). „Persistent Identifiers: Considering the Options“. In: *Ariadne* 56. URL: http://www.ariadne.ac.uk/issue56/tonkin (Abgerufen am 08. 12. 2017).

Vierkant, Paul u. a. (2017). „Workflows zur Bereitstellung von Zeitschriftenartikeln auf Open-Access-Repositorien – Herausforderungen und Lösungsansätze“. In: *O-Bib. Das offene Bibliotheksjournal* 4 (1), S. 151 –169. DOI: https://doi.org/10.5282/o-bib/2017H1S151-169.

Voigt, Michaela und Winterhalter, Christian (2016). *Open-Access-Anteil bei Zeitschriftenartikeln von Wissenschaftlerinnen und Wissenschaftlern an Einrichtungen des Landes Berlin*. DOI: https://doi.org/10.14279/depositonce-5570.

Walker, Lizzy und Armstrong, Michelle (2014). „Ï cannot tell what the dickens his name is”: Name Disambiguation in Institutional Repositories“. In: *Journal of Librarianship and Scholarly Communication* 2 (2). DOI: https://doi.org/10.7710/2162-3309.1095.

Walther, Marcus und Melsheimer, Bastian (2016). „Automated affiliation identification for Converis using Web of Science core collection data“. In: *CRIS2016: 13th International Conference on Current Research Information Systems*. (St. Andrews, 9.–11. Juni 2016). URL: http://hdl.handle.net/11366/547.

Wang, Jing, Smith-Yoshimura, Karen und Gatenby, Janifer (2015). *The complexity of scholar affiliation in ISNI and VIVO*. DOI: https://doi.org/10.6084/m9.figshare.2002122.v2.

Wiechmann, Brigitte (2014). „Normdaten“. In: *Praxishandbuch Bibliotheksmanagement*. Berlin, Boston: De Gruyter Saur.

Winterhager, Matthias, Schwechheimer, Holger und Rimmert, Christine (2014). „Institutionenkodierung als Grundlage für bibliometrische Indikatoren“. In: *Bibliometrie - Praxis und Forschung* 3. DOI: https://doi.org/10.5283/bpf.209.

Wissenschaftsrat (2016). „Empfehlungen zu einer Spezifikation des Kerndatensatz Forschung (Drs. 5066-16)“. Berlin. URL: https://www.wissenschaftsrat.de/download/archiv/5066-16.pdf (Abgerufen am 19. 12. 2017).

Zhao, Dangzhi und Strotmann, Andreas (2015). *Analysis and visualization of citation networks*. Synthesis lectures on information concepts, retrieval, and services 39. San Rafael: Morgan & Claypool.

A Anhang

OCLC *Addressing the Challenges with Organizational Identifiers and ISNI*	**Jisc Landscape** *Landscape Study for CASRAI-UK Organisational ID*	**Jisc/CASRAI** *Review of selected organisational IDs and development of use cases for the Jisc CASRAI-UK Organisational Identifiers Working Group*
Organization administrator: Diffuse accurate organization description Normierung Namensangabe	Research attribution Tracking von Outputs/Aggregation (Forschende/Forschungsinfrastrukturen)	OrgID-UC1 -Researcher -applying for funding Antragsellung (Forschungsförderung)
Funder and publisher: minimize conflicts of interest Gutachtenverfahren Peer Review	Automatic allocation of Article Processing Charges (APCs) APCs/Billing/OA Monitoring	OrgID-UC2 -Funder -minimising conflicts of interest Gutachtenverfahren Peer Review
Publisher: accurately identify researchers' affiliations in publications Affiliation/Korrekte Daten	BI for Research Performance Organisations (RPOs) Reporting/FIS	OrgID-UC3 -Funder- tracking published outputs Tracking von Outputs/Aggregation (Forschungsförderer)
Funder: track published output Tracking von Outputs/Aggregation (Forschungsförderer)	BI for research funders and regulators Reporting/Tracking von Outputs/Aggregation (Forschungsförderer)	OrgID-UC5 -Researcher or research manager -reporting academic impacts to funders Reporting/FIS/Tracking von Outputs/Aggregation (Forschende/Forschungsinfrastrukturen)
Academic administrator: aggregate the scholarly output of all my institution's researchers and selected subsets Tracking von Outputs/Aggregation (Forschende/Forschungsinfrastrukturen)	BI for publishers Tracking von Outputs/Aggregation (Verlage)	OrgID-UC6- Researcher -tracking organisations across time Affiliation

Researcher: ensure that all my output is accurately attributed to me, regardless of which institutional affiliation I may have had at the time Tracking von Outputs/Aggregation (Forschende/Forschungsinfrastrukturen)	Reporting to research funders Reporting/FIS/Tracking von Outputs/Aggregation (Forschende/Forschungsinfrastrukturen)	OrgID-UC7- Repository manager -populating repositories, managing automation Atomatisierung von Datenflüssen
Researcher: track organizations across time Affiliation	Reporting to other regulators Reporting (Forschungsförderer)	OrgID-UC8- Developer- directory services Automatisierung/Single Sign On/Authentifizierung
Research group: include multiple organizations related to my proposal for funding Antragstellung	Finance and billing Billing	
Repository manager: populate the repository and manage automation Automatisierung von Datenflüssen	Grant management Antragstellung	
Metadata specialist: add identifiers to national authority records Normierung/Normdaten	Authentication and authorisation Single Sign On Systeme/Rechteverwaltung	
Software developer: match organizations in a linked data environment LOD		
Rights agency: manage rights for the correct organization Rechteverwaltung		

Stand: 28.12.2017		Suche nach: Affiliation (GFZ Potsdam)				
Gesamt IDs: 27						
Affiliation ID	**Name**	**Other name formats**	**Adresse**	**Documents**	**Authors**	**Patent results**
60030245	Deutsches GeoForschungsZentrum (GFZ)	Geoforschungszentrum Potsdam	Telegrafenberg, Potsdam, Germany	7,981	1,134	46
		Gfz German Research Centre For Geosciences				
		Geoforschungszentrum				
		Helmholtz Centre Potsdam				
		Deutsches Geoforschungszentrum Gfz				
		Geoforschungszentrum Potsdam (gfz)				
		Helmholtz-zentrum Potsdam				
		Gfz				
		German Research Centre For Geosciences				
		German Research Centre For Geosciences				
110115752	GeoForschungsZentrum Potsdam	Geoforschungszentrum Potsdam	Telegrafenberg, Potsdam, Germany	10	0	45
		Geoforsckungszentrum Potsdam				
		Geoforschungszentrum-potsdam				
105456575	GFZ German Research Centre for Geosciences	Gfz German Research Centre For Geosciences	Telegrafenberg, Potsdam, Germany	31	7	0
112097984	GFZ German Research Centre for Geosciences	German Research Centre for Geosciences (GFZ)	Telegrafenberg, null14473 Potsdam, Germany	19	4	0
		GFZ German Research Centre for Geoscience				
		GFZ German Research Centre for Geosciences				
108688040	Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences	Helmholtz Centre Potsdam Gfz German Research Centre For Geosciences	Geomicrobiology, Germany	18	7	0
		Helmholtz Centre Potsdam-gfz German Research Centre For Geosciences				
109682212	GFZ German Research Centre for Geosciences	Gfz German Research Centre For Geosciences	Oberpfaffenhofen, Wessling, Germany	17	1	0
113372539	GFZ German Research Centre for Geosciences	Gfz German Research Centre For Geosciences	C/o DLR Oberpfaffenhofen, Wessling, Germany	17	7	0
109854044	GFZ German Research Centre for Geosciences	Gfz German Research Centre For Geosciences	Telegrafenberg, Potsdam, Brandenburg, Germany	16	0	0
		Potsdam Gfz German Research Centre For Geosciences				

		Potsdam German Research Centre For Geosciences Gfz				
		Organic Surface Geochemistry Laboratory Gfz German Research Centre For Geosciences Potsdam				
		Gfz-german Research Centre For Geosciences Potsdam				
		Gfz-german Research Centre For Geosciences				
		German Research Centre For Geosciences Gfz				
		German Research Centre For Geosciences				
112580423	German Research Centre for Geosciences	German Research Center for Geoscience	Telegrafenberg Potsdam, Potsdam, Germany	15	2	0
		German Research Centre for Geoscience				
		German Research Centre for Geosciences				
101323541	GeoForschungszentrum Potsdam	Deutsches GeoForschungsZentrum Potsdam	Telegrafenberg, Potsdam, Germany	13	2	45
		GeoForschungsZentrum Potsdam				
106751051	GFZ German Research Centre for Geosciences	Gfz German Research Centre For Geosciences	Potsdam, Potsdam, Germany	10	1	0
106728208	GFZ	Gfz	Muencher Strasse 20, Wessling, Germany	8	0	20
113225302	GFZ German Research Centre for Geosciences Potsdam	GFZ German Research Centre for GeosciencesPotsdam	DESY/Hasylab., Hamburg, Germany	7	1	0
		GFZ German Research Centre for Geosciences Potsdam				
117075395	GFZ German Research Centre for Geosciences Potsdam	Gfz German Research Centre For Geosciences Potsdam	Germany	6	2	0
		German Research Centre For Geosciences (gfz) Potsdam				
		Gfz German Research Centre For Geoscience Potsdam				
		German Research Centre For Geosciences Gfz Potsdam				
110071501	German Centre for Geisciences, Climate Dynamics and Landscape Evolution	Climate Dynamics and Landscape Evolution	Telegrafenberg, Potsdam, Germany	5	2	0
		Section 5.2: Climate Dynamics and Landscape Evolution German Research Centre for Geosciences Telegrafenberg Potsdam				
		German Centre for Geosciences, Climate Dynamics and Landscape Evolution				

117144899	GFZ	-	Potsdam, Leuna, Germany	5	2	20
114559295	GeoForschungsZentrum Potsdam	Geoforschungszentrum Potsdam	Adolf Schmidt Geomagnetisches O., Niemegk, Germany	4	3	45
112494449	GeoForschungsZentrum Potsdam (GFZ)	GeoForschungsZentrum Potsdam (GFZ)	Wessling, Germany	3	1	19
109901290	German Research Center for GeoSciences GFZ	German Research Center for Geosciences - GFZ	Berlin, Germany	3	2	0
		German Research Center for GeoSciences GFZ				
		German Research Center for Geosciences GFZ				
114765870	GeoForschungsZentrum Potsdam	Geoforschungszentrum Potsdam	C/o DLR, Munchner Str. 20, Wessling, Germany	3	1	45
106013964	Geraman research Centre for Geosciences	Geraman Research Centre for Geosciences	Telegrafenberg, Potsdam, Germany	3	1	0
		GFZ German Research Centre for Geosciences				
		Research Centre for Geosciences				
112025166	Deutsches GeoForschungsZentrum	Deutsches GeoForschungsZentrum	Telegrafenberg, Germany	2	0	20
107592761	GeoForschungsZentram Potsdam	GeoForschungsZentram Potsdam	Teiegrafenbcrg, Potsdam, Germany	2	2	19
		Deutsches GeoForschungsZentrum Potsdam				
114766739	GeoForschungsZentrum Potsdam	Geoforschungszentrum Potsdam	Am DESY, HASYLAB, Notkestr. 85, Hamburg, Germany	2	0	45
114995879	GeoForschungsZentrum Potsdam	Geoforschungszentrum Potsdam	Lessingstrasse 20, Heidelberg, Germany	2	0	45
118857660	GeoForschungsZentrum Potsdam	Geoforschungzentrum Potsdam	Germany	2	0	45
		Geoforschungzentrum Potsdam				
114873998	Helmholtz Centre Potsdam Gfz German Research Centre for Geosciences Potsdam	Helmholtz Centre Potsdam Gfz German Research Centre For Geosciences Potsdam	Germany	2	1	0
116754994	Helmholtz-Zentrum Potsdam GFZ	GFZ Helmholtz-Zentrum Potsdam	Germany	2	0	20
		Helmholtz-Zentrum Potsdam GFZ				

2018 Organization ID Stakeholder Meeting

ORG ID RFI Responses -- fulltext

1. **Library and Information Centre**, Hungarian Academy of Sciences
    a. www.konyvtar.mta.hu,
    b. Andras Holl
    c. holl.andras@konyvtar.mta.hu
    d. What is your organization interested in: providing open data
    e. Describe your organization's interest in organizational identifiers: We run the national scholarly publication database, which has scientometric use.Institutional evaluation and financing depends on publication output. Proper affiliation needs appropriate identification.",Hungarian RPO historical and hierarchical data. Experience with handling a fluid institutional landscape.,

2. **re3data** (DataCite)
    a. http://re3data.org
    b. Michael Witt
    c. mwitt@purdue.edu,
    d. What is your organization interested in: providing open data
    e. Describe your organization's interest in organizational identifiers: We maintain a registry of research data repositories.,"Each repository is associated with at least one organization. USE CASE: We would be interested in capturing / linking / sharing a common, open identifiers for these organizations for each data repository.

3. **California Digital Library**
    a. http://cdlib.org
    b. Guenter Waibel
    c. guenter.waibel@ucop.edu,
    d. What is your organization interested in: participating in a governance role; serving as technical and/or administrative host for the Registry organization; providing technology, staffing or marketing resources
    e. Describe your organization's interest in organizational identifiers: CDL has a long tradition in the identifier landscape through thought leadership, products/services, and community service. We also have connections throughout the academic institution and research community and understand the need for open, community-led organization identifiers. Our entire organization including our digital curation group, University of California Curation Center, are champions of networked research and the need to leverage persistent identifiers in the scholarly comms space.","CDL is interested in continuing our involvement in the Org ID project in either a governance role and/or as a host organization. We are willing to offer our expertise and our connections within the library/institution community to the success of this effort. We have experience in hosting administrative structures for regional/national/global efforts and see no problem with

allowing Org ID flourish as its own initiative within CDL. That being said, as a part of an academic institution, we do have resource constraints. For us to be able to host the Org ID effort we would need the project to be fully funded through grant or gift funds. We would also need the continuing support of the three steering organizations: ORCID, DataCite, and Crossref and we would need to better understand the short- and long-term governance plans. However, based on reading and discussing the outputs of the OrgID working groups, we see this as all possible/doable.

f. Describe the resources your organization could bring to the development of the open organization identifier registry: Our proposal is to be the host for Org ID, while maintaining extensive connections to the steering organizations. We would offer portions of our existing staff time to support the effort as needed. Before we got started, we would look for gift, grant, or in-kind contributions to fund any additional headcount required. And regardless of the decision on hosting, we are very interested in continuing our involvement in governing the new Org ID effort. Of course, before we could move forward on anything, we would also need to discuss all of these elements in more detail. Looking forward to it. Please let me know if you have any other questions.

4. **OCLC**
   a. http://oclc.org
   b. Jody DeRidder
   c. jody_deridder@oclc.org,
   d. What is your organization interested in: participating in a governance role; serving as technical and/or administrative host for the Registry organization; providing technology, staffing or marketing resources
   e. Describe your organization's interest in organizational identifiers: As a non-profit organization devoted to serving the library community, OCLC has demonstrated its commitment to cooperative cataloging, and sees cooperative identity management as a natural extension to existing services.

      OCLC already maintains the WorldCat Registry, a service for libraries and consortia, in which the organization has access to maintaining their own profile. We are very interested in leveraging our knowledge and insights to efficiently manage other types of organizational identifiers and their relationships to researchers, research products and publications, as this would improve access, context, and information retrieval for libraries, their patrons and the entire research community.

      OCLC also serves as the host for ISNI Institutional Identifiers and has thus demonstrated the capabilities for managing this type of information at an international level according to the specifications of the governing guidelines provided by others. Similarly, OCLC has experience in shared governance in external projects, as a member on the ISNI Council.

      As host to VIAF, OCLC has demonstrated the ability to provide unique, persistent IDs and information in both human and machine-readable formats, which are openly available, to support linked data and other applications.

   f. Describe the resources your organization could bring to the development of the open organization identifier registry: As a non-profit, library cooperative funded by our memberships use of shared services and programs in over 120 countries, OCLC has

over 50 years of experience managing metadata at a large scale. We have resources to develop and maintain metadata required to uniquely identify organizations and the relationships they have with researchers and other organizations.

Specifically, we are interested in:
--serving as a host for an organization registry under the auspices of a governing board
--participating as a member on the governing board
--providing support to openPIIR under a contractual agreement that could include incorporation of seed data from organization identifier providers who serve the research community, whose data meet the metadata requirements, and which data are available under a recommended Open Definition conformant license.
implementation of open criteria and documented processes for inclusion/exclusion, creating, merging, and deprecating an openPIIR record.

--openPIIR staff will have authority to enforce record management criteria
--OCLC will provide openPIIR record change tracking and recording using an open provenance model
--OCLC will support openPIIR staff curation, and the ability for organizations to curate their own information
--APIs to enable cross-talk between registry providers as appropriate

In reference to the proposed budget for phases 1-3 (per the Governance document), OCLC could assist with Technical Lead, Technical Contracting, Data Support and Technical Infrastructure resources.

5. **ORCID**
    a. https://orcid.org
    b. Laure Haak
    c. l.haak@orcid.org,
    d. What is your organization interested in: participating in a governance role; providing technology, staffing or marketing resources
    e. Describe your organization's interest in organizational identifiers: Organization identifiers are a key component of the ORCID vision and our trust program: the ability and responsibility for employers, educators, funders, associations, and publishers of researchers to assert affiliation, associating the personâ€°Ã›Âªs ORCID iD with the organization iD and stamping that relationship with a provenance link. That assertion, in turn, can be shared by the researcher as they interact with organizations in their research workflows. We started using organization IDs (including Ringgold, Funder Registry, and GRID) in 2013 for our users and members; each member of ORCID uses an organization ID when they make assertions into the ORCID Registry.
    f. Describe the resources your organization could bring to the development of the open organization identifier registry: We are primarily interested in serving in a governance role, to provide oversight in shaping the Registry features and services. In addition, we have developed some technology to manage organization identifiers for our members that may be useful for the Registry, to enable a front-end interface for organizations to manage their own record. Finally, given our broad user and member community that encompasses universities, funders, publishers, and researchers (and vendors who serve

these constituents) we may be able to assist with outreach and engagement toward adoption of Registry services.

g. Additional comments: ORCID has served as a member of the Organization ID Steering Committee and Working Group, has committed resources to these activities including managing the Product subcommittee, and looks forward to continued involvement in the initiative."

6. **Global Legal Entity Identifier Foundation** (GLEIF)
   a. www.gleif.org
   b. Stephan Wolf, CEO
   c. Stephan.Wolf@gleif.org
   d. What is your organization interested in: providing open data
   e. Describe your organization's interest in organizational identifiers: In 2011, the Group of Twenty (G20) called on the Financial Stability Board (FSB) to provide recommendations for a global Legal Entity Identifier (LEI) and a supporting governance structure. This led to the development of the Global LEI System which, through the issuance of LEIs, now provides unique identification of legal entities participating in financial transactions across the globe.

      The Legal Entity Identifier (LEI) is a 20-digit, alpha-numeric code based on the ISO 17442 standard developed by the International Organization for Standardization (ISO). The LEI code connects to key reference information that enables clear and unique identification of legal entities participating in financial transactions.

      Although the LEI had its origin is financial regulatory reporting, LEIs can be used in any use case in which a legal entity needs to be uniquely and persistently identified.

   f. Describe the resources your organization could bring to the development of the open organization identifier registry: The Global LEI System can be leveraged by the openPIIR as the only global free online source that provides open, standardized and high quality legal entity. By leveraging the LEI as the unique identifier, can save costs by eliminating duplicate data collection.","The RFI notes that the ORG ID working group is interested in ideas about business models for sustaining the Registry. Please find the following ideas from GLEIF regarding the cost recovery model of the Global LEI system. In the Global LEI System, generally costs are recovered by Local Operating Units (LOUs), also called LEI issuers, via fees paid by the legal entity to the LOU for LEI issuance or annual renewal. LOUs are able to structure their fee model as long as the LOU complies with the cost recovery principle. For example, one LOU does not charge any fee to legal entities registering for LEIs and instead cost is covered by a regulator in that country. Other fee models also could be considered such as a ""zero fee for the legal entity"" model. Firms that largely benefit from the operational improvements of having the LEI as a primary identifier could fund LEI issuance. GLEIF suggests an identifier system could use either model, one based on payment by the entity requesting the identifier or one based on the entity primarily benefiting from the standardization. GLEIF receives its funding from fees paid by the LOUs. The fee is calculated based on the number of revenue-generating LEIs maintained by the LOU. The structure of the Master Agreement between GLEIF and each LOU, allows flexibility as annually the contribution fee

structure and amounts are determined by GLEIF in consultation with its governing and oversight body, the LEI Regulatory Oversight Committee (LEI ROC).

7. **Australian National Data Service**
    a. ands.org.au
    b. Adrian Burton and Natasha Simons
    c. adrian.burton@ands.org.au
    d. What is your organization interested in: providing open data
    e. Describe your organization's interest in organizational identifiers: We can support this initiative through: (1) awareness raising in the research and research support community; (2) undertaking pathfinder projects/ demonstrator integrations in our own systems (as we do/have done for ORCID); (3) including this initiative in the Identifier agenda of our funded research technical infrastructure development projects and partnerships.,"Organizations play a role in managing, storing, identifying, describing and sharing the research data created by their affiliated researchers. An open, independent, non-profit organization identifier registry would facilitate the disambiguation of researcher affiliations and help identify and link the custodianship of data collections.
    f. Describe the resources your organization could bring to the development of the open organization identifier registry: ANDS plays a leadership role in the adoption of PiDs for research data and related materials in Australia. We do this by providing technical infrastructure including Handle and DOI minting services, by building relationships with key stakeholders and facilitating discussions, and via a strong community engagement program, such as that encouraging ORCID uptake.

       Community awareness raising of this initiative, and pathfinder/demonstrator integrations in our own systems, could be done using our own resources in collaboration with the organization registry.

       Research organizations contribute records about research data collections to the ANDS Research Data Australia discovery portal, along with related information (people and the organisations/groups they are affiliated with, activities/projects, services). This information is openly available free of charge and could be used by an organization registry.
    g. Additional comments: Happy to discuss further.

8. **DataCite**
    a. www.datacite.org
    b. Patricia Cruse
    c. patricia.cruse@datacite.org,
    d. What is your organization interested in: participating in a governance role
    e. Describe your organization's interest in organizational identifiers: DataCite is a non-profit working with over 1300 data centers worldwide. Our community is diverse and includes government institutions, non-government institutions, national libraries, large and small research centers, and for-profit organizations. DataCite is a leading provider of persistent identifiers for users and creators of research. Regardless of their provenance, the DataCite community has identified the ability to associate ORG IDs with DataCite content as a critical need. Connect research outputs to other research objects and people and organizations is a top priority for our diverse community.

f. Describe the resources your organization could bring to the development of the open organization identifier registry: DataCite brings knowledge and expertise on how to grow a young organization into a trusted community service provider. An important factor in launching the new ORG ID initiative will be the simultaneous launch of a governance structure that is responsive, representative, trustworthy and inclusive. DataCite is at the ready to lend our expertise and community voices to the effort. ORG ID's stated principles and Governance Recommendations are in complete agreement with DataCite's vision, mission, and values.
g. Additional comments: We love identifiers.

9. **euroCRIS**
    a. www.eurocris.org
    b. Juergen Guedler
    c. juergen.guedler@dfg.de
    d. What is your organization interested in: data model (CERIF); QA-aspects. Also sent separate email with document outlining EuroCRIS vision,euroCRIS is an international not-for-Profit association that brings together experts on Research Information in General and Research Information Systems (CRIS) in particular. ORG-IDs are one of the fundaments of such Systems.,
    e. See separate document, sent to l.haak@orcid.org at Nov, 30th:

10. **Foundation for Science and Technology** - National Scientific Computation (FCCN) branch
    a. https://www.fccn.pt/
    b. CÌÁtia Laranjeira
    c. laranjeira@fccn.pt
    d. What is your organization interested in: providing open data
    e. Describe your organization's interest in organizational identifiers: PTCRIS is a program launched by the Portuguese Foundation for Science and Technology (Fundacao para a Ciencia e a Tecnologia - FCT, FCCN unit) with the aim of ensuring the creation and sustained development of a national integrated information ecosystem to support research management. Data flow in such an open and integrated information ecosystem relies on unique, persistent identifiers to address unequivocally the different players of the academic and scientific system, namely organizations. For that reason, PTCRIS initiated the Org-Ids project with four major goals: 1) to create rules, principles and processes regarding registration of organizations in Portugal; 2) to build a first version of an authoritative National Organizations Database; 3) to develop an organization reconciliation service and 4) to set up a registration system for those organizations. The main outcomes of this work so far are the full specification of an organization registration system and a mandatory national organizations database that synchronizes information between the main national and international registration systems like Ringgold and ISNI.
    f. Describe the resources your organization could bring to the development of the open organization identifier registry: Our organization may contribute to the development of the open organization identifier registry by providing consulting services as well as effective data curation regarding Portuguese organizations. We may also provide a reconciliation service to map the organizational identifier system used within the registry's scope with other organization identifiers of relevant databases (e.g., Tax Identification Number, Participant Identification Code - a unique identifier for organisations participating in European Union funding programmes).

11. **Ringgold**
    a. www.ringgold.com
    b. Laura Cox
    c. laura.cox@ringgold.com
    d. What is your organization interested in: providing open data; participating in a governance role
    e. Describe your organization's interest in organizational identifiers: Ringgold is a specialist in Organisational Identification, with associated deep metadata and hierarchies across the scholarly communications environment. Ringgold is also an ISNI Registration Agency and provides the open ISNI ID for over 450,000 organisational entities.
    f. Describe the resources your organization could bring to the development of the open organization identifier registry: As an ISNI Registration Agency specializing in the identification of organizations, Ringgold provides a free open access service to the ISNI data for over 450,000 organizations that it holds. This includes:
    --An API to obtain and resolve existing ISNIs for organizations;
    --A complete dataset download of ISNIs, organization names, locations, alternate names, and URLs in both JSON and CSV; and
    --A free online look-up service to search and obtain ISNI records.

    For access to this service and for more information please see: http://isni.ringgold.com/

    Ringgold specialises in the identification of and curation of data about organistions involved in scholarly communications. In its capacity as an ISNI Registration Agency, Ringgold has operated this service since November 2016 as a conduit to ease access, increase adoption, and improve utility of ISNI for the good of the scholarly community. Ringgold currently operates the free service under the ISNI-IA open access license (http://www.isni.org/content/isni-international-agency-information-license) and will change the licensing terms in accordance with the planned changes to the ISNI license to CC0 when notified to do so by the ISNI-IA.

12. **Crossref**
    a. https://www.crossref.org/
    b. Ed Pentz
    c. epentz@crossref.org,"
    d. What is your organization interested in: providing open data; participating in a governance role; serving as technical and/or administrative host for the Registry organization; providing technology, staffing or marketing resources
    e. Describe your organization's interest in organizational identifiers: Addressing the affiliation name ambiguity problem is a strategic priority for Crossref and its members. This is so that we can add affiliation to the metadata we collect to be used by our members and all the services and tools that use our metadata. To do this we need an open, community governed organization identifier. Also, we want to combine the Open Funder Registry of 13,000 funders with this new registry. We recognize that the registry must address other stakeholder use cases. Here is our response to the Working Group principles:

    Governance:

1. <Adopt a hosted non-profit model so that there is independent governance for the Registry..> Crossref supports creating a non-profit joint venture collaboration to fulfill this recommendation and is prepared to act as the host organization with a suitable governing board.
2. <That the WG identify existing organizations that would be willing to act as a host..> Crossref is willing to act as the host organization under an appropriate governing body.
3. <That the WG identify existing organizations that would be willing to commit resources for phase 1..> Our board has authorized Crossref to provide resources (staff, financial, technical) for launching and running an independent registry.
4. <The new Registry organization should be non-profit..> Crossref supports a joint venture collaboration being established on a non-profit basis with stakeholder governance. An appropriate legal document (e.g. an MOU) can be used to create the joint venture collaboration with protections if it is ended.
5. <The host organization should be non-profit..> Crossref is non-profit, non-stock, transparently governed and well regulated as a US tax exempt entity. As a NY not-for-profit, if Crossref is wound up its assets must go to another non-profit with a similar mission.
6. <All software developed to run and support the Registry should be Open Source..> Crossref supports this and it can be part of the joint venture collaboration agreement.

Product:
1. <The openPIIR is intended for use by the research community..> Crossref is focused on the research community and we believe that for the service to succeed any data source used to seed the registry should also be focused on the scholarly research community.
2. <The openPIIR will derive utility through a hybrid staff- and organization-curation service model...>
3. <Access to organizations for managing openPIIR records shall be via permission..> Crossref has run the Open Funder Registry on a hybrid staff/organization curation basis and all our services support cross-talk between other providers. Crossref's current system manages permissions for depositing and updating metadata for over 8,500 member organizations.
4. <The openPIIR will focus on the organization levels that are most pertinent for the affiliation use case..> This is Crossref's primary use case and we believe - based on our own data analysis - that there are existing scholarly organization identifier registries that can be used to seed the registry.
5. <The openPIIR will require data elements...> Crossref has long experience in collecting metadata that uniquely identifies content and organizations and in assigning PIDs.
6. <openPIIR documentation and required data will be available for use under a recommended Open Definition conformant license, in human and machine readable formats.> Crossref already releases all of its Open Source software under an MIT license and we make copyrightable metadata available under Open Definition conformant licenses. This can be included in the legal document establishing the joint venture collaboration.
7. <The openPIIR will seek seed data from organization identifier providers who serve the research community..> the Open Funder Registry Data is available under an Open Definition Conformant license (CC0) and this should be a requirement for the new

registry (and any seed data that is used in the new registry).
8. <There will be open criteria and documented processes for inclusion/exclusion..> Crossref can ensure that all policy documentation for the registry's record are made available under a CC-BY licence. We have extensive experience overseeing metadata management policies for our 90 million+ metadata records.
9. <Record changes will be tracked and recorded using an open provenance model..> Crossref has extensive experience with this - our Event Data service provides open provenance data for millions of records and the Open Funder Registry deals with deprecation and other changes. We are also very experienced with community consultations to agree policies.
10. <The openPIIR will maintain a customer support ticketing system and an open knowledge base..> Crossref has extensive experience running a customer support ticketing system and providing an open technical support knowledge base for its own services.

f. Describe the resources your organization could bring to the development of the open organization identifier registry: Crossref is a well-established, financially stable organization with cash reserves and an established team of 34 staff. The Crossref board approved us providing resources to enable the startup and operation of a registry along with other stakeholders. The creation of the registry outlined in the Organization Identifier Working Group documents is a strategic priority for Crossref.

Crossref has a lot of experience (its own formation and running of infrastructure services and through having played a lead role in founding ORCID) that will be relevant to the development of the open registry. There are limited business model options to support an independent, open registry. The main models to review are:

--Membership fee - there are positives and negatives to this approach. Existing organizations already have membership fees and so we have some evidence of what works. However, there are already a range of organizations asking the community to pay membership fees and adding a new one - even if it's a joint venture rather than a new entity.
--Paid SLA API - combine a public API and a fee-based SLA API. An annual fee can be based on the size of the organization or the level of queries to the API. This annual fee based on organization size is the model Crossref is putting in place for 2018 and has the technical infrastructure in place and understands what would be required. The benefit of this is that those using the system pay for it and could be set up relatively quickly and more easily than a full membership offering as in number 1.
--Joint Venture Infrastructure Charge - have the members of the joint venture pay a fee based on a percentage of their own revenue to support the costs of the registry (For example, CERN runs a number of initiatives this way.) It would not be too difficult to see 10-15 organizations supporting operational costs. From Crossref's point of view this would recognize that this is infrastructure that will benefit our members and save our members the hassle of yet another membership or additional fees paid to the registry directly. ORCID and DataCite could also pay on behalf of their members as could other organizations.

Combination of 1, 2 and 3. Agree that 50% of registry costs will be funded out of the

infrastructure charge and that we would attempt to raise the remaining 50% with combination of: SLA fees, membership/support fees, grants (for startup, particularly)

Crossref is willing to comment financial, technical and staff resources to the organization identifier registry.

g. Additional comments: The launch and success of an independent, open organization identifier registry focused on the scholarly research space is a strategic priority for Crossref. The Crossref board unanimously approved the following points at its November 16th meeting:
- Crossref supports the Working Group principles and recommendations.
- Crossref is interested in being a host for the new registry.
- Crossref supports creating a joint venture non-profit collaboration of representative stakeholders to launch a registry. Crossref does not support creating a new entity (e.g. 501c3) at the start although this is something that could happen later.
- Crossref is interested in participating in the governance of the new registry
- Crossref can make significant resources available to help launch an independent registry.
- Crossref is well placed to undertake the above based on current experience running its own services and infrastructure including the Open Funder Registry.
- Crossref has developed an estimated expense budget and rationale for an independent registry that it is willing to share with the group of organizations that create the new joint venture collaboration.

Crossref has done a data analysis of existing organization identifier registries and verified that there is at least one data source that can be used for an open registry and meets the requirements of the governance and product principles of the Organization Identifier Working Group. Based on the analysis of different data sources (including current costs and rates of updating of records) and its own experience running the Open Funder Registry, Crossref has developed an initial expense budget for the launch and first year running costs of the registry of around $700k. We are happy to share the data analysis and expense budget with a new joint venture collaboration to move forward with standing up a community governed, independent, open registry.

In addition to solving our own use cases, Crossref recognizes that an identifier registry also has to address other stakeholder use cases. Crossref has demonstrated its commitment to a community effort by working with ORCID and DataCite in 2016 to review the organization identifier landscape - this involved extensive consultation and analysis and resulted in a set of three papers released at the end of October 2016 for public comment. Based on these papers we then supported setting up the Organization Identifier Working Group in 2017 that led to a set of governance and product principles and recommendations and this RFI process.

13. **Digital Science**
    a. https://www.digital-science.com/
    b. Andres Mori
    c. a.mori@digital-science.com

d. What is your organization interested in: providing open data; participating in a governance role; serving as technical and/or administrative host for the Registry organization; providing technology, staffing or marketing resources; Advising on initial policy decisions regarding the scope of the data held and the definition of an institution or organisation- i.e. who the registry would permit to register,
e. Describe your organization's interest in organizational identifiers: We are deeply involved in the scholarly communication space and have various needs for organisational identifiers - both internal and external (in our software). It is, in fact, this need that led us to develop our own solution (grid.ac). Noting that there was no open standard in the community at that time we were motivated to share our identifiers with the community. As both producers and consumers of organizational identifiers, we have a well-formed view of the challenges that come with developing, supporting and administering this type of project.
f. Describe the resources your organization could bring to the development of the open organization identifier registry: Digital Science and GRID - an open institutional identifier system

   Digital Science has developed GRID based on an internal need for solving the challenge to map scholarly outputs consistently to organisations. Specifically, the data are used across Digital Science's consultancy services and the products in its portfolio. As a conglomeration or cooperation of companies sharing certain services it is natural of Digital Science companies to collaborate on initiatives that benefit multiple companies in the portfolio. GRID was borne out of a common need to have a high-quality organisation identification system that could help to tie data together in many contexts. Today, GRID is used by Symplectic Elements, UberResearch Dimensions, Figshare, Altmetric, has featured in Digital Science Consultancy's projects and reports, and is even used in Digital Science's sales pipeline management software. Many external organisations (both for-profit and not-for-profit) have used the data for a range of purposes from research projects to data cleansing activities We took a pragmatic approach to developing policies around the data and took feedback from academic institutions governments, funders and external companies. Digital Science has invested over the years more than Â£2 million into the development of GRID. Since the beginning it was always the aim to make the efforts around GRID available to the community and Digital Science realised this first under a CC-BY license, switching to CC-0 based on feedback and input from the community. This is of course available to serve as basis for any new system.

   Digital Science and institutional identifiers going forward
   Due to the nature of the Digital Science products and services, Digital Science will need to continue a GRID-like undertaking even if an open alternative is developed and maintained externally. Most of the reasons are around speed to changes to policies, flexibility and ability to accommodate client use cases and needs. This means that Digital Science will have to continue to invest resources in this area, an open alternative is not eliminating this need. But we see the possibility to do this as an additional layer on an open identifier system.

   What Digital Science can contribute
   Digital Science has also developed a data curation platform to aid in the maintenance,

disambiguation and quality monitoring of data. Further, APIs and tools support an operating team of data curators - GRID is an ongoing effort. In addition to the public dataset, more data has been accumulated. This includes: mapped name variants of institutions. We could foresee scenarios where we make the APIs, the curation backend and processes and name variants available for the new effort, but given the massive investment we would need to explore scenarios which would allows to do this - mainly around saving the new organisation development and implementation efforts or performing as a contractor services - to have a counterweight of future revenues at least partially set against the investments.

The primary aim would be to enable that the new organisation would be fully functional, including the ability to provide APIs openly from day 1.

g. Additional comments: Digital Science is keen to participate and support an open and broad-scale industry initiative in this space. Our existing work in the field qualifies us to help any organisation involved in this activity to reach a solid and useful product that has the maximum chance to improve scholarly communications in a sustainable manner. We believe that issues of sustainability are well served by engagement with commercial for-profit providers such as ourselves, as has been demonstrated by the constitution of the ORCID and CrossRef boards. Digital Science has an established brand in open research through a number of initiatives including its donation of the technology and data for SureChEMBL, the creation of GRID and its support for Figshare.

14. **The British Library**
    a. https://www.bl.uk/
    b. Torsten Reimer
    c. torsten.reimer@bl.uk,
    d. What is your organization interested in: providing open data; participating in a governance role; serving as technical and/or administrative host for the Registry organization; providing technology, staffing or marketing resources
    e. Describe your organization's interest in organizational identifiers: Digital identifiers are critical for the global research ecosystem. Efficient workflows in scholarly communication and research information management depend on identifiers, and the use of robust and open IDs supports open science, research integrity and the discovery of and access to knowledge (and thereby the impact of research). As a national library with an international remit, the British Library is committed to supporting and sustaining international infrastructures and services that contribute to the improvement of research workflows and sharing of knowledge. We have set up a new Research Services department and a major change program (called Everything Available) to transform the Library's services to the global research community, and identifiers play a key role in this work. The British Library recognizes the importance of organizational identifiers alongside other IDs that refer to outputs, people etc.

    Supporting the development of an organization ID is directly relevant to the British Library's statutory obligation to provide information services that support research and also to its role in promoting the use of metadata standards. As an internationally trusted organization with a focus on longevity and persistence, we are interested in ensuring that the governance of an emerging organization identifier has a strong focus on the

long-term sustainability of that ID. The British Library is already playing a wider role in the development of persistent identifier and metadata standards infrastructure through its role as a Registration Agency within ISNI as well as through membership to DataCite, and hosting the UK ISSN Centre (http://www.bl.uk/bibliographic/issn.html).

The Library is already involved in a number of initiatives related to identification of organizations, specifically:
--Acting as the UK National Agency for the assignment of the International Standard
----Identifier for Libraries (ISIL) http://www.bl.uk/bibliographic/isilagency.html
--MARC Organization Codes (http://www.bl.uk/bibliographic/marcagency.html)

Involvement in organization ID development will further enhance this area of work. Our ISNI and DataCite UK services fed in to the outcomes of the Horizon2020 funded THOR project, led by the British Library. We are also a partner in THOR's successor FREYA, which will go on to support development of further requirements and wider advocacy for organization identifiers.

We have closely followed the recommendations of the Jisc-CASRAI study on Organizational IDs and have subsequently worked with Research Councils UK on a pilot to assign c.1000 ISNIs to fundable institutions. This work demonstrated that the BL could provide the support needed to deliver quality assured identifiers, but left open questions for funding full scale assignment to the whole RCUK databases and wider diffusion across CRIS systems in the UK. As for a broader organization ID, a sustainable cost model is needed, and is something that we are keen to explore further. We are currently modeling priced service options for handling individual or batch requests for ISNIs. We would also be interested in developing partnership models for funding.

f. Describe the resources your organization could bring to the development of the open organization identifier registry: As an organization with experience in supporting the incubation of a persistent identifier service, we have already provided in-kind contributions to the technical and administrative hosting of a new registration agency. We are also committed to using organizational and other IDs in our next generation of metadata services including the Linked Open British National Bibliography.

The British Library has much to contribute to the development of a functional, open and independent organization registry. The most important of these would be:
-- Experience of supporting persistent identifier infrastructure development
-- Strong track record in managing centers for international standards
-- Staff expertise and resources committed to ISNI operational development
-- Existing funds committed to technological development
-- Commitment to openness and the provision of open data services

The British Library already has staff with a responsibility for the management of services related to persistent identifiers. As well as staff responsible for providing DataCite UK, it has appointed an Identifier Development and Operations Manager to lead the development of its work with ISNI.

To support the international work of both DataCite and ISNI, the British Library has

provided additional staff where our relevant expertise was needed. At the inception of DataCite, the Library provided in-kind staff resource for the initial development of the DataCite infrastructure and metadata schema. This support allowed DataCite to grow its membership and demonstrate the utility of DOIs for research data and other outputs. For ISNI, we have provided quality assurance services, the Quality Team working closely with the ISNI Assignment Agency at OCLC on refining the matching rules for batch loads and setting the threshold for automated assignment. This provision of relevant expertise would be valuable in support of organization identifier development, resulting in similar advantages in terms of development, feedback and community advocacy.

Within the ODIN and THOR projects we have started to integrate global PID systems with each other, and our own systems, e.g. in the form of an ISNI-ORCID linker, or by embedding the DataCite citation formatter into our data.bl.uk service. We are committed to expanding technical deployment and interoperation of PID services and to promoting their use globally.

We are actively developing technical capability to support end-user interaction with records in the ISNI database. See detail on ISNI Registration Portal in the next section.

Currently the British Library is using sponsorship funds provided by the Mellon Foundation to build an ISNI Registration Service Portal to allow for the management of ISNI requests, both online and batch load. The service will provide a capability for logged-in users to search and edit existing records and request new ISNI records using a template request form, thereby introducing a more hybrid user-managed and curated approach. We are also using our involvement in the ISNI to transform the effort we have already invested in identity management through our Name Authority Files into providing shareable, open and persistent identifiers for persons and organizations. We have specific interest in leveraging this work to help support research information systems.

The British Library continues to make its data (https://data.bl.uk/) and metadata (http://www.bl.uk/bibliographic/datafree.html) available openly where possible, for instance the national bibliography as linked open data (http://bnb.data.bl.uk/), as well as providing metadata files for download from other services, such as UK theses (EThOS http://ethos.bl.uk/). The British Library has contributed its own authority (http://www.bl.uk/bibliographic/authority.html) records for organizations that are present in the ISNI dataset, where they have been matched and linked to records and identifiers from multiple sources, including c.50 national libraries' authority files. We also support a range of international metadata standards (http://www.bl.uk/bibliographic/service.html).

g. Additional comments: The British Library is keen to ensure that any emergent organization ID makes maximum use of existing systems and technology, where practical and appropriate. For instance, the new capabilities for interacting with ISNI that have been developed at the BL will provide the facility for organizations to edit their own IDs in the ISNI database. This development forms part of ISNI's roadmap for improvements to central database provision to support organization IDs.

Working with ISNI and its network of Registration Agencies and Members brings a ready-made platform and system for collaboration and the collective curation of a

Persistent Institution Identifier Registry. There is a network of active institutions that are working with and actively improving ISNI, which is interested in aligning its approaches with community-articulated organization identifier requirements. Several US research libraries who contributed to the OCLC Research Task Group on Organizations in ISNI (https://www.oclc.org/content/dam/research/publications/2016/oclcresearch-organizational-identifiers-and-isni-2016.pdf) are now members of a consortium (https://wiki.duraspace.org/display/PCCISNI/PCC+ISNI+Pilot+Home) that is working with ISNI. Other national libraries such as the Royal Library in the Netherlands and the National Assembly Library of Korea are also already working with ISNI to create more organization IDs. The British Library is also setting up Working Groups (http://www.isni.org/content/call-working-group-volunteers) to develop ISNI policy and best practice implementations with members of the Conference of European National Librarians. We are aware that euroCRIS and CASRAI are also interested in pursuing a solution involving ISNI based on National Libraries providing a trusted service. We are supportive of the role both organizations play in this environment, for example CERIF formats contributing to improved data quality and exchange of data about organizations and CASRAI's role in contributing software-neutral specifications for the semantics and business rules to share a common business understanding of organizations.

The ISNI-IA has responded to the growing interest in its own community of members as well as wider representations by groups such as Jisc and CASRAI by committing to a Roadmap of developments for organization identifiers. http://www.isni.org/content/isni-organizations-registry-identifying-organizations-scholarly-supply-chain

The British Library's view is that exploration of the funding model and governance issues associated with a brand new institutional registry could be brought to bear on a more robust collaboration across stakeholders to ensure that the solution it provides addresses the needs of the research community.

As an organization that must support anyone who wants to do research, our application of the researcher affiliation use case for organization IDs already goes well beyond the traditional academic organizations and related bodies. This broad scope already indicates a large community of potential users and so greater opportunities for sustainability. We are also aware of other use cases that are already actively looking for a solution, for instance we are working with the Publisher Licensing Association on use cases for the assignment of ISNIs to Publishers and Imprints to improve identification of rights owners in the publishing supply chain.

15. **CASRAI**
    a. https://casrai.org/
    b. David Baker
    c. dbaker@casrai.org
    d. What is your organization interested in: participating in a governance role; providing technology, staffing or marketing resources; providing added consultation on business requirements and rules. Also sent separate email document with more details about CASRAI dictionary and consulting processes.

e. Describe your organization's interest in organizational identifiers: As these new central registries come online they will also evolve to become central databases (profile systems) holding structured information (data) about the entity being uniquely identified. ORCID has grown to play this role for the Researcher entity and this new OrgID registry will likely come to play this role for the Organisation entity. These new ID/profile systems (whether for Persons, Organisations, Grants, or other key entities) will not be replacing local systems-of-record at institutions and funders around the world; so our interest is in ensuring that appropriate open standards (formal, informal, or defacto) exist to aid interoperability among all these systems storing and sharing data about organisations.
f. Describe the resources your organization could bring to the development of the open organization identifier registry: CASRAI chapters offer the following direct contributions to the success of your initiative:
-- Research Management End-user Consultations**. CASRAI chapters will convene & facilitate added consultations with research management end-users at funders and institutions in our national communities (e.g. target audiences for each national chapter, providing more granular information about pre-existing membership contacts). These consultations can augment your own outreach during requirements-gathering for the initial launch of your service as well as subsequent maintenance/upgrade releases. We will deliver an open specification of management/administrative use-cases and requirements for information about organisations. The spec will include recommended mandatory/optional data elements (incl. glossary definitions and business rules) for each use case plus any national or disciplinary extensions. This specification can inform your design and can also serve as an added user requirements input to your interfaces (UI or API).
-- Scoping Institutional User Buy-in and Requirements**. Institutional buy in needs to go beyond the library. Further, the value of the OrgID solution needs to be demonstrable not only to research administrators, but to ERP systems owners too, as they are the institutional unit that has most to gain and the most to do to make an OrgID work beyond the research office in any institution. Via consultation with the various units within our membership we would also offer technical expertise in higher education administration.
-- Policy Recommendations**. Provide input on recommended principles for what constitutes an organisation in the context of research information systems; what criteria must be met to be eligible for an org id (e.g., unique tax id? or legal entity for contract purposes). Also critical to clearly define different types of entity limits, e.g. the OrgID will declare the limits of the legal entity, the technical support (an inbox?), the operational support, access to resource (subscription access?), a brand, etc). The main reason for highlighting these different aspects is that they speak to different orders of conversation and the org that is identified will mean different things/have different limits in different contexts.
-- Formal Reference XSDs**. We will partner with EuroCRIS to develop a technical representation of the above-noted business specification in the new and simplified CERIF-XML format. This formal representation can inform the modelling of your transport exchange layer (XML or JSON). Together with EuroCRIS we would deliver a CERIF-XML XSD (XML Schema Definition) of the CASRAI Business Spec for Organisation Profile(s).","In addition to offering, with our partners, the above contributions, CASRAI strongly supports an overall approach that aligns with the ISNI/British Library/National Libraries model to solving the challenges of uniquely identifying organisations.

This submission has the endorsement of the Steering Committees of all CASRAI chapters (Canada, UK, Europe) and was prepared by the following representatives from the CASRAI community:
- Owen Roberson (UK), Research Information Analyst, Research Office, University of Cambridge;
- Geoff Harder (Canada), Associate University Librarian, University of Alberta Libraries;
- Sushanta Mitra (Canada), Executive Director & Professor, Waterloo Institute for Nanotechnology, University of Waterloo;
- David Baker (Canada), Executive Director, CASRAI International;
- Gail Clement (USA), Head of Research Services, Library, California Institute of Technology;
- Joao Mendes Moreira (Portugal/EU), Head of Scientific Information, Foundation for Science and Technology;
- Kristoffer Lindell (Sweden/EU), Project Manager, Umea University / Swedish Research Council"

16. **Rhine-Westphalian Library Service Centre** (hbz)
    a. https://www.hbz-nrw.de/
    b. Adrian Pohl
    c. pohl@hbz-nrw.de
    d. What is your organization interested in: providing open data; serving as technical and/or administrative host for the Registry organization; providing technology, staffing or marketing resources
    e. Describe your organization's interest in organizational identifiers: We are currently providing two organization registries: lobid-organisations & OER World Map
    lobid-organisations is a registry of libraries in German-speaking countries. It comprises entries for more than 20,000 organizations. The data is provided for end users via an HTML interface and via a JSON-LD-based API (that largely uses schema.org as its vocabulary). The data isn't curated in this system though, but comes from the German ISIL registry and the German Library Statistics (DBS). For more information on the dataset see http://lobid.org/organisations/dataset.

    The OER World Map collects and visualizes data on the growing number of actors and activities in the field of open education worldwide. Besides persons, events and projects, one type of data we collect is information on organizations involved in OER initiatives. The OER World Map is a read/write system with a JSON-LD API and data versioning. See the website at https://oerworldmap.org/ and the list of organizations under https://oerworldmap.org/resource/?q=&filter.about.%40type=Organization.

    Also, as a library institution the hbz has a lot of experience describing organizations as part of the German Integrated Authority File (GND, for which we also provide a service that is currently in beta at http://lobid.org/gnd), the central authority file for German libraries. Thus, we have a lot of experience with describing organizations and the problems that come up for an organization registry over time.

    Together with graphthinking (http://graphthinking.com), the hbz is doing further work on the underlying software of the OER World Map for another project where we build a

Linked-Data-based ""Customer Relationship"" tool for the hbz. Thus, we also have a lot of experience regarding the web standards and technology to run a web-based registry.

To sum it up, we think the hbz together with graphthinking would be a good fit to help developing an organization ID registry.

f. Describe the resources your organization could bring to the development of the open organization identifier registry: We could bring into the project expertise on data models and vocabularies and -- together with graphthinking -- could take part in the development of the technical basis for the registry. Furthermore, we could host the start-up version of the registry in our computer center.",

17. **IP-Registry** (PSI Ltd)
    a. https://app.theipregistry.org
    b. Andrew Pitts
    c. andrew@publishersolutionsint.com
    d. What is your organization interested in: Please see the full response below (sent email with data schema)
    e. Describe your organization's interest in organizational identifiers: In 2009, at the request of our customers, PSI began development of a proprietary system for identifying and resolving widespread problems/errors included in the global IP Address databases that govern access to publisher electronic content. A major part of this initiative required that we create a means of cross-referencing data from a large number of publishers held within a variety of different systems. We soon realised that the industry lacked a standard way to name Organisations. As a result, before we started reviewing IP Address data, we created a set of exact naming standards along with a set of standard data attached to an automatically generated, unique six digit Organisational ID. The schema for an organisation is included below in appendix one. The database stemming from that that effort --The IP Registry (http://theipregistry.org) -- includes records for more than 60,000 institutional libraries (government, academic, & corporate) and holds more than 1.5 billion vetted IP Addresses associated with same. The IP-Registry is the industry's only database of standardized Organisational IDs that includes an English standard name, local name, aliases and acronyms for every organisation utilizing electronic content. The IP-Registry allows libraries to register for free on the web and amend their own records, with PSI verifying changes to ensure the standards are adhered to. The database is searchable using Latin alphabets including diacritics, Arabic and Brahmic scripts as well as Chinese characters and derivatives. We also have parent to child relationships and affiliations and the ability for Organisations to add/amend or delete these relationships.

    The ethos behind the IP Registry is to facilitate communication between libraries and publishers and benefit the industry as a whole. To this end, we are unique in having our data curated by both Organisations and Publishers along with other important stakeholders, whilst maintaining the integrity of the data by utilizing PSI's standardized process of verifying all changes.

    The Registry is quickly becoming adopted by many Publishers, Organisations and Consortia and we believe that no other database will be as widely used in the near future due to the many enhancements we are bringing, including but not limited to enabling

Organisations to collect OA Publisher usage data, IP black lists, and compliance data. Publishers will also be able to reach all organisations globally by pushing data out to them via the Registry. Perhaps the best recent example of the IP-Registry's breadth of functionality is evidenced by the fact that we are now working with Research4Life to enable them to use the IP-Registry to be the place that validates access to their collections.

f. Describe the resources your organization could bring to the development of the open organization identifier registry: As part of the IP Registry, PSI has created an organisational identifier structure and allocated this Org ID to each of the 60,000 institutions within the IP Registry. The IP Registry is currently Live and is freely available to librarians.

   In line with the requirements of the openPIIR, the IP Registry is able to provide the following services:
   -- The functionality already exists within the IP Registry enabling libraries to review and/or edit their pre-populated records. The IP Registry therefore supports record self-management by organisations and has dedicated staff in place to manage records and approve inputs thus maintaining the integrity and uniformity of the database.
   -- The IP Registry has capacity to incorporate all necessary Org IDs. PSI is both willing and able to sync with all the relevant Org ID providers and has currently matched all IDs to Ringgold. Plans are already in place to incorporate relevant ISNI data into the IP-Registry.
   -- The IP Registry provides an existing, and solid framework upon which to adopt a larger set of functionalities to support the user type, metadata, organisation type, hierarchical and functional needs of the openPIIR project.
   -- An API (currently being developed) to communicate with library service providers. This technology can easily be expanded for the requirements of this project.
   -- Seeded data will not be required as the IP Registry is already populated with the required information. However, if the working group requires comparison with other data sources, PSI will be happy to work towards comparison and inclusion of this data.
   -- The metadata set for the IP Registry includes the minimal requirements as follows; physical location, electronic address, persistent Organisational ID, organisation name in official language and in Latin characters, and provenance. Further information including optional fields such as physical address, crosswalks to other IDs, variant names, and organisation sector is also available.
   -- PSI will also provide the staff resource to verify changes made to the Registry and also for questions about APIs and the downloading of the freely available dataset.
   -- PSI will provide a mirror copy of the database for the host and it will also provide them with a copy of the source code, to be held in escrow.

g. Additional comments: Responses to the Principles listed in the Product Principles and Recommendations document
   Principle 1: By effectively leveraging the existing content of the IP Registry, we are able to provide a database that is easily and freely available for use by the research community. The IP Registry already provides a unique organisational identifier that meets many of the openPIIR requirements.

Principle 2: The functionality already exists within the IP Registry enabling libraries to review and/or edit their pre-populated records, the IP Registry therefore supports record self-management by organisations. The IP Registry data can be synchronised with other relevant Org ID providers as required and already has been matched to Ringgold who are also going to provide us with the relevant ISNI identifier. We would advise standardising methods of data cross-walking for all organisational data.

Principle 3: This is already in operation with the Registry and PSI has dedicated staff in place to manage records and approve inputs thus maintaining the integrity and uniformity of the database. Evidence will be required to support significant amendments and what can and cannot be amended is already built into the database.

Principle 4: PSI will enhance the existing IP Registry publishing-focused organisational hierarchy with additional organisation level concepts, to enable us to meet the requirements of each affiliation use case, needed by openPIIR. We are ready to add the standards for funders from Cross-Ref into the database and create another code for them on the Org ID schema.

Principle 5: Please refer to the Schema below in Appendix 1

Principle 6: PSI are willing to license all relevant data within the IP Registry (excluding IP data) according to a CC0 license for the purposes of this project.

Principle 7: The IP-Registry will accept data from other providers as necessary to enhance the data but it should be kept in the same standard format and therefore will need to go though a clean-up process first.

Principle 8: Should PSI be chosen as the openPIIR provider, the IP Registry with all data necessary to comply with the Open PIIR, will be made open. Processes for inclusion/exclusion, creating, merging, and deprecating records will be fully documented. These are already used within the IP Registry and the user manuals have already been created. PSI has dedicated staff in place to manage records and approve inputs thus maintaining the integrity and uniformity of the database. Evidence will be required to support significant amendments but the governing board for OpenPIIR will be able to decide on any changes to the openPIIR records.

Principle 9: The IP-Registry will enable a history of all changes will to be kept within the Registry to a standard to be discussed; this will include details of the change, the source and the evidence provided where applicable.

Principle 10: This is a process which can be introduced as required. We already have a system for customer service to react to customer questions by email and phone.

Response to the Governance Recommendations document
-- The IP-Registry will open the parts of the Registry that are necessary to comply with the OpenPIIR project.
-- There will be the ability to download the data at no charge under a CC0 license to be decided by the Governors

-- There is no fee for Organisations to come into the Registry and amend their records with verification and checks carried out by IP-Registry staff
-- A mirrored copy of the Open Registry will be offered to the host, details to be arranged if this is adopted by the group
-- If chosen, the source code will also be held in Escrow to comply with the wishes of the Governance Breakout Group
-- A new governing board will be organised if we are accepted, and a new non profit, transparent, non stock company set up to handle the Open Registry. PSI has discussed same with Roy Kaufman Managing Director of New Ventures CCC; Paul Doda, General Counsel, Litigation & Compliance at Elsevier; Andrew Richardson, ex Managing Director and Vice President of Business Development Wolters Kluwer. We would also recommend inviting members from Cross-Ref, Datacite and ORCID.
-- If the company is wound up, protections will be put in place for the Open Registry parts to be passed to Cross Ref and for the mirrored copy to be opened and the source code removed from Escrow and passed to Cross Ref.

Business Model

We expect initial development costs to be minimal as a large part of the development has already been carried out in the creation of the IP Registry.

The IP Registry provides an existing, and solid framework that the openPIIR project could leverage to deliver the target solution without the requirement to initiate a -- from scratch--product build. The advantage of this approach would be that much of the heavy lifting with regards to organisational hierarchies has already been done. To support the openPIIR requirements, the existing solution would need to be expanded, but the level of change would be primarily focused on inclusion of metadata, user types and other organisations attributes and funding for this should be relatively easy to obtain as opposed to a significant capital drive necessary to fund the creation of a solution that encompasses the offerings already included in the IP-Registry.

The IP-Registry already solves problems for Organisations who have issues with the fact that their data, held in Publisher systems, is inconsistent and often incorrect -- causing problems such as incorrect usage statistics and duplication of accounts etc. Using the IP-Registry to distribute Open Access usage metrics directly to libraries, the distribution of Black Listed IP addresses, the use of the Registry to enable OA Publications to add Affiliations to research papers, and assisting the publishing industry in reducing author and peer review fraud make the IP-Registry the natural place to hold an OpenPIIR. Nowhere else can immediately allow over 60,000 Organisations to register and amend for free, their pre-populated Organisational details, add hierarchies and affiliations and manage their IP addresses.

PSI suggest that the necessary parts of the IP Registry to comply with the OpenPIIR, is also mirrored, and the existing solution evolved into the openPIIR solution. We can leverage our existing development team, who are familiar with the data structures and the specificities of managing this type of data and delivering a product for scalable levels of consumer use.

Taking into consideration how we could apply more progressive technologies to the openPIIR project, we would suggest that blockchain is considered as either a mechanism for change audit within the system, but also as a mechanism of change management.

The application of blockchain could be used to support an immutable 'audit leger' of change within the openPIIR system. A larger application of blockchain could be used to synchronise the management of organisational data - where each organisation manages their own blockchain leger and the openPIIR system is given read access to this leger.

PSI are open to discussing how this technology could be leveraged to form a transparent and persistent mechanism for self-management of organisational data.

Schema sent separately as not able to place a chart here"

18. **Clarivate**
    a. Analytics,clarivate.com
    b. Christopher Burghardt
    c. chris.burghardt@clarivate.com
    d. What is your organization interested in: participating in a governance role
    e. Describe your organization's interest in organizational identifiers: Clarivate provides research discovery and analytics primarily through the Web of Science and InCites platforms. Organizational identifiers are an important attribution link for our data products.,
    f. Describe the resources your organization could bring to the development of the open organization identifier registry: Clarivate has expertise in the collection, indexing, and use cases for organization data. We would like to participate initially in a governance role.
    g. Additional comments: We look forward to participating in future stakeholder meetings.

19. **ISNI International Agency** (ISNI-IA)
    a. www.isni.org
    b. Tim Devenport
    c. tim@editeur.org,
    d. What is your organization interested in: providing open data; participating in a governance role; serving as technical and/or administrative host for the Registry organization; Collaborating actively with other interested parties, with ISNI providing a platform, database and methodology for ensuring unique assignment of persistent identifiers and other stakeholder organizations supplying scholarly-domain expertise, business and technical governance, and advisory board direction.
    e. Describe your organization's interest in organizational identifiers: ISNI-IA proposes that the infrastructure envisaged in the RFI be implemented by leveraging the existing mature ISNI registry in combination with specialist registration services provided through existing ISNI Registration Agencies and new entities where specialist sector knowledge makes this appropriate.

       Follow changes announced earlier this year, ISNI-IA meets the governance requirements set out in the RFI. It was already a multi-stakeholder not-for-profit entity,

but, in addition, it will now offer core organizational data under a CC0 license. New APIs are planned to allow organizations and their representatives to manage their own records. Downloads of the registry will be available for those needing them. These and other changes will complete the creation of an ISNI Organizations Registry within the ISNI system.

A new advisory board with strong representation from the scholarly communications community will assist the ISNI-IA Board in tailoring the new service to the needs of that community.

ISNI (the International Standard Name Identifier -- ISO 27729) is an established standard supported by an organizational infrastructure, dedicated to providing unique and persistent public identifiers to parties (organizations and individuals) active in the creative supply chain. The ISNI International Agency (ISNI-IA) operates and promulgates the ISNI standard under license from ISO.

During operations over the past five years, ISNI identifiers have already been assigned to in excess of 650k organizational entities, the majority being in the scholarly and associated library domains. These identifiers are already being accessed and cited by a variety of users and ISNI-IA is currently preparing to open up access and usage still further, by making available ISNIs and core associated metadata in the form of Open Linked Data.

f. Describe the resources your organization could bring to the development of the open organization identifier registry:
-- An established infrastructure, specializing in knowledge and technical procedures for disambiguation and identifier assignment.
-- Significantly lower entry costs for the scholarly community that would otherwise have to fund any entirely new approach, in that the standard, database and surrounding software already exist.
-- A dedicated Quality Team that constantly reviews and monitors the algorithmic and manual functions that underpin operations.
-- Established partnerships with both a large technology provider and a series of registration agencies operating in a variety of domains.
-- An ISNI Board and flexible organizational infrastructure that can expand to encompass both numerical scaling and the appearance of new entrants in existing or novel domains.
-- A structure that allows ISNI itself to run as a non-profit operation, but with a sustainable economic model based upon both membership and low identifier-assignment fees.

g. Additional comments: Rather than seeking to run an entire service itself, ISNI-IA feels that the approach most likely to be of value to the scholarly community will be to partner with bodies with detailed knowledge of the workings and needs of particular players in the scholarly domain, whether these be functionally (libraries, funders, authors, researchers, publishers, etc.) or territorially defined. ISNI-IA has extensive experience in reconciling the different ambitions of actors in diverse sectors.

Experience in hosting an established and large-scale identifier assignment service is ISNI-IA's core expertise, complementing the proposals we understand to be under development by other interested parties, including ISNI Registration Agencies the British

Library and Ringgold, ISNI's technology provider and Board Member OCLC, as well as CASRAI, EuroCRIS and perhaps others.

In its capacity as an ISNI Registration Agency specializing in the identification of organizations, Ringgold already provides a free open access service to the ISNI data it holds for over 450,000 organizations, as a conduit to ease access, increase adoption, and improve utility of ISNI for the good of the scholarly community. The service includes:

-- An API to obtain and resolve existing ISNIs for organizations;
-- A complete dataset download of ISNIs, organization names, locations, alternate names, and URLs in both JSON and CSV; and
-- A free online look-up service to search and obtain ISNI records.

ISNI-IA is very supportive of work underway at another of its Registration Agencies, the British Library, providing additional facilities including an ISNI Registration Portal and acting as a focal point for its own activities as well as those of wide range of national libraries that seek to implement an ISNI-based solution to issues of organization identification.

Finally, and with the objective of striving for maximum interoperability, ISNI-IA also supports the role that the CERIF formats of EuroCRIS might play in contributing to greater data quality and standardised storage and exchange of data about organisations, as well as the role that CASRAI could play in contributing software-neutral specifications for the semantics and business rules required for all players in the ecosystem to share a common business understanding of organisations.

20. **Inter-university Consortium for Political and Social Research (ICPSR)**
    a. http://www.icpsr.umich.edu/
    b. Jared Lyle
    c. lyle@umich.edu
    d. What is your organization interested in: participating in a governance role
    e. Describe your organization's interest in organizational identifiers: An international leader in data curation and management, ICPSR curates, preserves, and disseminates over 10,000 data collections of original social science data for research, instruction, and policy evaluation. We are interested in unique, persistent organizational identifiers for several reasons, including:
    -Enhancing the organizational records embedded in our existing metadata.
    -Enabling data linkage within our collections, as well as to external resources.
    -Making administrative data interoperable.
    -Verifying researcher credentials to facilitate access to (multiple) restricted data sets.
    -Improving or replacing an existing agency identifier registry maintained by the DDI Alliance, whose secretariat is based at ICPSR.
    f. Describe the resources your organization could bring to the development of the open organization identifier registry: ICPSR encourages the establishment of a modified existing organization, as summarized in the Governance Recommendations, and is interested in serving on the Advisory Board. ICPSR can tap its international network of over 760 consortium members and more than 20 funding agencies to encourage strong

and early research institution and funder buy in. Additionally, ICPSR is informed by a broad user community, which is helpful for the recommended broad set of stakeholders in governance.",

21. **OSTI, DOE** (via email)
    a. Sherline Crystal
    b. sherlinec@osti.gov
    c. What is your organization interested in: I am interested in participating in the ORG ID working group. Please let me know where I can be of service. I currently work at OSTI as a liaison and with the data team and software policy. I can see the importance of ORG IDs for funding agencies, universities, publishers, etc.

22. **European Directory of Marine Organisations EDMO** (via email)
    a. https://www.seadatanet.org/Metadata/EDMO-Organisations
    b. Peter Thijsse,peter@maris.nl
    c. What is your organization interested in: Open Data,
    d. Describe your organization's interest in organizational identifiers: Recently I came across your initiative to create an Organisation Identifier directory. As potential source for this directory I would like bring to your attention the European Directory of Marine Organisations - EDMO.

       EDMO contains up-to-date addresses and activity profiles of research institutes, data holding centres, monitoring agencies, governmental and private organisations, that are in one way or another engaged in oceanographic and marine research activities, data & information management and/or data acquisition activities. Currently, EDMO lists and describes more than 4.000 organisations. The content is maintained via the NODCÃ•s (National Oceanographic Data Centers) that each have control over their national subset of organisations.

       EDMO has been developed and maintained as part of the SeaDataNet/SeaDataCloud projects in last 10 years, but it is now widely used throughout the European Marine domain, e.g. in EMODNet, AtlantOS, etc. As a result of actions in the ODIP project (www.odip.eu) where European partners cooperate with marine data management experts in USA and Australia you will find in EDMO also the main marine organisations from USA and Australia and beyond. A mapping to national codes as in use in USA and Australia is foreseen to be added soon.

       You will find access to the directory at:
       https://www.seadatanet.org/Metadata/EDMO-Organisations

       Please let me know if you have a need for further information.

23. **RNSR**
    a. RÃ©pertoire national des structures de recherché
    b. https://appliweb.dgri.education.fr/rnsr/,
    c. Xiaofeng (Xavier) Chen
    d. xiaofeng.chen@education.gouv.fr,

e. I'm the project manager of the french RNSR, RÃ©pertoire national des structures de recherche, the french national registry of reasearch structres. I'm interessted by your OrgID project. We are definitely interested in your next meeting in January.

24. **Deutsche Forschungsgemeinschaft (DFG)**
    a. Christian Fischer
    b. www.dfg.de
    c. See paper submitted via email

B Abbildungen

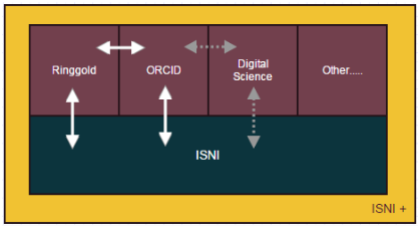

Abbildung 1: ISNI+ als hybrider Organizational Identifier. Aus: Ferguson, Moore und Schmoller, 2014, S. 19

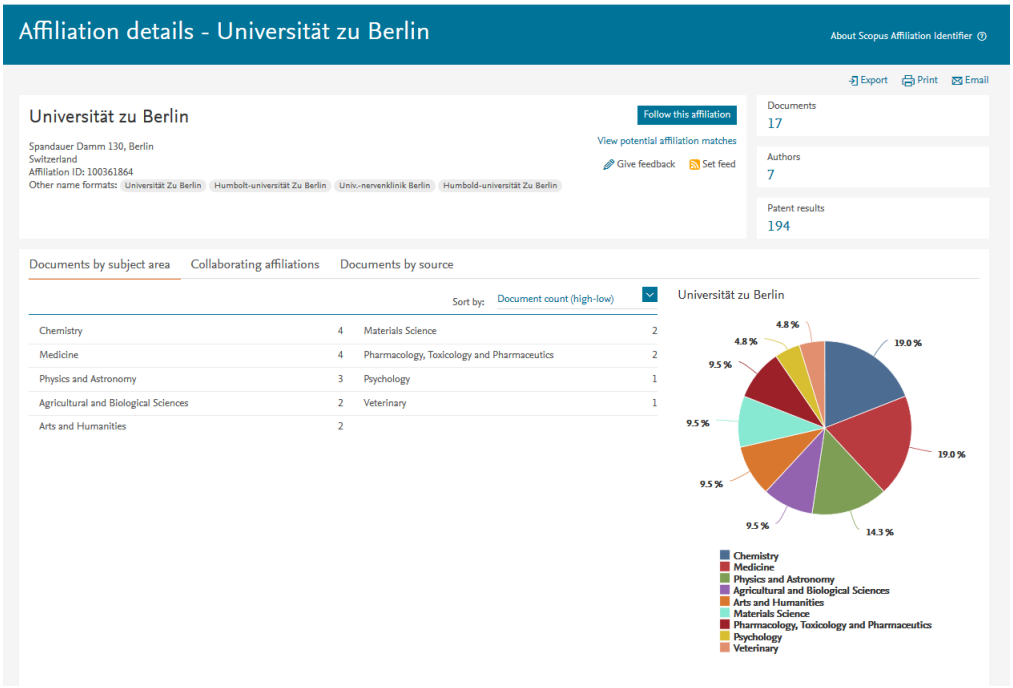

Abbildung 2: Affiliation Profile der *Universität zu Berlin* in Scopus