



**MITTEILUNGEN DER VEREINIGUNG  
ÖSTERREICHISCHER  
BIBLIOTHEKARINNEN & BIBLIOTHEKARE**



**70 (2017) 2**

**Schwerpunktthema:  
Metadata – Metadaten**

**e-infrastructures  
austria**

**ISSN 1022-2588**

**Redaktionsschluss für Heft 3/4 (2017): 31. September 2017**

## IMPRESSUM

### **Medieninhaber, Hersteller und Herausgeber:**

Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare

Geschäftssitz: Universitätsbibliothek Graz,

Universitätsplatz 3a, A-8010 Graz

Telefon: +43 (0)316 380 - 1419

E-Mail: [werner.schlacher@uni-graz.at](mailto:werner.schlacher@uni-graz.at), [voeb@ub.tuwien.ac.at](mailto:voeb@ub.tuwien.ac.at)

Website: <http://www.univie.ac.at/voeb>

### **Redaktionsteam:**

Stefan Alker, Monika Bargmann, Bruno Bauer, Patrick Danowski,  
Andreas Ferus, Andreas Heppinger, Michael Katzmayer, Peter Klien,  
Klaus Niedermaier, Otto Oberhauser, Josef Pauser und Bernhard Schubert

### **Redaktion (Schwerpunktthema):**

Susanne Blumesberger, Bernhard Schubert und Alexander Zartl

**E-Mail der Redaktion:** [voeb-mitt@uibk.ac.at](mailto:voeb-mitt@uibk.ac.at)

### **Elektronische Ausgabe unter:**

<https://ojs.univie.ac.at/index.php/voebm>

**Indiziert/indexed in:** BASE (Bielefeld Academic Search Engine), e-lis (e-prints in library and information science), DABI Datenbank Deutsches Bibliothekswesen, EBSCO® Library & Information Science Source, EBSCO® Library, Information Science & Technology Abstracts (LISTA), ProQuest® Library Science, Elsevier® Scopus, European Reference Index for the Humanities and the Social Sciences (ERIH PLUS)

### **Druck:**

Steiger Druck, Lindenweg 37, A-6094 Axams

Tel.: +43-5234-68105, Fax: +43-5234-68105/11

E-Mail: [steigerdruck@tirol.com](mailto:steigerdruck@tirol.com)

### **Preise:**

*Jahresabonnement der Mitteilungen ab 2007:* EUR 50,-; *Einzelheft:* EUR 15,-

*Anzeigenpreise:* 1/1 Seite: EUR 360,- (Teile entsprechend)

Beilage pro 1.000 Stück bzw. Gesamtauflage: pro Heft: EUR 360,-

Alle in den „Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare“ veröffentlichten Texte stellen die Meinung der Verfasser\_innen und nicht unbedingt die der Redaktion dar.



## ■ Editorial

|   |     |
|---|-----|
| Schwerpunktthema „Metadata – Metadaten“ ..... | 153 |
|---|-----|

## ■ Schwerpunktthema

|   |     |
|---|-----|
| <i>Anna Fensel</i> : Towards Semantic APIs for Research Data Services<br>(Auf dem Weg zu semantischen APIs für Forschungsdatendienstleistungen) .....   | 157 |
| <i>Walter Koch &amp; Gerda Koch</i> : Aggregation and Management of Metadata in the Context of Europeana (Aggregation und Management von Metadaten im Kontext von Europeana) .....  | 170 |
| <i>Georg Neubauer</i> : Visualization of Typed Links in Linked Data<br>(Visualisierung von typisierten Links in Linked Data) .....  | 179 |
| <i>Barbara Petritsch</i> : Metadata for Research Data in Practice<br>(Metadaten für Forschungsdaten in der Praxis) .....  | 200 |
| <i>Nina Rannharter &amp; Sarah Teetor</i> : Challenges of Building and Maintaining an Image Database: A Use Case Based on the Digital Research Archive for Byzantium (DIFAB)<br>(Herausforderungen beim Aufbau und bei der Pflege einer Bilddatenbank – Ein Anwendungsbeispiel aus dem Digitalen Forschungsarchiv Byzanz (DIFAB)) ..... | 208 |
| <i>Sándor Kopácsi<sup>†</sup>, Rastislav Hudak &amp; Raman Ganguly</i> : Implementation of a Classification Server to Support Metadata Organization for Long Term Preservation Systems (Implementierung eines Klassifikationsservers für Metadatenorganisation für Langzeitarchivierungssysteme) .....                                  | 225 |
| <i>José Luis Preza</i> : Data Science and Analytics in Libraries<br>(Data Science und Datenanalytik in Bibliotheken) .....  | 244 |
| <i>Susanne Blumesberger &amp; Alexander Zartl</i> : Umgang mit Metadaten in Repositorien – Eine österreichweite Umfrage (Handling Metadata in Repositories – An Austria-wide Survey) .....  | 249 |

## ■ Beiträge

|  |  |
|--|--|
| <i>Magdalena Andrae &amp; Márton Villányi</i> : Der Springer Compact-Deal – Ein erster Einblick in die Evaluierung einer Offsetting-Vereinbarung |  |
|--|--|

|   |     |
|---|-----|
| (The Springer Compact Deal – First Insights into the Evaluation of an Offsetting Agreement) ..... | 274 |
|---|-----|

## ■ Aus der VÖB

|   |     |
|---|-----|
| <i>Werner Schlacher: Aus Präsidium und Vorstand</i> ..... | 281 |
|---|-----|

*Eva Ramminger, Magdalena Andrae und Andrea Torggler:*

|   |     |
|---|-----|
| „Libraries.Solidarity.Society“ – Schritte zu einer Intensivierung der Kooperation zwischen den deutschsprachigen und internationalen Bibliotheksverbänden ..... | 284 |
|---|-----|

*Helmut Hartmann, Eugen Rott, Klaus Niedermair und Eva Ramminger:*

|   |     |
|---|-----|
| Dr. Eveline Pipp (12.12.1956–05.05.2017): Ein Nachruf ..... | 290 |
|---|-----|

## ■ Berichte

|   |     |
|---|-----|
| <i>Stefan Alker-Windbichler und Bruno Bauer: Treuhänderische Übernahme und Verwahrung – international und interdisziplinär betrachtet: Tagung zur NS-Provenienzforschung an der Universität Wien (Wien, 2.–4. Mai 2017)</i> ..... | 302 |
|---|-----|

## ■ Rezensionen

|   |     |
|---|-----|
| Bertot, John Carlo [ed. et al.]: Accessibility for Persons with Disabilities and the Inclusive Future of Libraries. Emerald Publishing Group 2015 (Advances in Librarianship 40).<br>( <i>Katharina Flicker</i> ) ..... | 314 |
| Hannah Frey: „Zuckerfrei – Die 40 Tage Challenge“. GU Verlag 2017.<br>( <i>Marlene Kersten</i> ) .....  | 316 |

|                                |     |
|--------------------------------|-----|
| ■ <b>Veranstaltungen</b> ..... | 317 |
|--------------------------------|-----|

# ■ AUTOR\_INNENRICHTLINIEN DER MITTEILUNGEN DER VEREINIGUNG ÖSTERREICHISCHER BIBLIOTHEKARINNEN UND BIBLIOTHEKARE

## Voraussetzungen für die Veröffentlichung

Neben dem Fachbeitrag, der einen substantiellen Beitrag zu einem Thema aus dem Bibliotheks-, Informations- oder Dokumentationswesen leisten soll, können unter anderem folgende Artikeltypen veröffentlicht werden: Editorial, Interview, Report, Bericht, Personalien, Rezension, Veranstaltungshinweis oder sonstige Mitteilung. Die eingereichten Manuskripte oder wesentliche Teile daraus dürfen nicht bereits an anderer Stelle veröffentlicht oder zur Veröffentlichung eingereicht worden sein. Die Präsentation der Inhalte als Kurzfassung auf einem Kongress gilt dabei nicht als Vorveröffentlichung.

## Autor\_innenschaft und Urheberrecht



Alle Autorinnen und Autoren versichern, dass sie einen substantiellen Beitrag zum Artikel erbracht haben und mit Form und Inhalt des Manuskriptes einverstanden sind. Mit der Einreichung des Manuskriptes und des sonstigen Materials bestätigt die Autorin oder der Autor, dass sie/er über die urheberrechtlichen Nutzungsrechte am Werk und den mitgelieferten Text- und Bildvorlagen verfügt. Die Autorin oder der Autor räumt im Fall der Veröffentlichung das Recht auf zeitlich unbegrenzte Einspeicherung in Datenbanken, Verbreitung und Wiedergabe des Beitrages in elektronischer als auch gedruckter Form ein. Für alle veröffentlichten Beiträge kommt die Creative-Commons-Lizenz Namensnennung 4.0 International zur Anwendung.

## Gestaltung der Manuskripte

Die Manuskripte der Fachbeiträge sollen in Deutsch oder Englisch eingereicht werden. Der Umfang soll zwischen 1.000 und 5.000 Wörter betragen. Längere Artikel sind nach Rücksprache möglich. Zusätzlich sollen die Manuskripte ein deutsches und englisches Abstract mit ca. 100 Wörtern sowie den deutschen und englischen Titel beinhalten. Weiters sollte die Autorin oder der Autor geeignete (deutsche und englische) Schlüsselwörter

ter angeben. Das Manuskript soll klar und übersichtlich gegliedert sein. Die Überschriften der Hauptabschnitte sollen möglichst kurz sein. Im Literaturverzeichnis ist die im Text zitierte Literatur in eindeutig nachvollziehbarer und konsistenter Form anzuführen. Als Grafikformate für die mitgelieferten Dateien können verwendet werden: die Formate TIFF und BMP (verlustfreie Bitmap-Formate); GIF und PNG (komprimierte Bitmap-Formate) für Schaubilder, JPG (komprimierbares Bitmap-Format) für Fotos. Auch wenn die Grafiken in den Text eingebunden sind, sollten sie zusätzlich als separate Dateien mit eindeutigem Dateinamen mitgeliefert werden.

## **Einreichung**

Beiträge sind an die E-Mail-Adresse des Redaktionsteams zu übermitteln: [voeb-mitt@uibk.ac.at](mailto:voeb-mitt@uibk.ac.at). Über die Aufnahme entscheidet das Redaktionsteam.

### ■ SCHWERPUNKTTHEMA „METADATA – METADATEN“

von *Susanne Blumesberger, Bernhard Schubert, Alexander Zartl*

Wir widmen diese Ausgabe unserem lieben Kollegen Sándor Kopácsi, der am 4. August 2017 unerwartet verstorben ist.

Die sichere und dauerhafte Archivierung und Bereitstellung von elektronischen Publikationen, Multimedia-Objekten und anderen digitalen Daten aus Forschung und Lehre ist eines der dringendsten Bedürfnisse des gegenwärtigen Forschungs- und Universitätsbetriebes. Um dieser Anforderung gerecht werden zu können, rief das Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFW) im Jänner 2014 das Partnerprojekt *e-Infrastructures Austria* ins Leben. Ziel des von der UB Wien koordinierten und auf drei Jahre angelegten Projekts war der Aufbau und die Weiterentwicklung digitaler Repositorien und forschungsunterstützender Services im Bereich der österreichischen Universitätslandschaft. Durch Vernetzung und Bündelung von Know-how und Ressourcen sollte ein Wissens- und Kompetenznetzwerk entstehen, das Hilfestellung bei der Implementierung von Repositorien und der Erweiterung technischer Systeme, Services und begleitenden Fragestellungen leisten kann. 25 Universitäten beteiligten sich an verschiedenen Arbeitsgruppen, Veranstaltungen und Workshops, die – unter Einbezug von Bibliotheken, IT-Services und Wissenschaftlerinnen und Wissenschaftlern der mitwirkenden Institutionen – einen regen Erfahrungsaustausch zu den technisch, organisatorisch und rechtlich relevanten Themen gewährleisteten.<sup>1</sup>

Die Aufgaben des Projekts wurden auf 12 nach inhaltlichen Gesichtspunkten strukturierten Work-Package-Cluster verteilt, von denen Cluster I sich mit dem Thema „Metadatenkomplex aus nicht-technischer und technischer Sicht“ beschäftigte. Ziel war nicht nur der Aufbau einer Wissensplattform über Metadatenfragen im obigen Sinn, sondern auch die Erstellung von Richtlinien zur Beantwortung konkreter Fragen mit Bezug auf Metadaten. Nach einer Erhebung des Ist-Zustandes an den Partnerinstitutionen lag der

Schwerpunkt vor allem bei den in Frage kommenden Metadaten-Standards (Dublin Core, MODS, METS, ...) und deren Vor- und Nachteilen hinsichtlich bestimmter Datentypen. Weitere Themen betrafen Linked (Open) Data und deren Umsetzungsmöglichkeiten, eine Analyse von Best-Practice-Modellen im Fachbereich, den Einsatz von kontrollierten Vokabularen und nicht zuletzt eine Bedarfsanalyse bezüglich der Ansprüche der ForscherInnen an Metadaten zur Beschreibung der von ihnen erzeugten bzw. verwendeten Daten.<sup>2</sup>



Abb. 1: Workshop „Metadata Management – the Way to Open Science“ (Foto: Susanne Blumesberger)

Krönender Schlusspunkt der Arbeit des Metadatenclusters<sup>3</sup> war ein am 22. Juni 2016 am Campus der Universität Wien abgehaltener Workshop mit internationaler Beteiligung unter dem Titel „Metadata Management – the Way to Open Science“ (Abb. 1).<sup>4</sup> 13 Vortragende beleuchteten das Thema unter den verschiedensten Gesichtspunkten mit Beiträgen, die – soweit sie uns zur Verfügung standen – den überwiegenden Teil der Rubrik "Schwerpunktthema" ausmachen:

- Anna Fensel (Universität Innsbruck) zeigt in „Towards Semantic APIs for Research Data Services“ semantische Technologien und den Einsatz von Programmierschnittstellen zur Verknüpfung von Linked und Open Data.

- Walter Koch (Technische Universität Graz) und Gerda Koch (AIT Forschungs-GesmbH) erklären in „*Aggregation and Management of Metadata in the Context of Europeana*“ das offene Datenmodell der European Digital Library *Europeana* und zeigen, wie dieses Modell die gemeinschaftliche Indizierung von Metadaten aus verschiedenen kulturellen und wissenschaftlichen Bereichen unterstützt.
- Georg Neubauer (Donau-Universität Krems) berichtet in „*Visualization of Typed Links in Linked Data*“ über eine Benutzerstudie, die die Auswirkung unterschiedlicher grafischer Darstellungsweisen von Linked Data auf die Benutzerfreundlichkeit untersucht.
- Barbara Petritsch (IST Austria) stellt in „*Metadata for Research Data in Practice*“ das Repositorium *DataRep* vor, das sich direkt an den Bedürfnissen des wissenschaftlichen Personals orientiert und mit dem das IST Austria dem wachsenden Bedarf an digitalen Publikationen entgegenkommen will.
- Sarah Teetor und Nina Rannharter (Universität Wien) erläutern in „*Challenges of Building and Maintaining an Image Database: a Use Case Based on the Digital Research Archive for Byzantium (DiFAB)*“ die Metadatenstruktur des Digitalen Forschungsarchivs Byzanz am Beispiel eines Kulturdenkmals und seiner fotografischen Dokumentation.
- Sándor Kopácsi<sup>†</sup>, Rastislav Hudak und Raman Ganguly (Universität Wien) beschreiben in „*Implementation of a Classification Server to Support Metadata Organization for Long Term Preservation Systems*“ einen Klassifikationsserver für PHAIDRA auf der Basis von SKOSMOS und dem Jena Fuseki SPARQL-Server.
- José Luis Preza (ehem. Universität Wien) zeigt in „*Data Science and Analytics in Libraries*“, wie moderne Analysetechniken dabei helfen, die umfangreichen Datenmengen in Bibliotheken besser zu erschließen und das Durchsuchen erleichtern.

Den Abschluss des Schwerpunktthemas bildet der Beitrag „*Umgang mit Metadaten in Repositorien – eine österreichweite Umfrage*“ von Susanne Blumesberger und Alexander Zartl (Universität Wien). Dabei handelt es sich um eine Momentaufnahme von 2015, die den Repositoriumsverantwortlichen an den Partnerinstitutionen zur Dokumentation des Ausgangszustandes von *e-Infrastructures Austria* präsentiert wurde. Für die mittlere Zukunft (Zeiträumen: ca. fünf Jahre) ist im Rahmen des Nachfolgeprojekts *e-Infrastructures Austria plus*, das von der Universität Innsbruck koordiniert wird, eine Wiederholung dieser Befragung geplant, die es ermöglichen soll, den Fortschritt und damit die Auswirkungen des Projekts *e-Infrastructures Austria*

aufzuzeigen. Im Rahmen dieses Projekts werden auch die anderen Themen im Bereich Metadaten weiterverfolgt.

Mag.<sup>a</sup> Dr.<sup>in</sup> Susanne Blumesberger, MSc  
ORCID: <http://orcid.org/0000-0001-9018-623X>  
Universität Wien, Bibliotheks- und Archivwesen  
E-Mail: [susanne.blumesberger@univie.ac.at](mailto:susanne.blumesberger@univie.ac.at)

Mag. Bernhard Schubert, BA MSc  
Universität Wien, Bibliotheks- und Archivwesen  
E-Mail: [bernhard.schubert@univie.ac.at](mailto:bernhard.schubert@univie.ac.at)

Dr. Alexander Zartl  
Universität Wien, Bibliotheks- und Archivwesen  
E-Mail: [alexander.zartl@univie.ac.at](mailto:alexander.zartl@univie.ac.at)

- 1 Der Projektendbericht ist unter <http://phaidra.univie.ac.at/o:460779> abrufbar.
- 2 Siehe auch: Blumesberger, Susanne: Die Welt der Metadaten im Universum von Repositorien. In: Mitteilungen der VÖB 68 (2015) 3/4, S. 515–528. Online unter: <https://ojs.univie.ac.at/index.php/voebm/article/view/1295>
- 3 Die Ergebnisse des Clusters sind unter <http://phaidra.univie.ac.at/o:441674> abrufbar.
- 4 Informationen über den Workshop findet man unter: <http://phaidra.univie.ac.at/o:441281>.



Dieses Werk ist lizenziert unter einer  
Creative-Commons-Lizenz Namensnennung 4.0 International



## ■ TOWARDS SEMANTIC APIS FOR RESEARCH DATA SERVICES

by Anna Fensel

**Abstract:** Rapid development of Internet and Web technology is changing the state of the art in communication of knowledge, or results of research activities. Semantic technology as well as linked and open data in particular are becoming key enablers for successful and efficient progress in research. At first, I define the research data service (RDS) and discuss typical current and possible future usage scenarios involving RDSs. Furthermore, I discuss the state of the art in the areas of semantic service and data annotation and API construction, as well as infrastructural solutions applicable for RDS realization. Finally, innovative methods of online dissemination, promotion and efficient communication of research are discussed.

**Keywords:** Research Data Service (RDS); Research Data; RDS Metadata; Web API; Semantic Web Service; Semantic Technology; RDS Publication; Research Dissemination

## AUF DEM WEG ZU SEMANTISCHEN APIS FÜR FORSCHUNGSDATENDIENSTE

**Zusammenfassung:** Die schnelle Entwicklung der Internet- und Web-Technologie verändert den Stand der Technik in der Kommunikation von Wissen oder Forschungsergebnissen. Insbesondere semantische Technologien sowie verknüpfte und offene Daten werden zu entscheidenden Faktoren für einen erfolgreichen und effizienten Forschungsfortschritt. Zuerst definiere ich den Research Data Service (RDS) und diskutiere typische aktuelle und mögliche zukünftige Nutzungsszenarien dafür. Darüber hinaus bespreche ich den Stand der Technik in den Bereichen semantische Dienstleistung, Datenannotation und API-Konstruktion sowie infrastrukturelle Lösungen, die für die RDS-Realisierung anwendbar sind. Zum Schluss werden noch innovative Methoden der Online-Verbreitung, Förderung und effizienten Kommunikation der Forschung diskutiert.

**Schlüsselwörter:** Forschungsdatendienste (RDS); Forschungsdaten; RDS Metadaten; Web API; Semantische Technologie; RDS Publikation; Verbreitung von Forschung



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## **Contents**

1. *Introduction to and Motivation for Research Data Services (RDSs)*
2. *Modeling RDSs*
3. *Managing and programming RDSs*
4. *Disseminating RDSs*
5. *Conclusion*

## **1. Introduction to and Motivation for Research Data Services (RDSs)**

Acceptance of the open science principles<sup>1</sup> entails open access not only to research data, but also to tools that allow researchers to perform various types of activities over these data including mining, visualization, and analysis. The data and tools can be called Research Data Services (RDSs), enabling researchers to conduct their research activities efficiently and effectively.

One of the challenges faced by researchers in a globally networked scientific world is to be able to locate RDSs that fulfil their research needs. RDSs should be discoverable, i.e. have a feature at the semantic service description level that enables automatically locating research data services that fulfil a researcher goal. Making a RDS discoverable enables service (re-)use. Research data and service infrastructures are becoming increasingly interlinked, and semantic modeling and linked data are playing an important instrumental role in this process (Thanos, 2016).

Essentially, RDSs should have the following characteristics:

- they are subclasses of Services in a general sense (have a service provider and a service consumer, added value, ...),
- they are data services, part of a data economy,
- they are applicable in scenarios implementing some part of the research process,
- they may be delivered by a program/IT system, but also via other means, e.g. a human.

Wikipedia defines "Research" as "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of humans, culture and society, and the use of this stock of knowledge to devise new applications." Therefore, RDSs modeled as semantic Application Programming Interfaces (APIs) can increase efficiency of research in a broad sense, including the following tasks in particular:

- discovering new knowledge facilitating the process of research,
- delegation and contracting of research tasks to humans (e.g. in crowdsourcing, interdisciplinary research, etc.) as well as machines,
- combination of data and information coming from heterogeneous sources,
- seamless benchmarking and integration of stand-alone research efforts,
- etc.

According to the RDA Europe project (Thanos and Candela, 2016), a "research data service is a rule of correspondence between two sets", or "a Concrete Research Data Tool on which there exists an Institutional Commitment in the form of a Service-Level Agreement".

Technically, RDSs can be based on Web services, as the latter implement a service-oriented architecture in a specific manner, and are essentially a programmatic layer on top of distributed systems. Therefore, RDSs, as defined here earlier:

- May or may not be implemented as Web services.
- In any case have specific characteristics related to research.

This paper is organized as follows: Section 2 discusses the modeling aspects of RDSs, the details of RDSs' technical management and programming in Section 3, and innovative dissemination techniques for RDSs in Section 4. Section 5 concludes the paper.

## 2. Modeling RDSs

As a typical Web service, an RDS API would have the following types of properties:

- *Functional* (contain the formal specification of *what* exactly the service can do.),
- *Behavioral* (describe the functionality of the service that can be achieved in terms of interaction with the service and in terms of functionality required from other Web services.), and
- *Non-functional properties* (capture constraints over the previously mentioned properties.)

Similarly, the tasks one can perform with RDSs would, as in the world of Web services, be as follows (Fensel et al., 2011; Cardoso et al., 2014):

- **Discovery:** "Find services that match the service requester specifications".
- **Selection and Ranking:** "Choose the most appropriate services among the available ones".
- **Composition:** "Assembly of services based in order to achieve a given goal and provide a higher order of functionality".
- **Mediation:** "Solve mismatches among domain knowledge used to describe the services, protocols used in the communication, data exchanged in the interaction (types used, and meaning of the information) and business models of the different parties".
- **Execution:** "Invocation of a concrete set of services, arranged in a particular way following programmatic conventions that realize a given task".
- **Monitoring:** "Supervision of the correct execution of services and dealing with exceptions thrown by composed services or the composition workflow itself".
- **Handover:** "Replacement of services by equivalent ones, which solely or in combination can realize the same functionality as the replaced one, in case of failure while execution".

As to any linked data services, the semantic web and linked data principles, e.g. on linked data publishing in particular (Heath and Bizer, 2011), would be applicable to RDSs. Scientists and librarians, who comprise the relevant communities here, are de facto more on the early adopter side of semantic web and linked data principles. In the bibliographic domain, semantic mark-up complying with specialized vocabularies such as Dublin Core has already been in use for many years, even decades (Klee, 2013). Many research efforts around using linked data for open science, such as in distribution of educational and research content – especially the adoption of the practices –, are fostered by educational and research institutions, universities in particular (Mouromtsev & d'Aquin, 2016). Semantic formalism has also been applied for decades in some fields, e.g. the life sciences<sup>2</sup>. Another relevant development here is schema.org. Schema.org provides a collection of vocabularies for sharing information relevant in the context of the Web. It was launched in June 2011 by Bing, Google and Yahoo!, further joined by Yandex in November of the same year. Its purpose is to create a common set of schemas for webmasters to mark-up their websites with structured data. It has proven to be a very large success: eventually everything that can be consumed or booked through the Web is semantically annotated with schema.org – there are over 4.8 million schema.org

annotations describing hotels available on the Web, for example (Kärle et al., 2016). However, as with any linked data, data quality is and remains a large issue that needs to be addressed in the near future (Zaveri et al., 2016; Kärle et al., 2016).

The advantages of schema.org are as follows:

- Webmasters can use schema.org to mark up their web pages (creating enriched snippets) in a way that is recognized by major search engines.
- The enriched snippets enable search engines to understand the information on web pages that results in richer and more attractive search results for the users. Hence it is easier for users to find relevant and right information on the web.
- Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results.
- It helps webmasters achieve higher rankings of their pages in search results.
- This markup has the potential to enhance the CTR (click through ratio) from the search results from anywhere between 10–25%<sup>3</sup>.
- Schema.org can be also used for structured data interoperability.
- Its usage can also lead to the development of new tools, for example Google Recipe Search, which may open up other marketing channels if not now, then in the near future.
- Query/Answer based search engines will be improved, particularly in terms of semantic search, by making use of structured data, i.e. the search engine can understand the content of a website and make use of it to give a direct and accurate search result.

In conclusion, schema.org is also obviously relevant for RDSs, as research is being done on a variety of objects, most of these existing in real life and already annotated with schema.org.

Metadata needed for the description of the RDS should contain an "input set" (domain) and the "output set" (co-domain). Whereas the service itself is typically a process, which is not really possible to describe formally. Examples of details that can be annotated as input and output of the service are shown in Figure 1.

Ideally, we need to specify the syntax and semantics of the elements of the domain and co-domain. Establishment and spread of such specifications is also a realistic development path as the amount of research data services increase and the research steering service economy is becoming more interdisciplinary and it becomes more difficult to identify and find relevant services.

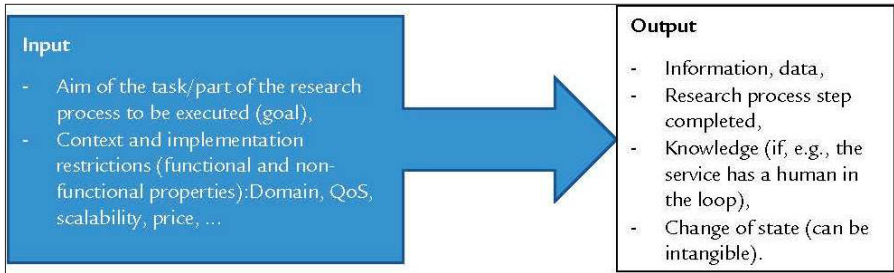


Fig. 1: Inputs and outputs of RDSs

In the real world, however, simpler things and models spread better. For example, despite many developments in semantic web services, the largest service registries, like ProgrammableWeb<sup>4</sup>, still do not have semantic descriptions for service APIs. Adding these clearly constitutes progressive usage potential, and is also a likely to happen here first since the research community tends to be an early adopter of new technology.

An RDS instance definitely has the dynamicity aspect as well, and changes in time in the following ways:

- Its quality may alter, its non-functional properties change,
- Its context and usage may vary,
- It may appear and disappear,
- Its implementation may change, etc.

Modeling such state changes programmatically is difficult, and making them widely used is even more difficult. Therefore, RDSs should be designed as stateless.

In order to appropriately describe its functionality, the RDS profile description should include:

- Aimed dataset or service,
- Scientific discipline,
- Scientific method,
- Domain,
- Information about quantity, quality, availability, creator(s)/ provider(s),
- Access and license policies,
- Origin and annotation of reused/subcontracted sources (if applicable).

Some or even all of the above attributes can be optional. Examples for a service can include:

- "compare performance of my semantic repository according to criteria X",
- "find datasets with energy consumption of fridges in Vienna".

Scientific workflows are appropriate for describing the process model of a RDS, for example in conjunction with research method description. Existing models, such as BPEL<sup>5</sup>, USDL<sup>6</sup> and Linked USDL (Cardoso et al., 2014) can be used to interlink business models, service systems, service models, service instances and service descriptions (Cardoso et al., 2014).

### 3. Managing and programming RDSs

There are existing service description frameworks like OWL-S<sup>7</sup> and WSMO (Fensel et al., 2011), but in most practical cases they are too complex, except where the related technologies are already being built on RDF infrastructures. In genetics, for example, some data annotations are made using OWL, and building a service on top of it would be a natural extension. But in many cases it would not be the best choice, as most data are not shared on the Web with OWL, and its semantic and syntactic complexity and expressivity are higher than necessary for developers.

Using real data in research is essential, so the frameworks containing the real data are the most important and should therefore comprise and rely on linked (open) data and schema.org annotations. For the modeling level, a feasible approach could be applying the Linked Service System (LSS) model structure, defining a human-friendly way to model a service by representing key aspects of the service by answering the essential questions about it (Who/Role, Why/Goal, What/Resource, How/Process, When/Time, Where/Location), and linking to data represented in existing formalisms like USDL and Linked USDL (Cardoso et al., 2014).

It is also worth noting that schema.org has Actions as a part of its model, which makes modeling typical actions relevant in the Web context possible. Using it of course also ensures direct compatibility with the multitude of real data available in schema.org.

Given that many service representation languages are still evolving at present, it is important to note the main characteristics and capabilities of a knowledge representation language appropriate for the description of the functionality of a data service as well as for effectively supporting reasoning in the matchmaking process. These include the following parts:

- Presence of the ways to annotate the functionality, domain, ...
- Assumption of "incorrectness and incompleteness", or the need to combine heterogeneous techniques in reasoning (Fensel & Van Harmelen, 2007) – given that working with RDS would have an open world assumption as well as any reasoning on the Web, the principles of reasoning as performed before in "closed" infrastructures (databases) would become irrelevant.
- Much of matchmaking and reasoning should be moved to the applications – but: semantics can support maintenance of community-generated reusable mapping (e.g. stating that two service parts are the same).

Discipline-specific classification of data services (classes of data services) supported by discipline-specific ontologies would be considered state of the art. We particularly need it, because:

- Data sets vary from domain to domain, and often the research is domain-specific.
- It facilitates discoverability.

This would also be in agreement with other trends in the services area, e.g. microservices they have a very narrow focus – typically very domain-specific ones, but with semantic APIs that enable easy integration into more complex services that, in turn, can be modified, re-created and re-published by the developers as well as the end users (Davies et al., 2011).

The role of registries, directories, and catalogs of services – or essentially, infrastructures, is very important as they provide a) a single point to make services discoverable, b) meta-Research Data Services in themselves, as a collection of services. Ideologically they may be constructed architecturally similarly to UDDI from the past, or like ProgrammableWeb for Web APIs now.

Once semantically annotated, RDSs can be found by various features e.g. "stateless/state-based", "type of input data: discrete data/vectors/functions/streaming data", "types of output data", etc. – by data, by domain, by functionality, ... The classifications do not have to be created a priori, but could be created ad hoc once the annotations are there in order to avoid restricting usage. De facto, RDSs are classified by provider platforms, e.g. in the areas the platforms are operating in mainly: like Linked Open Vocabularies (LOV) or datahub.io for structured data, or in the library domain, where numerous repositories for publications are provided by various publishers, such as Zenodo or Google Scholar.



Last but not least, being able to identify and reference RDSs is essential, and “citation” is instrumental in making research data services discoverable. Efforts in bridging such data to semantic formats are already ongoing, e.g. opening up bibliographic data (Peroni et al., 2015). Of course, eventually identification and citation data will become a necessary accompanying part of research data and RDSs as well.

Citation information will be used in processes, e.g. in ranking. As in other cases where citation numbers influence ranking (e.g. in Google Scholar- most cited publications are displayed on top), this is not optimal, as searches according to these criteria may overlook data with closer matches, and the output of such ranking processes itself also impacts citations as the items that are ranked higher are more visible and cited more. To ameliorate this situation, several ways for choosing semantic annotations have been suggested. These include taking into consideration the provenance specifics (e.g. author reputation) of newly appearing annotations (Stavrakantonakis et al., 2016) as well as inclusion of end-user perspectives in the ranking process (Klan & König-Ries, 2014).

#### **4. Disseminating RDSs**

Nowadays, research is disseminated not only via classical channels like digital libraries, but also in a multi-channel manner, e.g. via social media, collaborative infrastructures, and a multitude of other diverse communication and dissemination channels (Fensel et al., 2014). Arguably, social media are used by younger generations more commonly than email, and social media networks like Facebook, LinkedIn, Xing, Twitter and Google+ have become a mainstream mode of communication in general. Channels focusing on research communication exist as well, such as ResearchGate, Academia.edu, Google Scholar and SlideShare, to name a few. They are increasingly used by researchers, and research can be directly followed by interested parties, e.g. by following accounts of relevant groups and relevant researchers.

There are even solutions to make the multi-channel communication of research more efficient and less time-consuming. For example, ONLIM<sup>8</sup> is an online tool based on semantic technologies that aims to facilitate managing various social media platforms by means of publishing posts and tracking the feedback given by other users. ONLIM supports several social media platforms such as Facebook, Twitter, Youtube, LinkedIn, Xing and Flickr. It also allows users to schedule their posts to enable more effec-

tive social media management and marketing. In contrast to other similar tools, ONLIM also has an automatic post generation feature that creates posts for publication in social media from external sources. RDSs that are annotated with schema.org can be converted to social media posts by ONLIM automatically, for example. In addition, these posts can be automatically forwarded to all appropriate channels to disseminate the research. While they are already being used in mainstream marketing (Fensel et al., 2014; Fensel et al., 2016), such developments are becoming a game changer in RDS discoverability and in how research is being disseminated and accessed.

With content and data, there are a few particularities hampering their potential (re-)use in the research data value chain. Data licensing in particular is still complicated, and semantic formats for licensing data, as well as the tools processing data license annotations are under-defined or nonexistent. Semantic standards for licenses are being developed right now, and include efforts such as ODRL<sup>9</sup> and RightsML (Ermilov and Pellegrini, 2015). Another example of an ongoing Austrian project focusing on semantic languages and tools for data licensing is DALICC<sup>10</sup>.

## 5. Conclusion

RDSs are data services specifically addressing the production processes needed for research and education, and as such belong to the data service value chain. In the long run, they would be accessible via and manageable with semantically annotated APIs, over a network of RDS repositories and e-Infrastructures.

Promising semantic languages and technologies exist that can be applied to solutions of RDS modeling and discovery problems, e.g. linked services, linked data and schema.org. These solutions would facilitate the use of semantic data already available in terms of research, discoverability and applicability.

Efficient dissemination of research is very important. Dissemination also needs to be multi-channel now, and new kinds of channels appear, e.g. social media and collaborative environments. Eventually, the Web of RDSs would be accessible over a semantic API, so that the most relevant RDSs can be delivered or activated by the user merely as a result of a query. Such functionality, as well as RDS delivery and activation across different communication and dissemination channels, would facilitate conduction of interdisciplinary research.

Relevant data value chain languages, techniques and tools are in development, e.g. on (semantic) data licensing. These solutions would facilitate implementation of the data value chain in research. In particular, data contributions from professional researchers as well as unprofessional researchers (e.g. data generated via crowdsourcing) would be clearly annotated in terms of rights, permissions and obligations associated with its usage.

Ass. Prof.<sup>in</sup> Dr.<sup>in</sup> Anna Fensel

ORCID: <http://orcid.org/0000-0002-1391-7104>

University of Innsbruck, Semantic Technology Institute (STI)

E-Mail: [anna.fensel@sti2.at](mailto:anna.fensel@sti2.at)

## References

- Cardoso, J., Lopes, R., & Poels, G. (2014). Service systems: concepts, modeling, and programming (pp. 1–91). Springer. DOI: <https://dx.doi.org/10.1007/978-3-319-10813-1>.
- Davies, M., Carrez, F., Heinilä, J., Fensel, A., Narganes, M., & Carlos dos Santos Danado, J. (2011). m: Ciudad: enabling end-user mobile service creation. *International Journal of Pervasive Computing and Communications*, 7(4), pp. 384–414. DOI: <https://doi.org/10.1108/17427371111189683>.
- Ermilov, I., & Pellegrini, T. (2015, September). Data licensing on the cloud: empirical insights and implications for linked data. In *Proceedings of the 11th International Conference on Semantic Systems* (pp. 153–156). ACM. DOI: <https://doi.org/10.1145/2814864.2814878>.
- Fensel, A., Akbar, Z., Toma, I., & Fensel, D. (2016). Bringing Online Visibility to Hotels with Schema.org and Multi-channel Communication. In *Information and Communication Technologies in Tourism 2016* (pp. 3–16). Springer International Publishing. DOI: [https://doi.org/10.1007/978-3-319-28231-2\\_1](https://doi.org/10.1007/978-3-319-28231-2_1).
- Fensel, A., Toma, I., García, J. M., Stavrakantonakis, I., & Fensel, D. (2014). Enabling customers engagement and collaboration for small and medium-sized enterprises in ubiquitous multi-channel ecosystems. *Computers in Industry*, 65(5), pp. 891–904. DOI: <https://doi.org/10.1016/j.compind.2014.02.001>.
- Fensel, D., Facca, F. M., Simperl, E., & Toma, I. (2011). *Semantic web services*. Springer Science & Business Media. DOI: <https://doi.org/10.1007/978-3-642-19193-0>.

- Fensel, D., & van Harmelen, F. (2007). Unifying Reasoning and Search to Web Scale. *IEEE Internet Computing*, 2(11), pp. 96–95. DOI: <https://doi.org/10.1109/MIC.2007.51>.
- Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), pp. 1–136. DOI: <https://doi.org/10.2200/S00334ED-1V01Y201102WBE001>.
- Kärle, E., Fensel, A., Toma, I., & Fensel, D. (2016). Why Are There More Hotels in Tyrol than in Austria? Analyzing Schema.org Usage in the Hotel Domain. In *Information and Communication Technologies in Tourism 2016* (pp. 99–112). Springer International Publishing. DOI: [https://doi.org/10.1007/978-3-319-28231-2\\_8](https://doi.org/10.1007/978-3-319-28231-2_8).
- Klan, F., & König-Ries, B. (2014, June). A user-centered methodology for the evaluation of (semantic) web service discovery and selection. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*. ACM. DOI: <https://doi.org/10.1145/2611040.2611069>.
- Klee, C. (2013). Vokabulare für bibliographische Daten. (Open) Linked Data in Bibliotheken, S. 45–63. DOI: <https://doi.org/10.1515/9783110278736.45>.
- Mouromtsev, D., & d'Aquin, M. (Eds.). (2016). *Open Data for Education: Linked, Shared, and Reusable Data for Teaching and Learning* (Vol. 9500). Springer. DOI: <https://doi.org/10.1007/978-3-319-30493-9>.
- Peroni, S., Dutton, A., Gray, T., Shotton, D. (2015). Setting our bibliographic references free: towards open citation data. *Journal of Documentation*, 71 (2): pp. 253–277. DOI: <https://doi.org/10.1108/JD-12-2013-0166>.
- Stavrakantonakis, I., Fensel, A., & Fensel, D. (2016, September). Linked Open Vocabulary ranking and terms discovery. In *Proceedings of the 12th International Conference on Semantic Systems* (pp. 1–8). ACM. DOI: <https://doi.org/10.1145/2993318.2993338>.
- Thanos, C. (2016). A Vision for Open Cyber-Scholarly Infrastructures. *Publications*, 4(2), 13. DOI: <https://doi.org/10.3390/publications4020013>.
- Thanos, C., Klan, F., Kritikos, K., & Candela, L. (2016). White Paper on Research Data Service Discoverability. *Publications*, 5(1), 1. DOI: <https://doi.org/10.3390/publications5010001>.
- Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, 7(1), pp. 63–93. DOI: <https://doi.org/10.3233/SW-150175>.
- Zhdanova, A. V. (2008, May). Community-driven ontology evolution: Gene ontology case study. In *International Conference on Business*

Information Systems (pp. 118-129). Springer Berlin Heidelberg. DOI: [https://doi.org/10.1007/978-3-540-79396-0\\_11](https://doi.org/10.1007/978-3-540-79396-0_11).

- 1 Open Science – Digital Single Market: <https://ec.europa.eu/digital-single-market/en/open-science>.
- 2 See, for example, an analysis of the community-driven development of Gene Ontology (Zhdanova, A. V, 2008).
- 3 See, for example, an analysis of a hotel web presence annotated with schema.org (Fensel et al., 2016).
- 4 ProgrammableWeb – APIs, Mashups and the Web as Platform: <http://www.programmableweb.com>.
- 5 BPEL: <https://www.oasis-open.org/committees/wsbpel/>.
- 6 USDL: <https://www.w3.org/2005/Incubator/usdl/XGR-usdl-20111027/>.
- 7 OWL-S: <https://www.w3.org/Submission/OWL-S/>.
- 8 ONLIM – Social Media Management Tool: [www.onlim.com](http://www.onlim.com).
- 9 ODRL – Open Digital Rights Language: <https://www.w3.org/ns/odrl/2/>.
- 10 DALICC – Data Licenses Clearance Center: <https://dalicc.net>.

## ■ AGGREGATION AND MANAGEMENT OF METADATA IN THE CONTEXT OF EUROPEANA

by *Walter Koch* & *Gerda Koch*

**Abstract:** *The creation of connected content and the linking of metadata are basic requirements for the realisation of the semantic web. Semantic linkage of data enables the joint search of heterogeneous databases and facilitates future machine learning. The present article outlines the metadata management and metadata linking activities of the European Digital Library. A short overview on the current core research areas and implementation strategies in this field is presented. Various projects and metadata services tailored to natural history data, regional cultural heritage data and audio collections are described.*

**Keywords:** *Linked data; data enrichment; microservices; data aggregation; Europeana; cultural heritage metadata; natural history data; audio collections; data annotation*

### AGGREGATION UND MANAGEMENT VON METADATEN IM KONTEXT VON EUROPEANA

**Zusammenfassung:** *Mit dem In-Beziehung-Setzen und Verlinken von Daten im Internet wird der Weg zur Umsetzung des semantischen Webs geebnet. Erst die semantische Verbindung von heterogenen Datenbeständen ermöglicht übergreifende Suchvorgänge und späteres „Machine Learning“. Im Artikel werden die Aktivitäten der Europäischen Digitalen Bibliothek im Bereich des Metadatenmanagements und der semantischen Verlinkung von Daten skizziert. Dabei wird ein kurzer Überblick zu aktuellen Forschungsschwerpunkten und Umsetzungsstrategien gegeben und einzelne Projekte und maßgeschneiderte Serviceangebote für naturhistorische Daten, regionale Kultureinrichtungen und Audiosammlungen werden beschrieben.*

**Schlüsselwörter:** *Verlinkung von Daten; Anreicherung von Daten; Microservices; Datenaggregation; Europeana; Kulturerbedaten; Naturhistorische Daten; Audiosammlungen; Annotationen*



Dieses Werk ist lizenziert unter einer [Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## **Contents**

1. *Linking of heritage data in Europe*
2. *Natural history data in Europeana*
3. *Serving the data development and aggregation of regional and cultural heritage institutions*
4. *Annotating content*

## **1. Linking of heritage data in Europe**

The European Digital Library, Europeana<sup>1</sup>, was first launched in 2008 and is the cross-domain central portal and single entry point to Europe's digitized cultural heritage. The Europeana Data Model (EDM)<sup>2</sup> is the metadata model that is used for aggregating and ingesting data from the diverse European cultural and natural heritage repositories into the joint digital library. EDM describes the data using the Resource Description Framework (RDF)<sup>3</sup> and re-uses the Simple Knowledge Organization System (SKOS)<sup>4</sup> framework for the integration of vocabulary concept terms. The Resource Description Framework (RDF) provides the syntax and rules that build the basis of linked data and structures information as triples of subjects, predicates and objects. In addition, ontologies and vocabularies reduce the complexity of the world and help to put vast amounts of data into structured forms.

Popular use cases often list the following benefits of publishing Linked Open Data:

- Being findable on major search engines and social platforms in the first place
- Cross-language retrieval based on vocabularies
- Vocabulary based facets with multilingual facet labels
- Auto completion with semantic disambiguation and suggestions of related searches
- Cross-linked entity pages for concepts, agents, places and periods

Recognizing these advantages, Europeana has started to strongly encourage its partners to provide open metadata, so called Linked Open Data (LOD), with an increasing proportion of links to content marked Public Domain or bearing Creative Commons licenses.

A basic premise of the LOD approach is to reuse, wherever possible, terms from existing standard vocabularies, rather than reinvent them when publishing Resource Description Framework (RDF) data. This maximizes

the probability that the data can be used without additional modifications by applications that may be tuned to well-known vocabularies. Vocabulary standards for creating (ISO 25964)<sup>5</sup> and publishing vocabularies (SKOS) on the web offer solutions to interconnect isolated data silos and support cross search and data comparison. SKOS (Simple Knowledge Organization System) has become an accepted standard model for expressing the structure and content of concept schemes like vocabularies, authority lists and thesauri on the web using RDF.

Data enrichment with vocabulary terms can be done right when the data object is registered in the local cataloguing system, or "afterwards", when the metadata is processed further. Vocabulary web services provide the functionalities to enrich metadata (catalogue data) with standard vocabulary information that further allow the automatic linking of data from various online data sets. When data is ingested into joint virtual catalogues (like Europeana), these automated enrichment functionalities are often used in order to support semantic linking of heterogeneous data and easy browsing through the entire repository. The following paragraphs describe recent tools and services for the semantic enrichment of data that have been developed by European research projects.

## **2. Natural history data in Europeana**

OpenUp!<sup>6</sup> started in 2011 as a three year project of the European Commission with the main goal to provide online access to a wide range of natural history collection objects and to connect the cultural heritage and natural history domains. Today OpenUp! is a constantly growing network and Europeana's aggregator for the natural history domain. OpenUp! is using international vocabularies within a common names web service that draws common names from about 25 different vernacular names lists from various countries. The enrichment with common name information opens up the natural history research data to a considerable number of other user communities, like tourists, pupils or nature enthusiasts. The OpenUp! technical aggregator uses the open source Business Intelligence tool Pentaho Kettle<sup>7</sup> for data integration and ETL (Extract, Transform, Load) mechanisms. The chosen tool provides the ETL functionalities needed for the processing of the distributed natural history datasets across Europe. As a result, the entire data management process for the Europeana data provision was tailored to three steps: transform, validate, and OAI<sup>8</sup>-import.





- Vocabulary service (enriches data with various vocabulary terms)
- Historic Place Names service (adds historic place names information)
- Geo-coding application (tool to manually enrich data with geoinformation)
- Vocabulary matching service (automated vocabulary matching)
- Background linking service (automated enrichment with Wikipedia information)
- Wikimedia application (capture data from selected Wikimedia resources)

The LoCloud vocabulary microservice, for instance, provides the business capability for vocabulary management and enrichment of metadata. The LoCloud vocabularies have been imported into the TemaTres<sup>10</sup> tool in order to use them via web services in the aggregation process. That way the LoCloud microservice "Generic enrichment" automatically receives the vocabularies available in the vocabulary tool and uses them during the automated enrichment process conducted in the LoCloud aggregation facility MORE. Further on, instead of matching the vocabularies automatically to the metadata, it is also possible to select individual terms from the LoCloud vocabularies that should be added to the metadata.

One major benefit of microservices is that these small and independent services may be plugged in into other applications and can be used via the APIs they provide. The LoCloud vocabulary tool can also be used to create vocabularies from scratch or to import already existing vocabularies. By importing existing vocabularies to the tool the vocabularies become available in the SKOS format which has its own web presence that can be used for further semantic linking afterwards. Moreover, the LoCloud TemaTres vocabulary installation allows online collaboration when creating and extending vocabularies (for example in order to create new translations of existing vocabularies).

Summing up, all LoCloud cloud-based SaaS (Software as a Service) tools serve the cultural heritage community in many ways. The enrichment services help to improve data quality for the content providers and Europeana. The data capturing service allows re-use of already existing online cultural content and its proper ingestion into Europeana. The scalability and various options for deploying and using the microservices help tailor the services offered to the needs and capabilities of cultural heritage institutions.

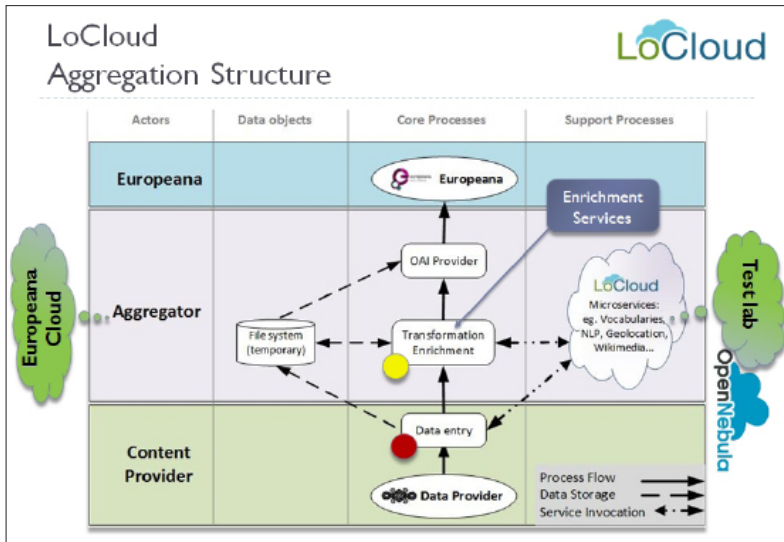


Fig. 2: The LoCloud Aggregation structure

#### 4. Annotating content

Several Europeana related projects have recently dealt with the annotation of content. Among these projects are the PATHS project, investigating automatic semantic enrichments, the DM2E project (scholarly annotations using the Pundit toolset), the SEALINCmedia project (expert annotation "nichesourcing", developed the Accurator tool), the Europeana Creative project (pre-alpha version of the Annotations API), Europeana 1989 (annotations in HistoryPin.org), Europeana V3 (roundtripping of annotations with HistoryPin.org), Europeana Food and Drink (annotations in HistoryPin.org) and Europeana Sounds with a wider range of user scenarios for annotations. Within these activities Europeana adopted the W3C Web Annotation Data Model<sup>11</sup> which is based on RDF and defines JSON-LD<sup>12</sup> as its default serialization format. This offers a model for exchanging annotations across platforms but needs further investigation into its flexibility to support complex scenarios.

The "Moments of Interest" Approach of the audio aggregation platform DISMARC<sup>13</sup> adopts the W3C recommendation on Media Fragments<sup>14</sup> and defines annotations as "has annotation" relationships between data objects instead of following the complex W3C Web Annotation Data Model (see figure 3).

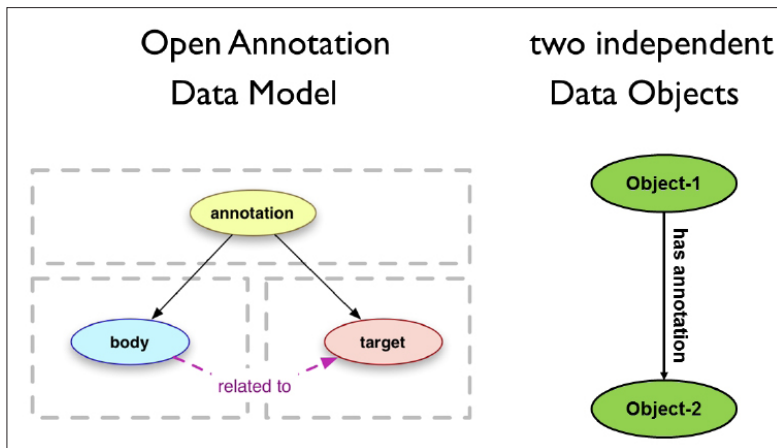


Fig. 3: Annotation Data Models

Moments of Interest can be chosen from a timeline, and therefore annotations may refer to media fragments that can be selected from audio, video or text streams etc. (see figure 4).

▶ **W3C Recommendation:**  
<https://www.w3.org/TR/media-frags-reqs/>

▶ **Use cases and requirements for Media Fragments:**

|                |                        |  |  |
|----------------|------------------------|--|--|
| General Header | Track Specific Headers |  | Example VIDEO Track<br>(Constant Frame Rate)               |
|                | Track Specific Headers |  | OTHER Tracks Types<br>(Further tracks possible)            |
|                | Track Specific Headers |  | Example AUDIO Track<br>(Constant Sampling Rate)            |
|                | Track Specific Headers |  | Example TEXT Track<br>(Discontinuous, Overlapping in time) |
|                |                        |  | Example IMAGE Track<br>(Discontinuous in time)             |

Fig. 4: Media Fragments

Different types of fragments refer to different media types. Video and audio streams lend themselves to temporal fragments, images to spatial fragments, remixed content to track fragments and data objects to named

fragments. According to this approach all Moments of Interest become independent objects and can be recorded using whatever type of metadata scheme that best fits the content.

The DISMARC v2 platform offers novel applications for the development of connected content (see figure 5) and seeks to prove whether this approach could be a better choice for linking independent data objects and annotating parts of existing data objects with additional information.



Fig. 5: Selecting content and describing Moments of Interest (DISMARC)

In the future the semantic internet will be capable of developing ontologies and connections between content without the need for human interaction. Linked information will be found and presented even more rapidly. The foundation of this process is the availability of semantically rich and open data sets. This article presented a selection of European tools, services and projects that aim at blazing the trail.

Univ. Prof. Dr. Walter Koch  
Steinbeis Innovation Transfer Center IMCHI  
E-Mail: [walter.koch@stw.de](mailto:walter.koch@stw.de)

Mag.<sup>a</sup> Gerda Koch, MBA  
AIT Angewandte Informationstechnik ForschungsgesmbH Graz  
E-Mail: [kochg@ait.co.at](mailto:kochg@ait.co.at)

## Notes

- 1 The Europeana Data Portal. <http://www.europeana.eu>
- 2 Europeana Data Model Documentation. <http://pro.europeana.eu/page/edm-documentation>
- 3 Resource Description Framework. <https://www.w3.org/RDF/>
- 4 Simple Knowledge Organization System. <https://www.w3.org/TR/skos-reference/>
- 5 ISO 25964 – the international standard for thesauri and interoperability with other vocabularies. <http://www.niso.org/schemas/iso25964/>
- 6 OpenUp! <http://open-up.eu/en>
- 7 Pentaho Data Integration. <http://www.pentaho.com/product/data-integration>
- 8 Open Archives Initiative Protocol for Metadata Harvesting. <https://www.openarchives.org/pmh/>
- 9 LoCloud. <http://www.locloud.eu/>
- 10 TemaTres. <http://www.vocabularyserver.com/>
- 11 Web Annotation Data Model. <http://www.w3.org/TR/annotation-model/>
- 12 JavaScript Object Notation for Linked Data. <http://json-ld.org/>
- 13 DISMARC. <http://www.dismarc.org>
- 14 Use cases and requirements for Media Fragments. <https://www.w3.org/TR/media-frags-reqs/>

## ■ VISUALIZATION OF TYPED LINKS IN LINKED DATA

by Georg Neubauer

**Abstract:** *The main subject of the work is the visualization of typed links in Linked Data. The academic subjects relevant to the paper in general are the Semantic Web, the Web of Data and information visualization. The Semantic Web, invented by Tim Berners-Lee in 2001, was announced as an extension to the World Wide Web (Web 2.0). The actual area of investigation concerns the connectivity of information on the World Wide Web. To be able to explore such interconnections, visualizations are critical requirements as well as a major part of processing data in themselves. In the context of the Semantic Web, representation of information interrelations can be achieved using graphs. The aim of the article is to primarily describe the arrangement of Linked Data visualization concepts by establishing their principles in a theoretical approach. Putting design restrictions into context leads to practical guidelines. By describing the creation of two alternative visualizations of a commonly used web application representing Linked Data as network visualization, their compatibility was tested. The application-oriented part treats the design phase, its results, and future requirements of the project that can be derived from this test.*

**Keywords:** *Visualization; Visual representations; Linked Open Data; Analytics; Semantic Web*

## VISUALISIERUNG VON TYPISIERTEN LINKS IN LINKED DATA

**Zusammenfassung:** *Das Themengebiet der Arbeit behandelt Visualisierungen von typisierten Links in Linked Data. Die wissenschaftlichen Gebiete, die im Allgemeinen den Inhalt des Beitrags abgrenzen, sind das Semantic Web, das Web of Data und Informationsvisualisierung. Das Semantic Web, das von Tim Berners Lee 2001 erfunden wurde, stellt eine Erweiterung zum World Wide Web (Web 2.0) dar. Aktuelle Forschungen beziehen sich auf die Verknüpfbarkeit von Informationen im World Wide Web. Um es zu ermöglichen, solche Verbindungen wahrnehmen und verarbeiten zu können sind Visualisierungen die wichtigsten Anforderungen als Hauptteil der Datenverarbeitung. Im Zusammenhang mit dem Semantic Web werden Repräsentationen von zusammenhängenden Informationen anhand von Graphen gehandhabt. Der Grund des Entstehens dieser Arbeit ist in erster Linie die Beschreibung der Gestaltung von Linked Data-Visualisierungskonzepten, deren Prinzipien im Rahmen einer theoretischen Annäherung eingeführt werden. Anhand des Kontexts führt eine schrittweise Erweiterung der Informationen mit dem Ziel, praktische Richtlinien anzubieten, zur*

Vernetzung dieser ausgearbeiteten Gestaltungsrichtlinien. Indem die Entwürfe zweier alternativer Visualisierungen einer standardisierten Webapplikation beschrieben werden, die Linked Data als Netzwerk visualisiert, konnte ein Test durchgeführt werden, der deren Kompatibilität zum Inhalt hatte. Der praktische Teil behandelt daher die Designphase, die Resultate, und zukünftige Anforderungen des Projektes, die durch die Testung ausgearbeitet wurden.

**Schlüsselwörter:** Visualisierung; Visuelle Repräsentation; Linked Open Data; Analyse; Semantic Web



Dieses Werk ist lizenziert unter einer [Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

The aim of this work is to construct and evaluate different graphic representations of Linked Data that can be compared in terms of better handling and simultaneously as a main question have a more suitable design as a result.

This leads to three main research questions:

1. *How can typed links in Linked Data be visualized for a broader number of users?*
2. *Which graphic representation do typed links have in Linked Data?*
3. *Is there an interrelation between the visualization and the functionality of graphs when it comes to interpretation?*

## 1. Data and Information

There are two important parts of processes of data transformation concerning the exchange of information<sup>1</sup>:

- *Formalization in the beginning*
- *Reconstruction after data transfer*

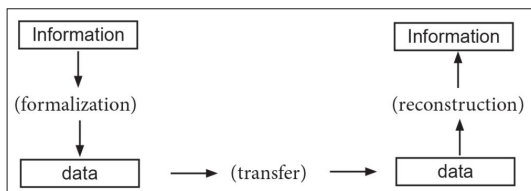


Fig. 1: Exchange of information<sup>2</sup>

Data preservation necessitates data storage. Nowadays data can be stored digitally using computers, magnetic memory media and the Internet.



## 2. Ontology

Data formalization is generally referred to as an ontology. It derives from the Greek *onto* (being) and *logia* (written or spoken discourse). In information technology, ontology is the *working model of entities and interactions* in a particular domain of knowledge or practices, such as electronic commerce. A good definition is the following sentence. "Ontology is a formal specification of a shared conceptualization"<sup>3</sup>.

The construction of ontologies is a cyclic process and consists of the following steps<sup>4</sup>.

- *Definition of area and domain*
- *Analysis of available ontologies for possible reutilization*
- *Identification of relevant terms*
- *Production of the draft hierarchy*
- *Definition of relations and formalization of classes*

## 3. The Web

The web builds on a gigantic amount of information and is accessible via the Internet. It is made up of three components:

- *HTML (Hyper Text Markup Language)*
- *HTTP (Hyper Text Transfer Protocol)*
- *URL (Uniform Resource Locator)*

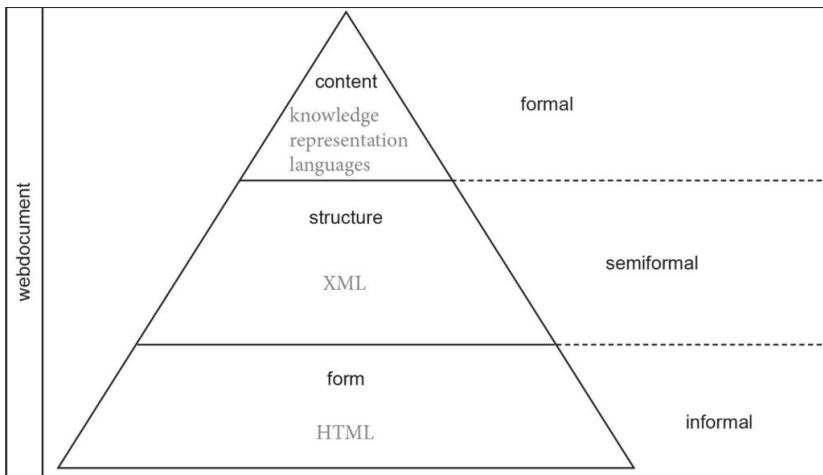


Fig. 2: Levels of structuralization of information on the web<sup>2</sup>

The Semantic Web is a suggestion for unorganized growing on the web. The idea of the Semantic Web is that information depending on content and relations can be processed. In order to bind information it is necessary to put the information in a confirmed, structured form.

#### 4. RDF

The Resource Description Framework (RDF) was developed for representing information on the web. RDF is a basic language for the description of structured information on the web and of course for the exchange of information<sup>5 6</sup>. URI = scheme ":" authority [path][ "?" query][ "#" fragment]

RDF is used to formulate logical statements. *Every resource is assigned a URI*. RDF is described by a number of triples (RDF graphs) which is often indicated by arrows<sup>6</sup>.

There are three components of these arrows:

- *The subject (is a URI or a "Black Node")*
- *The predicate (is a URI)*
- *The object (can be a URI, a "Black Node" or a "Literal")*

A Literal is used to identify a simple data type (numbers, data and alphanumeric characters).

A superset of RDF is RDF N3 (Notation 3). It extends the RDF data model by adding formulae, variables, logical implications, functional predicates and provides a textual syntax alternative to RDF/XML.

#### 5. SPARQL

SPARQL and the RDF Query Language are semantic query languages for databases and constitute key technologies for the Semantic Web. They can retrieve and manipulate data stored in the RDF format<sup>7</sup>.

SPARQL allows for a query to consist of *triple patterns, conjunctions, disjunctions, and optional patterns*<sup>8</sup>.

Implementations for multiple programming languages exist. There are tools that allow connection and semi-automatic construction of a SPARQL query for a SPARQL endpoint (i.e.: ViziQuer). Additionally, there are tools that are able to translate SPARQL queries to other query languages (e.g.: SQL).

#### 6. OWL

The W3C Web Ontology Language (OWL) is a computational logic-based

Semantic Web language and specification of the World Wide Web consortium that is intended to *represent complex knowledge about things, groups of things, and relations* between things. OWL documents (ontologies) can be published in the World Wide Web and may refer to or be referred to from other OWL ontologies. OWL is part of the W3C's Semantic Web technology area, which includes RDF, RDFS, SPARQL, etc.

The current version of OWL is OWL 2 (published in 2012) and is defined by five core specification documents describing its conceptual structure: the primary exchange syntax (RDF/XML), two alternative semantics (direct and RDF-based), and conformance requirements. In addition, three specification documents contain optional features that may be supported by some implementations: the language profiles and two other syntaxes (OWL/XML, Manchester).

## 7. Typed Links

The term "typed link" can be explained as a link relation and has a descriptive feature connected to a hyperlink that defines the type of the link or the relationship between the source and target resources<sup>9</sup>.

RDF typed links are essential in Linked Open Data (LOD) datasets to identify the relationship type of RDF-triples and thus contribute to automatic processing of machine-readable statements of the Semantic Web. Typed links in RDF are referred to as the value of the `rdf:type` property, defining the relationship type using vocabulary terms or definitions from LOD datasets<sup>9</sup>.

The mentioned *URLs of a relation are connected in the form of graphs*. Such implementations therefore preserve the semantic variability of content types by using the same syntax in a similar way. In Linked Data, a set of information uses this syntax, which provides an interconnection via the main three components of natural languages (subject, predicate, object). Figure 3 describes the main dependencies.

To offer further prospects of the functionality, *the predicate defines the relation of an RDF-triple*. This will be treated later on.

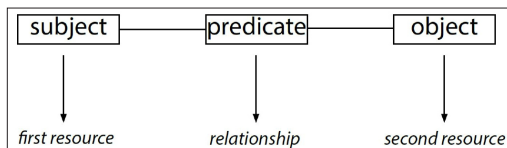


Fig. 3: RDF-triple (general)

## 8. Visualization

For the process of analysis, evaluation and identification of datasets visualizations are based on computational graphic systems<sup>10</sup>.

*Computer-based visualization systems provide visual representations of datasets* intended to help people carry out some tasks better. These visualization systems are often – but not always – interactive. Resource limitations include the *capacity of computers, of humans, and displays*. The area of possible visualization system design is huge and full of tradeoffs<sup>11</sup>.

## 9. Information visualization

Information visualization, a term coined by the User Interface Research Group at Xerox PARC, concentrates on the use of computer-assisted tools to explore a large number of abstract data. Useful applications of information visualization in computer programs involve selection, transformation and representation of abstract data in a form that facilitates human interaction for exploration and understanding. For this reason, information has to be *prepared, condensed and filtered* to enable effective utilization. In addition to this, perception, psychology and linguistics have to be integrated to deliver a genuine increase in surplus value for users. Central aspects of information visualization are the dynamics of visual representation and interactivity. Advanced technologies enable the user to modify visualization in real-time.

There are many tools available to create graphic representations of data nowadays. The integration of different scientific disciplines – like linguistics, information design, psychology of perception – necessitates preparation to render it generally understandable<sup>12</sup>.

## 10. Knowledge visualization

Knowledge visualization is used to represent and interpret complex information, text in particular, at the intersection of knowledge, art and cultural heritage. The use of visual representations to transfer knowledge between humans aims to improve the transfer of knowledge by using computer and non-computer-based visualization methods in a complementary fashion<sup>13</sup>.

While information visualization concentrates on the use of computer-assisted tools to gain new insight, knowledge visualization focuses on transferring insight and creating new knowledge in groups.

## 11. Visual communication

Visual communication is the communication of ideas through the visual display of information. It includes alphanumeric or artistic signs and electronic resources. Latest research in the field mainly concerns web design and graphically-oriented usability.

Visual analytics focuses on human interaction with visualization systems as part of a larger process of data analysis. Its focus is on human information discourse within dynamically changing information spaces. Visual analytics research concentrates on support for perceptual and cognitive operations that allow users to become aware of the expected and find out the unexpected in complex information spaces.

The concept visualization type deals with the meaning of a picture within a specific social and/or cultural context.

Visual perception laws include "Gestalt laws" which are commonly used.

Position as the most important part of design principles was researched by McKinlay<sup>24</sup>.

The figure shows a distinct change in the ranking of these tasks from quantity to quality.

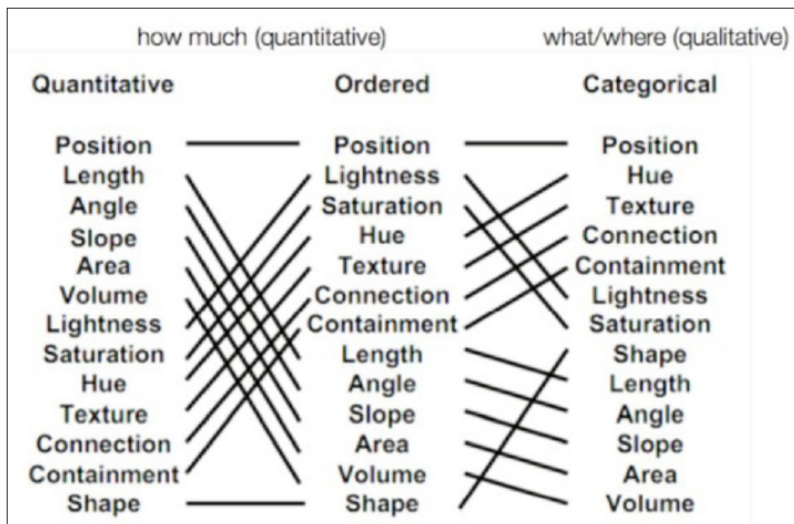


Fig. 4: Ranking of perceptual tasks<sup>14</sup>

To display the relevance of information components, positioning is highly important and therefore was examined in connection with the ranking of

the layout of all other graphical elements. As the figure shows, this layout changes when the quality of content increases.

## 12. Conceptual Models in Visualization

The science of visualization is in its infancy and involves the combination of graphics, imaging, data management and human perception. Two divergent trends in visualization have emerged. While visualization is often available and practical, the supporting technology is usually not designed for these applications. The second trend stems from the data glut problem. This means that *typical ad hoc approaches do not scale up to accommodate large, complex problems*. Despite competing requirements, access to data is the common barrier. A first step is to decompose visualization into a set of transformations that can highlight these limitations by defining a conceptual model and develop a taxonomy.

## 13. Linked Data Visualization Model

Applying information visualization techniques to the Semantic Web helps users to explore large quantities of data and interact with them. The main objectives of information visualization are to *transform data into and present data as a visual representation* in such a way that users can obtain a better understanding of it<sup>15</sup>.

Most existing work related to visualizing RDF is focused on concrete domains and concrete data types. The Linked Data Visualization Model (LDVM) is a formal base that exploits the Linked Data principles to ensure interoperability and *compatibility of compliant analytic and visualization components*. In short, LDVM allows users to create data visualization pipelines that consist of four stages:

- *Source data*
- *Analytical abstraction*
- *Visualization abstraction*
- *View*

The aim of LDVM is to provide means of creating reusable components at each stage that can be put together to create a pipeline even by non-expert users who do not know RDF. The *typical use case for visualization abstraction is to facilitate reuse* of existing analyzers and existing visualizers that work with similar data, only in different formats. For that purpose, LDVM uses a transformer. In the view stage, data is passed to a visualizer that creates

a user-friendly visualization. The components, when connected, create an analytic and a visualization pipeline which, when executed, takes data from a source and transforms it to produce a visualization at the end.

Linked Data browsers such as Tabular or Explanator allow users to navigate the graph structures and usually display property-value pairs in tables, but offer no broader view of the dataset<sup>16 17</sup>.

Fresnel is a vocabulary for rendering RDF to HTML, but its focus is on rendering instance data rather than on creating visualizations<sup>18</sup>.

Rhizomer provides an overview of the datasets and allows interacting with data through information architecture components such as navigation menus, breadcrumbs and facets<sup>19</sup>. It also provides visualizations such as maps and timelines.

DERI Pipes is an engine and graphical environment for general web data transformations and mashups<sup>20</sup>. It is not intended for lay-persons and requires software expertise.

The Linked Data Visualization Model (LDVM) uses the Data State Reference Model (DSRM) proposed by Chi as a conceptual framework.

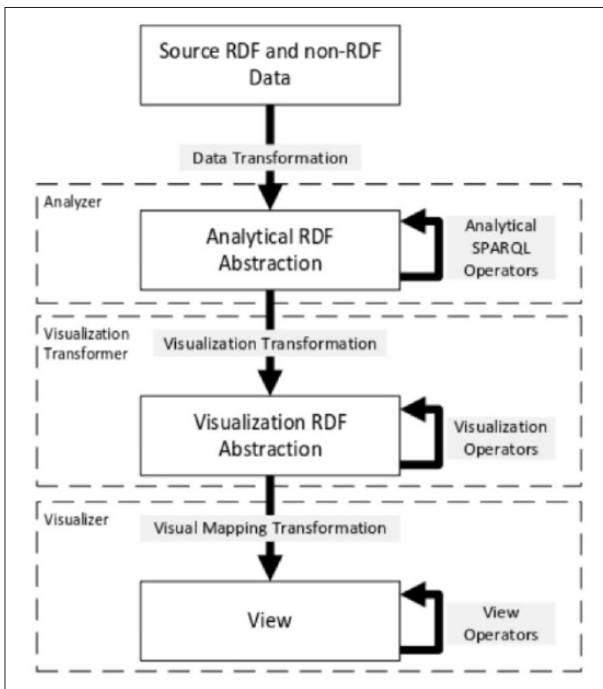


Fig. 5: High level LDVM overview<sup>15</sup>

DSRM describes the visualization process in a generic way<sup>15</sup>.

## 14. Prototyping the new visualization

To gain drafts of graph visualizations it is necessary to analyze why the existing layout standards are commonly used. The most important insights to construct visualization are gained by the components a visualization exhibits, and by their *noticeable design restrictions*. Components and restrictions concerning the visualization of Linked Data relations are itemized in the following list.

### Components

- *Nodes*: labeled components (signifiers that are extracted parts of the URLs of RDF-triples)
- *Edges*: visual lines between these labeled components (graphs that show the dependencies of a relation by orientation)

### Restrictions

- *Node size*: restricts the number of nodes displayed
- *Node position*: restricts the arrangement of nodes displayed
- *Edge length*: restricts the space between nodes
- *Edge formation*: restricts readability of information
- *View size*: restricts the space of viewable information
- *Node control*: restricts operations of manipulating nodes
- *View control*: restricts operations of manipulating view

Heuristics of general design principles treating application views give vertices for possible visualizations.

Figure 6 shows the process of such a refinement of one of these drafts.

The figure shows the first draft of the adapted visualization. The relations are arranged in a scrollable list here. The draft also includes an option to toggle the content of the abstract of each subject and object in the representation view.

From the input nodes (subjects) or the given nodes view additional found nodes are similar to objects in grammar. It is then possible to *filter identical predicates for direct relations* of two or more given input nodes while preserving the completeness of the path.

To normalize the relations I suggested limiting *identical predicates to a unique representation* that replaces several appearances of nodes having similar predicates in the relations. Semantically this method preserves the



meaning of information given by the relations. If the predicates pertaining to the relation have identical terms *these predicates can be normalized and the information is preserved.*

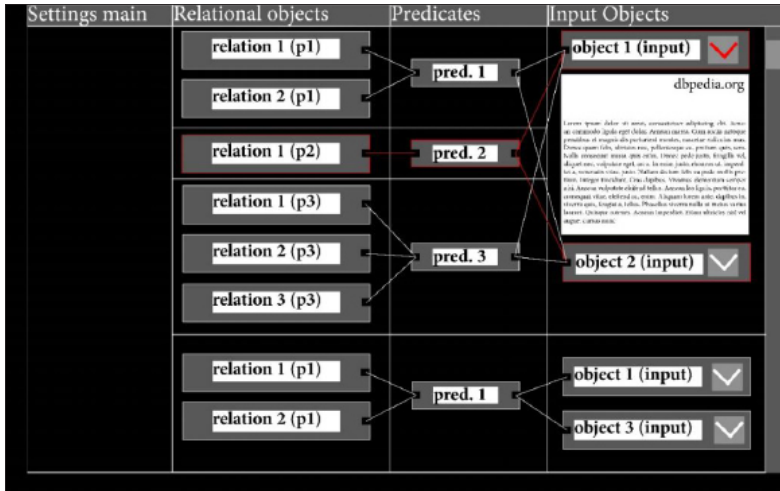


Fig. 6: Draft visualization

## 15. Introducing the views

**Node-link view (RelFinder)** – <http://www.visualdataweb.org/relfinder.php>

RelFinder is a graph visualization software based on "Adobe Flash" showing relations to specific objects or subjects of SPARQL queries using simple input fields. It calculates the relationships between two or more input phrases.

The software is based on flex framework 3.5 including the Spring graph visualization core. The Spring graph visualization enables software to link nodes (components of data) providing a network view.

Relations in RelFinder consist of RDF-triples representing the subject, predicate and object using the string component of the rdflabel already described as extracted part of these URLs.

### Linear view (RelFinder)

The linear view visualizes the relations in a list. Each node leads to a unified predicate and connects the given nodes over this predicate with the found nodes of a search query.

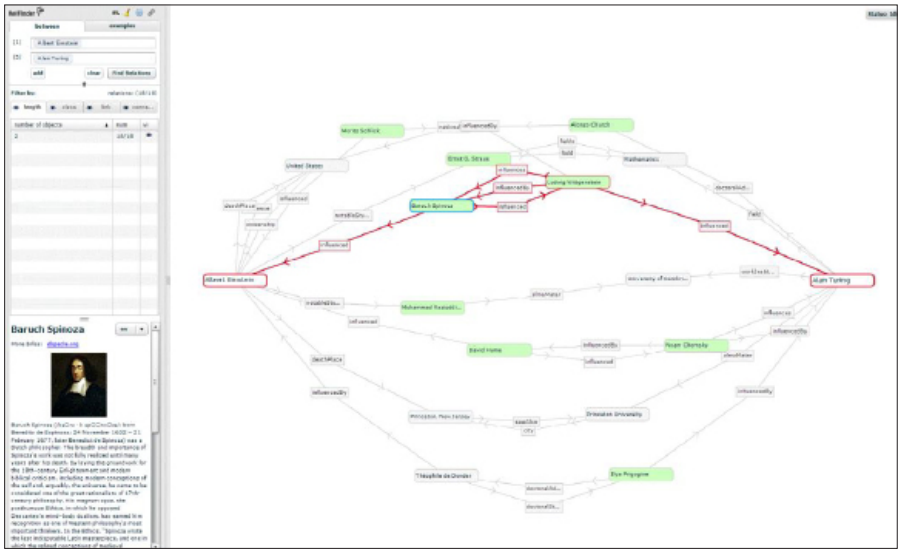


Fig. 7: RelFinder (node-link view)



Fig. 8: RelFinder (linear view)

Three main columns stand for three parts of a relation. The first column shows the given nodes, the second one the predicates while the third column is preserved for the found nodes of the relations. The button with the arrow next to the nodes opens an abstract (a box including some short information about the object).

### **Radial view (RelFinder)**

The radial view visualizes the found nodes' positions in a (semi)circle.

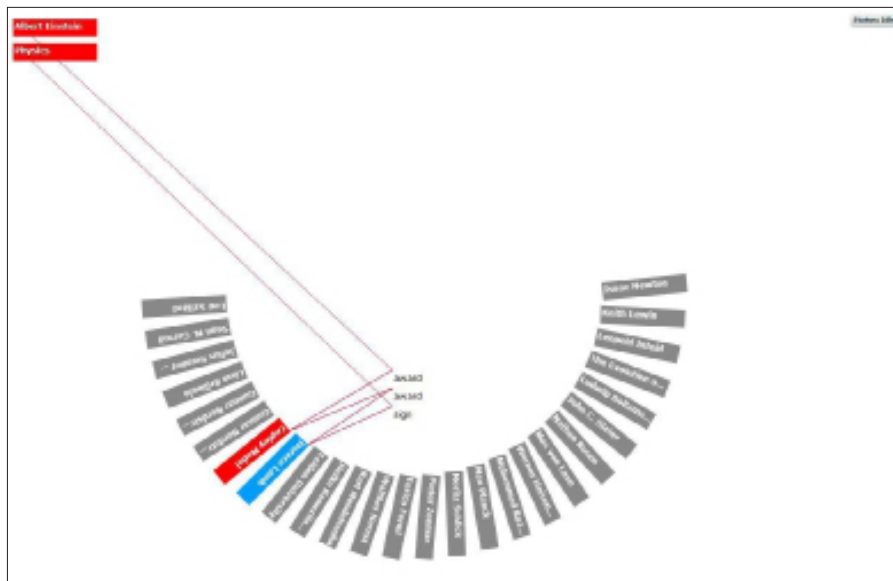


Fig. 9: RelFinder (radial view)

To describe the main points of reprogramming an equal list for the implementation of the list view is given below.

- *Normalizing predicates: Unification of similar predicates*
- *Toggable abstracts: Show and hide abstracts in the view*
- *Repositioning nodes as list/(semi)circle: Set ascendant x- and y-positions of nodes*
- *Highlighting specific paths: Hiding paths of unselected sub-relations*
- *Controllable components: Applying scrollbars to the view and drag nodes vertically*
- *Layout of components: Changing the visualization of labels and nodes*

Both views differ from the standard visualization of a node-link view common in Linked Data visualizations and therefore had to be tested in a prepared workflow in order to gain knowledge about interaction advantages and corresponding design principles.

## 16. Methods of testing

To find relevant methods concerning these questions the paper "*Task taxonomy for graph visualization*"<sup>21</sup> proved very helpful since it provided *essential tasks for the specification of results*. On the basis of the following listing the executed tasks describe the propositions for the methods.

### **Topology-Based Tasks**

- *Adjacency (direct connection): Find the set of nodes adjacent to a node*
- *Connectivity: Identify connected components*

### **Attribute-Based Tasks**

- *On the nodes: Find the nodes having a specific attribute value*
- *On the links: Given a node, find the nodes connected only by certain types of links*

### **Browsing Tasks**

- *Follow a given path*

### **Test subjects**

20 test subjects participated in the test. To describe their general characteristics a survey was held in addition to the test tasks. Exactly 10 of the test subjects were male and 10 female. 15 of the test subjects were between 21 and 39 years old.

11 of them had a general qualification, a bachelor's or a master's degree.

### **Materials**

The user test took place in the usability laboratory of the University of Applied Science St. Pölten, using an eye tracking device from SMI (Sensorimotoric Instruments).

It included the 2 different types of questions: one for navigation and one for information included in the abstract. To guide the test subjects

a textual briefing was given for each of the tasks before showing the test screen. The maximum duration of any task was set to 3 minutes, after which the next briefing faded in. The format of the eye tracking data was a *screen recording which included the scan path of the test user and the mouse clicks on the test screen.*

## 17. Test results

Each test subject had to learn to understand the visualization given and was asked to provide answers in text form. The results showing the correctness of the answers of all test subjects are shown below.

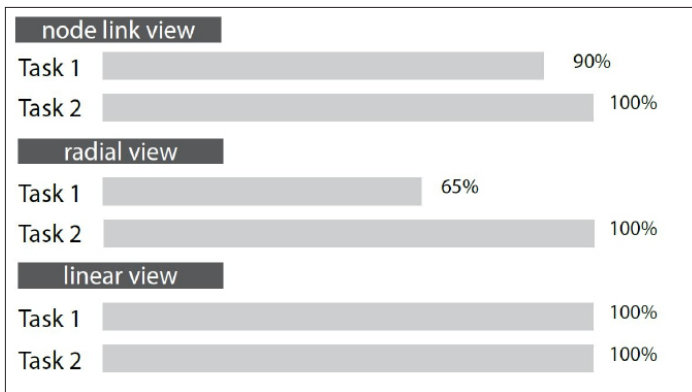


Fig. 10: Correctness of task answers

The second task depending on the information of the abstracts was always answered correctly.

The information specified by each test subject was denoted as true or false. Therefore the results clearly show that the answers given using the linear view were always right while the answers using the node-link view were wrong in 2 cases. The radial view had 7 wrong answers.

The results of task 1 that the following figure shows include the time the test subjects needed to find and click one node. The results as quartiles show that the *nodes were found fastest using the radial view and the linear view.*

Task 2 shows the time a person needed to find the information provided by the abstract. The linear view was the only view where the abstract had to be opened by clicking a button. Nevertheless the linear view could be handled very quickly and the concept was understood very easily.

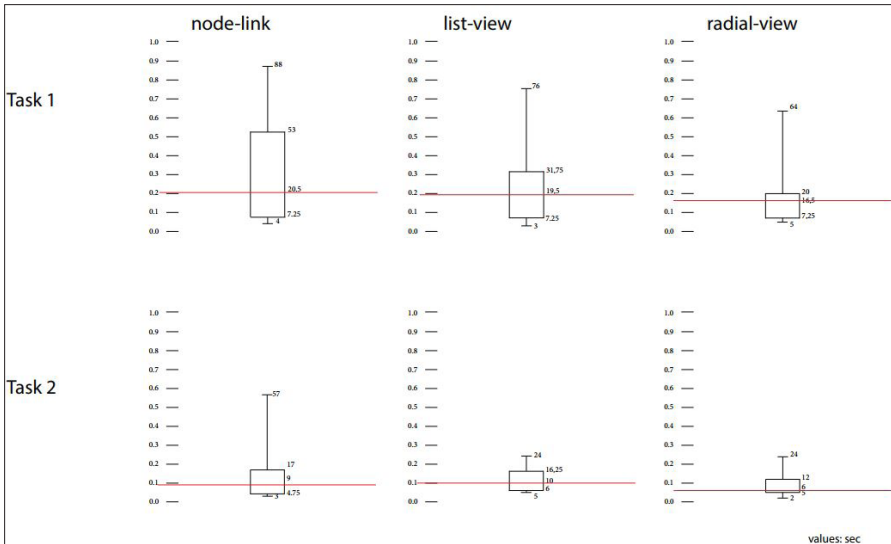


Fig. 11: Times (sec) to solve task 1 (find a node) and task 2 (abstract info)

In comparison to the node-link view we clearly see that the time needed for solving the first task in the *linear view and the radial view has very similar minima*. In the third quartile we see that the test subjects needed less time in both new views.

The main survey treated the three given visualizations in terms of preference, clearness and understanding and handling. According to this information *the visualization type preferred by the test subjects could be evaluated in relation to its clarity and handling*. To inform the test subjects about the visualization types a printed poster including a figure for each visualization type and a label naming these views was offered in order to make the test subject's choice easier. The charts below describe the results.

Question 1: Did you know Semantic Web and Linked Data before? (War Ihnen Semantic Web und Linked Data vor der Testung ein Begriff?)



Fig. 12: Semantic Web publicity

Question 2: Which version of RelFinder would you prefer? (Welche Version von RelFinder würden Sie bevorzugen?)

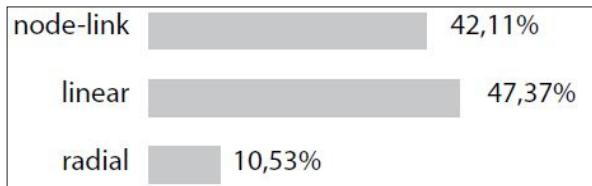


Fig. 13: Preference of visualization

Question 3: In your opinion, which view of RelFinder is more clearly arranged? (Welche Version von RelFinder ist Ihrer Meinung nach bezüglich der Darstellung übersichtlicher?)

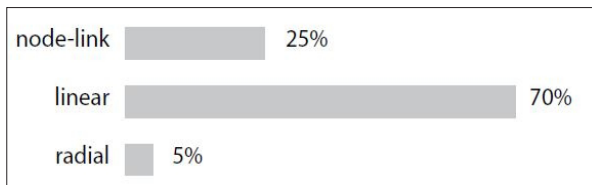


Fig. 14: Clarity of visualization

Question 4: In your opinion, which version of RelFinder is easier to handle? (Welche Version von RelFinder ist Ihrer Meinung nach leichter handhabbar?)

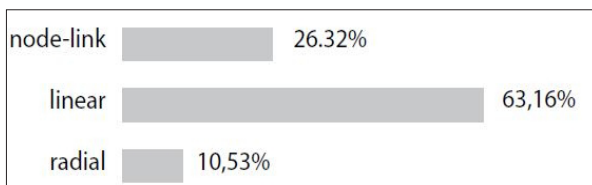


Fig. 15: Clarity of visualization

## 18. Conclusion and future work

The criteria of good design were utilized and compared with the test and the answers of the survey. In *terms of handling, preference and clarity, the line-*

*ar view's design was evaluated as best.* This may be due to its *minimalistic concept, which reduces the noise of unfiltered paths and nodes by repositioning them.*

*Faster handling for a broader number of users could be demonstrated for the radial and the linear view via the time the tasks took* in comparison. The nodes were found very much faster than in the node-link view. Proof of concept, however, would demand a larger sample of test subjects.

*The placement of the abstract as a toggleable container of the main view was often recognized intuitively* and needed no further explanation. In times of responsive design, the vast number of at times very large screen resolutions profits immensely from toggleable information boxes to organize nodes.

Another find is that *no statement referring to the normalized predicates* of the linear view was given, which might indicate that this feature was not recognized. One interesting point is that the users construed Linked Data in a very abstract way. Most of the time they had difficulties to put the visualized sentence into words. In terms of the translation of information into another language some common Linked data predicates were misunderstood.

*The unordered loading of the nodes found was also criticized* by some test subjects. The test subjects suggested that the nodes have to be ordered alphabetically in the linear and the radial view especially.

*The visualization of nodes in the radial view should be turned 180° so that readability is better.* As mentioned, the performance of findability is very good in the radial view, but the visualization space for nodes positioned in such a way is limited.

Nevertheless a *graph oriented visualization of data presented as a list seems to have synergies with the presentation of regular lists* when it comes to user habits and user demands.

An improvement already implemented is the finalization of the radial view by adding graph directions represented as arrows. All data of the prototypes developed and the most important information is hosted on a github repository available here: <https://github.com/geonb/Visualization-of-typed-links-in-Linked-Data>.

Where the Semantic Web in general is concerned, the *grammar of relative clauses visualized as oriented graphs cannot be used exclusively in dependence on certain vocabularies that represent equal exchange of information.* A group of predicates like "communicate", "meet", "talk" *could be linked without including the direction of the graphs.*

To outline future work on the prototypes of RelFinder, it is necessary to follow the rules of visual representation as well as the results of this evaluation. Better design seems to include *hiding of unnecessary informati-*



*on and better positioning of the nodes.* To briefly illustrate what could be improved concerning the linear and the radial view, one might say that the found nodes (objects) of a *normalized list of predicates should be grouped around similar normalized predicates and could appear in an alignment centered around a clicked found node.* One topic in need of further investigation is optimizable graph length, i.e. the distance of the node lists between the nodes.

From the perspective of a developer, all the modern visualizations of Linked Data in connection to a model view controller concept should *prepare the result set of a query inside the controller completely* and draw the entire visualization after this calculation.

Finally, to preserve the semantic sense of the relationships of typed links, I suggest a presentation where the *composition of all relationships, including subjects and normalizable predicates dependent on sortable and filterable object groups, represents a result set as a vertically positioned node-list view.*

Dipl.-Ing. Georg Neubauer, BSc  
Danube University Krems, Library  
E-Mail: [georg.neubauer@donau-uni.ac.at](mailto:georg.neubauer@donau-uni.ac.at)

## Notes

- 1 N. Bartelme, „Einführung in Geoinformatik“. Springer Berlin Heidelberg, 2005, pp. 1–42. DOI: [http://dx.doi.org/10.1007/3-540-27201-1\\_1](http://dx.doi.org/10.1007/3-540-27201-1_1)
- 2 Hernando Sáenz-Sánchez, „Visualisierung von Linked Open Data mit Orts- und Zeitbezug“, Institut für Informatik der Freien Universität Berlin, 2012. <http://www.inf.fu-berlin.de/inst/ag-se/theses/Saenz-Sanchez13-visual-geo.pdf>
- 3 W. Borst, "Construction of Engineering Ontologies for Knowledge Sharing and Reuse", 1997, Centre for Telematics and Information Technology, University of Twente, Enschede-Noord, the Netherlands, 2006. <http://eprints.eemcs.utwente.nl/17377/01/t0000004.pdf>
- 4 Heiner Stuckenschmidt, „Erstellen von Ontologien“, in Ontologien Konzepte, Technologien und Anwendungen, ser. Informatik im Fokus. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. 5, pp. 155–205.
- 5 T. Berners-Lee, R. Fielding, and L. Masinter, RFC 3986, "Uniform Resource Identifier (URI): Generic Syntax", 2005.
- 6 F. Manola and E. Miller, "Resource description framework (RDF) primer", W3C Recommendation, vol. 10, 2004.

- 7 Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak, "*Linking open data cloud diagram*", 2014. <http://lod-cloud.net/>
- 8 Robin Cover, "*XML and Web Services In The News*", xml.org. 6 October 2006. <http://www.xml.org/xml/news/archives/archive.10062006.shtml>
- 9 Dubost Karl, "*Web site metadata profile: favicon*", World Wide Web Consortium. October 2005. <http://www.w3.org/2005/10/profile>
- 10 Thomas A. Runkler, „*Visualisierung in Data Mining: Methoden und Algorithmen intelligenter Datenanalyse*“, 2010, pp. 35–55. [http://link.springer.com/content/pdf/10.1007%2F978-3-8348-9353-6\\_4](http://link.springer.com/content/pdf/10.1007%2F978-3-8348-9353-6_4)
- 11 T. Munzner, "*Visualization Design and Analysis: Abstractions, Principles, and Methods*", 2012. <http://web.cse.ohio-state.edu/~raghu/teaching/CSE5544/ClassLectures/PDF-old/02-Chapter1.pdf>
- 12 C. Moser, "*Informationsdesign, in User Experience Design*", ser. X.media.press. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 105–120. <http://www.springerlink.com/index/10.1007/978-3-642-13363-3;> <http://www.springerlink.com/index/10.1007/978-3-642-05404-4>
- 13 Remo Aslak Burkhard, "*Tube map: Evaluation of a visual metaphor for inter-functional communication of complex projects*", 2004. [https://www.researchgate.net/publication/229100758\\_Tube\\_Map\\_Evaluation\\_of\\_a\\_Visual\\_Metaphor\\_for\\_Interfunctional\\_Communication\\_of\\_Complex\\_Projects](https://www.researchgate.net/publication/229100758_Tube_Map_Evaluation_of_a_Visual_Metaphor_for_Interfunctional_Communication_of_Complex_Projects)
- 14 J. D. Mackinlay, "*Automating the Design of Graphical Presentations of Relational Information*", ACM Transactions on Graphics 5(2):110–151, 1986. [https://www.researchgate.net/publication/220184008\\_Automating\\_the\\_Design\\_of\\_Graphical\\_Presentations\\_of\\_Relational\\_Information](https://www.researchgate.net/publication/220184008_Automating_the_Design_of_Graphical_Presentations_of_Relational_Information)
- 15 M. Brunetti, S. Auer, R. Garcia, J. Klímak, M. Necsély, "*Formal Linked Data Visualization Model*", ACM New York, NY, USA 2013. [http://ceur-ws.org/Vol-914/paper\\_29.pdf](http://ceur-ws.org/Vol-914/paper_29.pdf)
- 16 T. Berners-Lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer, and D. Sheets, "*Tabulator: Exploring and analyzing linked data on the semantic web*", In 3rd Int. Semantic Web User Interaction WS, 2006.
- 17 S. Araujo, D. Shwabe, and S. Barbosa. "*Experimenting with explorer: a direct manipulation generic rdf browser and querying tool*", In WS on Visual Interfaces to the Social and the Semantic Web (VISSW2009), 2009.
- 18 E. Pietriga, C. Bizer, D. R. Karger, and R. Lee, "*Fresnel: A browser-independent presentation vocabulary for rdf*", In I. F. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, The Semantic Web - ISWC 2006, 5th International Semantic

- Web Conference, ISWC 2006, Athens, GA, USA, November 5–9, 2006, Proceedings, volume 4273 of Lecture Notes in Computer Science, pp. 158–171. Springer, 2006.
- 19 J. Brunetti, R. Gil, and R. Garcia, "*Facets and Pivoting for Flexible and Usable Linked Data Exploration*", In Interacting with Linked Data Workshop, ILD'12, Crete, Greece, May 2012.
- 20 Danh Le Phuoc, Axel Polleres, Giovanni Tummarello, Christian Morbidoni, "*DERI Pipes: visual tool for wiring Web data sources*", 2008. [https://www.researchgate.net/publication/228733128\\_DERI\\_Pipes\\_visual\\_tool\\_for\\_wiring\\_Web\\_data\\_sources](https://www.researchgate.net/publication/228733128_DERI_Pipes_visual_tool_for_wiring_Web_data_sources)
- 21 Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry, "*Task taxonomy for graph visualization*", 2006. In Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization (BELIV '06). ACM, New York, NY, USA, pp. 1–5. DOI: <http://dx.doi.org/10.1145/1168149.1168168>

## ■ METADATA FOR RESEARCH DATA IN PRACTICE

by *Barbara Petritsch*

**Abstract:** *What data is needed about data? Describing the process to answer this question for the institutional data repository IST DataRep.*

**Keywords:** *Data repository; data publication; Eprints; DataCite; metadata*

### METADATEN FÜR FORSCHUNGSDATEN IN DER PRAXIS

**Zusammenfassung:** *Welche Daten über Daten brauchen wir? Die Beschreibung des Prozesses diese Frage für das institutionelle Datenrepositorium IST DataRep zu beantworten.*

**Schlüsselwörter:** *Datenrepositorium; Datenpublikation; Eprints; DataCite; Metadaten*

#### Contents

1. Introduction
2. IST DataRep project
  - 2.1. Data Publication
  - 2.2. Data Format & Data Types
  - 2.3. User Tests
  - 2.4. Adaptation
3. Dublin Core, DataCite Metadata Schema and IST DataRep metadata fields
4. Conclusion



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

The *Institute of Science and Technology Austria (IST Austria)* is a young international institution dedicated to basic research and graduate education in the fields of life sciences, formal and physical sciences, located in Klosterneuburg on the outskirts of Vienna. Still growing, IST Austria is committed to conducting world-class research. By 2026, up to 90 research groups will perform research in an international state-of-the-art environment.

## 1. Introduction

In order to be prepared for the growing need of data publication, IST Austria decided to build a data repository in 2012. The library of IST Austria engaged in the tasks of researching the field and implementing the repository, which is in operation since August 2015. In this article I describe IST DataRep in regard to its metadata and how this was shaped by our approach.

For us it was crucial that the users of this tool – the researchers – were involved throughout the whole process of implementation. At first, a survey on the actual state of research at IST Austria was carried out to get a better idea of the requirements for and needs of our scientists. Later in the process, when we had already decided on the open source software Eprints, they took part in extensive user tests. The scope of these tests was to adapt the depositing workflow to our researchers' needs and wishes. Implicit to this process was the design of the metadata set.

## 2. IST DataRep project

In the course of the IST DataRep project various aspects determined the final metadata set:

### 2.1. Data Publication

The top priority for the institutional data repository was to provide the opportunity for affiliated scientists of all disciplines to publish their data (underlying publications) especially if they lack a suitable subject repository. One key feature of academic publishing is that it renders scientific content citable. As advancement on URLs – which tend to be inconsistent and unreliable – the concept of persistent identifiers was introduced for digital publishing as a permanent addressing mechanism. The most popular and best-known persistent identifier in academic publishing is the Digital

Object Identifier (DOI). Therefore the obvious choice for IST DataRep was to register with DataCite, which provides DOIs especially for research data and university publications (theses, technical reports etc.).

## **2.2. Data Format & Data Types**

Due to the interdisciplinary scope of research conducted at IST Austria, the institutional survey showed that a high variety of data types and formats are created.<sup>1</sup> With this precondition we were facing the usual role of an institutional repository: to take care of the "long tail" of research<sup>2</sup>. This implies universal values for data description and the consideration of all kinds of data types and formats. With this set of requirements we were equipped to evaluate the available proprietary and open source products, which resulted in us choosing the open source software Eprints.<sup>3</sup>

## **2.3. User Tests**

After implementing Eprints in combination with the plugin ReCollect<sup>4</sup> we ran extensive user tests with the barely surprising result that researchers wish to have an easy-to-use tool, which enables a quick upload. For the purpose of comparison they also tested Figshare. One distinctive feedback was that the test persons greatly appreciated metadata presets during upload. We therefore adopted this feature for IST DataRep.

## **2.4. Adaptation**

Adapting Eprints was a process of considering the institutional concept on the one hand and complying with external factors on the other. One influential factor of metadata design was the metadata schema of DataCite<sup>5</sup> as a precondition for minting DOIs. It consists of mandatory, recommended and optional terms, which provides (us with) some orientation, flexibility and limits. It is based on Dublin Core<sup>6</sup> but with additional regulations and restrictions.

We customized our fields considering the institutional needs and our researchers' request for a convenient workflow. Therefore our initial effort was to limit the amount of fields required to fill in during deposit.<sup>7</sup> Because the researchers themselves, who most likely have no specific cataloging skills, carry out the data upload we focused on implementing a tool which is usable for them. This also implied to rethink field names to maximize clarity from the researcher's/user's perspective.

### 3. Dublin Core, DataCite Metadata Schema and IST DataRep metadata fields

| Dublin Core<br>Metadata terms |            |
|-------------------------------|------------|
| Term Names                    | Obligation |
| Contributor                   | O          |
| Coverage                      | O          |
| Creator                       | O          |
| Date                          | O          |
| Description                   | O          |
| Format                        | O          |
| Identifier                    | O          |
| Language                      | O          |
| Publisher                     | O          |
| Relation                      | O          |
| Rights                        | O          |
| Source                        | O          |
| Subject                       | O          |
| Title                         | O          |
| Type                          | O          |

Fig. 1: Dublin Core Metadata terms

| Data Cite<br>Metadata schema |                      |       |
|------------------------------|----------------------|-------|
| Obligation                   | Metadata Properties  | Order |
| M                            | Identifier           | 1     |
| M                            | Creator              | 2     |
| M                            | Title                | 3     |
| M                            | Publisher            | 4     |
| M                            | Publication Year     | 5     |
| R                            | Subject              | 6     |
| R                            | Contributor          | 7     |
| R                            | Date                 | 8     |
| O                            | Language             | 9     |
| R                            | Resource Type        | 10    |
| O                            | Alternate Identifier | 11    |
| R                            | Related Identifier   | 12    |
| O                            | Size                 | 13    |
| O                            | Format               | 14    |
| O                            | Version              | 15    |
| O                            | Rights               | 16    |
| R                            | Description          | 17    |
| R                            | GeoLocation          | 18    |

Fig. 2: DataCite Metadata schema

Tracing back the schema of the metadata fields used in IST DataRep, one ends up with Dublin Core. Dublin Core is a metadata schema developed by the DCMI (Dublin Core Metadata Initiative, founded in 1994) at the OCLC/NCSA Metadata Workshop in Dublin, Ohio in 1995. It is a simple set of 15 core elements for resource description intended to facilitate the discovery of digital objects. All of these elements are optional (O), repeatable, and they have no specific order (Fig. 1). It was designed specifically for non-catalogers.<sup>8</sup>

DataCite was founded as a consortium by leading research libraries and information centers in 2009.<sup>9</sup> Its aim is to allow easy online access to research data and improve its citability. DataCite's Metadata Schema is based on Dublin Core with more regulations and restrictions, which clearly indicates a librarian's approach. It has 18 elements (terms) set

in a specific order with three different levels of obligation: mandatory (**M**), recommended (R) and optional (**O**) (Fig. 2). The element types are definable via subfields, e.g. the term contributor has to be refined as researcher, supervisor, advisor etc.<sup>10</sup>

The metadata form of IST DataRep consists of 28 fields in total with different qualities: free text fields to fill in manually (10), option menus to select (5), presets (5) which are still modifiable, and generated elements (8) (see exemplary fields in Fig. 4). Most fields can be mapped to DataCite's elements (Fig. 3). One option for the mapping process is automatically via the DataCite DOI plugin for Eprints. Subsequently, this information is transferred via an API to DataCite for DOI registration. The alternative is to prepare an XML file manually and register a DOI via DataCite's online metadata shop.

| IST DataRep       |                     |                               |                    |
|-------------------|---------------------|-------------------------------|--------------------|
| Metadata fields   |                     |                               |                    |
|                   | DataCite Obligation | Field Name                    | Field Type         |
| <b>File</b>       |                     |                               |                    |
| R                 | 17                  | File description              | O Free Text        |
|                   |                     | Visible to                    | M Preset/ Menu     |
|                   |                     | Content                       | M Menu             |
| O                 | 16                  | Licence                       | M Preset/ Menu     |
| R                 | 10                  | Type                          | M Preset/ Menu     |
| R                 | 8                   | Embargo Date                  | O Menu             |
| <b>Collection</b> |                     |                               |                    |
| M                 | 3                   | Title                         | M Free Text        |
| R                 | 17                  | Collection Discription        | M Free Text        |
|                   |                     | Keywords                      | M Free Text        |
| M                 | 2                   | Creators                      | M Free Text        |
| R                 | 7                   | Contributors                  | O Menu + Free Text |
| R                 | 12                  | Original Publication Details: |                    |
|                   |                     | Status                        | M Menu             |
|                   |                     | DOI                           | O Free Text        |
|                   |                     | URL                           | O Free Text        |
| R                 | 7                   | Research Funders              | O Menu + Free Text |
|                   |                     | Research Project Title        | O Free Text        |
| R                 | (7)                 | Contact Email Adress          | M Preset           |
| O                 | 9                   | Resource Language             | M Preset           |
| R                 | 8                   | Time Period                   | O Menu             |
| M                 | 4                   | Publisher                     | M Free Text        |
| M                 | 1                   | Data DOI                      | G Generated        |
| R                 | 6                   | Subjects                      | M Menu             |
| <b>Eprints</b>    |                     |                               |                    |
|                   |                     | Depositing User               | G Generated        |
| M                 | 5                   | Date Deposited                | G Generated        |
| R                 | 8                   | Last Modified                 | G Generated        |
|                   |                     | Metadata Revision             | G Displayed        |
| O                 | 11                  | URI                           | G Generated        |
|                   |                     | Citation                      | G Generated        |
| O                 | 14                  | Format                        | G Displayed        |
| O                 | 13                  | Size                          | G Displayed        |

Fig. 3: IST DataRep Metadata fields

#### 4. Conclusion

For the library of IST Austria it was crucial to create a tool that is usable, understandable, practical and tailored to specific needs. One main approach to the project was to focus on only one challenge at a time. Our first priority was data publication. After the repository was available for data deposit/publication we were aiming at a long-term storage solution. At the moment we are participating in the H2020 EUDAT project with the objective to establish off-site long-term storage for the deposited data, which is also crucial to the warranty of DOIs: to facilitate permanent access. In 2017 we will focus on two projects: one of which is very hands-on,



IST AUSTRIA  
Institute of Science and Technology

IST DataRep

Home About Browse by Subject Browse by Author Browse by Research Group Browse by Year Repository Policies

Dropdown menu / M

- Upload File.png  
362Kb

License: Creative Commons: Public Domain Dedication - CC0 1.0  
Recommended licence, which reduces legal and technical impediments to the reuse of data by waiving copyright and related rights to the extent permitted by law. Is used as default standard by leading research data repositories. License summary

Free text / M

Collection Description  
Please describe your data as well as you can. Add as much context as possible so that others can interpret your research and reproduce it. Data collection and processing method can be described here.

Options menu / M + Free text / O

Original Publication Details

Status: Please select the current status of related publication.  
 Published  
 In Press  
 Submitted  
 Unpublished

DOI: Related publication DOI

URL: Related publication URL

System generated / M

Contact Email Address  
The contact email address for this item. If the full-text is not available to the public, then requests to view the full-text will be sent to this email. The email address will not be made public.  
repository.manager@ist.ac.at

Generated in review process

Data DOI  
DOI of this Data Collection

Fig. 4: Upload form

focusing on seamless workflows for data publication and another, which is debating the various aspects of research data on an institutional basis to achieve a consent in definition but also an understanding of the distinctive perspectives on research data.

After an initial phase of little awareness of the RDM (Research Data Management) services at IST Austria, researchers are now approaching the library for advice and cooperation. The increasing interest in data manage-

ment, publication and preservation is certainly due to the recent uptake of these issues in national and international funding policies as well as publishers' data policies. For us this proves that the implementation of the data repository happened at the right time and that the library of IST Austria is well prepared for future tasks of data management.

Barbara Petritsch

ORCID: <http://orcid.org/0000-0003-2724-4614>

Institute of Science and Technology Austria (IST Austria), Library

E-Mail: [barbara.petritsch@ist.ac.at](mailto:barbara.petritsch@ist.ac.at)

- 1 Porsche, Jana: *Actual state of research data @ ISTAustria*. Project Report, IST Austria 2012. Online: <https://repository.ist.ac.at/103/>.
- 2 Tail (small, medium) data is, in contrast to head (big) data, described as heterogeneous, hand generated and created and processed under individual procedures. Therefore the curation is not likely to be organised centrally and/or automatically (i.e. subject repository). Furthermore, accessibility is a major issue due to legal uncertainties and lack of effective platforms (i.e. institutional repositories). See also: Heidon, Bryan P: *Shedding Light on the Dark Data in the Long Tail of Science*. Library Trends 2008, 57, 2, 280–299. Online: <http://hdl.handle.net/2142/10672>.
- 3 For further information on this process see: Porsche, Jana. *Technical requirements and features*. Project Report. IST Austria 2013. Online: <https://repository.ist.ac.at/135/>.
- 4 The ReCollect plugin transforms an EPrints install into a research data repository with an expanded metadata profile for describing research data (based on DataCite, INSPIRE and DDI standards) and a redesigned data catalogue for presenting complex collections. Developed by the UK Data Archive and the University of Essex, as part of the JISC MRD Research Data @Essex project. See also: <http://bazaar.eprints.org/367/> (28.3.2017).
- 5 <https://schema.datacite.org/>.
- 6 <http://dublincore.org/documents/dces/>.
- 7 ReCollect has 46 fields by default. We dropped the fields: alternative title, corporate creators, collection date, Bounding Box North, Bounding Box East, Bounding Box West, Bounding Box South, Data Collection Method, Legal and ethical issues, Lineage, Additional Information, Original Data Publisher, Restrictions, Copyright Holder, Note. See also: [http://www.data-archive.ac.uk/media/375386/rde\\_eprints\\_metadata\\_profile.pdf](http://www.data-archive.ac.uk/media/375386/rde_eprints_metadata_profile.pdf) (28.3.2017).

- 8 Weibel, Stuart: *The Dublin Core: A Simple Content Description Model for Electronic Resources*. Bul. Am. Soc. Info. Sci. Tech. 1997, 24, 9–11. DOI: <https://dx.doi.org/10.1002/bult.70>.
- 9 Brase, Jan: *European Initiative to Facilitate Access to Research Data*. D-Lib Magazine 2009, 15, 5/6. DOI: <https://dx.doi.org/10.1045/may2009-inbrief>.
- 10 Depending on the specific field subfields are mandatory or optional.

## ■ CHALLENGES OF BUILDING AND MAINTAINING AN IMAGE DATABASE: A USE CASE BASED ON THE DIGITAL RESEARCH ARCHIVE FOR BYZANTIUM (DIFAB)

by *Nina Rannharter* & *Sarah Teetor*

**Abstract:** *Due to the complex nature of archival images, it is an ongoing challenge to establish a metadata architecture and metadata standards that are easy to navigate and take into consideration future requirements. This contribution will present a use case in the humanities based on the Digital Research Archive for Byzantium (DiFAB) at the University of Vienna. By tracing one monument and its photographic documentation, this paper will highlight some issues concerning metadata for images of material culture, such as: various analog and digital forms of documentation; available thesauri – including problems of historical geography, multilingualism, and culturally specific terminologies –; and the importance of both precise and imprecise dating for cultural historians and their research archives.*

**Keywords:** *digital archive; metadata standards; image database; Digital Humanities; controlled vocabulary*

## HERAUSFORDERUNGEN BEIM AUFBAU UND BEI DER PFLEGE EINER BILDDATENBANK – EIN ANWENDUNGSBEISPIEL AUS DEM DIGITALEN FORSCHUNGSARCHIV BYZANZ (DIFAB)

**Zusammenfassung:** *Mit Metadaten versehene, digital archivierte Bilder sind aufgrund ihrer hohen Komplexität eine ständige Herausforderung für den Aufbau von Metadatenstrukturen und für das Beibehalten von adäquaten Metadatenstandards in Bilddatenbanken. Besonders bei Bildern über eine materielle Kultur zeichnet sich diese Komplexität in den diversen Ebenen ab. Zur Verdeutlichung dieser Problematik dient ein einzelnes Monument mit seiner fotografischen Dokumentation als geisteswissenschaftliches Anwendungsbeispiel aus dem Digitalen Forschungsarchiv Byzanz (DiFAB) der Universität Wien. Unter der grundsätzlichen Berücksichtigung der künftigen Anforderungen und der einfachen Navigation von Metadaten beinhaltet dieser Beitrag u.a. die Herausforderungen mit den diversen analogen und digitalen Formen der Dokumentation, der historischen Geographie, der kulturspezifischen Terminologie sowie mit der kunsthistorischen Bedeutung einer exakten und unscharfen Datierung für Forschungsarchive.*

**Schlüsselwörter:** *digitales Archiv; Metadaten-Standards; Bilddatenbank; Digital Humanities; controlled vocabulary*



## **Contents**

1. *Introduction*
2. *Forms of documentation*
3. *Thesauri and categorization schemes*
4. *Dating*
5. *Closing remarks and future prospects*

## **1. Introduction**

The signs and words that constitute a language are our main form of communication regardless of historical period, the medium of their transmission, or their context. When art historians, who work primarily with physical objects and images, are charged with the task of translating their visual material, or data, into a black-on-white or a binary language, several challenges must be met. In addition to the task of registering a complex material object in few words or numbers, it can prove a challenge first to find a common language with colleagues from other disciplines, and second to convey our requirements and aspirations when it comes to generating metadata and establishing metadata standards for objects of material culture.

This paper, which was originally presented in June 2016 at the workshop of Cluster I of the e-Infrastructures Austria project, sponsored by the Austrian Federal Ministry of Science, Research and Economy (BWF) <sup>1</sup>, will explore some specific issues pertaining to metadata for object-centered and visual humanities based on the use case of the Digital Research Archive for Byzantium (DiFAB) <sup>2</sup>. It will focus on three main aspects of generating appropriate, high-quality, and sustainable metadata for a digital research archive rooted in material culture: the different forms of documentation; the use of thesauri and categorization schemes, specifically for historical geography, multilingualism, and culturally specific terminologies; and the importance of precise and imprecise dating.

The Digital Research Archive for Byzantium presents a particularly complicated case due to the nature of the object of investigation – the cultural sphere of the Byzantine empire, a largely Greek-speaking, Christian empire rooted in Roman state ideology that ruled much of the eastern Mediterranean area from the 4<sup>th</sup> to 15<sup>th</sup> centuries and whose influence persists until today in religious and artistic traditions from modern Italy to Egypt,

and from the Middle East and the Balkans to Russia. This initiative was founded in 2005 by Lioba Theis, professor for Byzantine Art History at the Department of Art History of the University of Vienna. At first based on UNIDAM<sup>3</sup>, the project adopted PHAIDRA<sup>4</sup> for its long-term digital asset management and archiving in 2008, and contributed actively to early discussions of metadata requirements including the integration of internationally recognized metadata standards, such as Dublin Core. The main aims of DiFAB are:

1. to compile photographic documentation of the material culture of the Byzantine cultural sphere in digital form; and
2. to make these images available to researchers, students and interested members of the general public both now and in the future.

The principal sources of visual material constitute the historic photographs collection of the Department of Art History of the University of Vienna, which date from the late 19<sup>th</sup> century to the 1960s, private archives of eminent scholars in the field from the 1950s to around 2005, and born-digital photographs from field trips undertaken by project members since 2006.

On the following pages, we will discuss the example of one monument, the Church of St. Clement in Ohrid, Former Yugoslav Republic of Macedonia (FYROM), and several digital assets that describe it in order to highlight certain necessary considerations for functional and sustainable metadata for the digital documentation of tangible cultural history.

## **2. Forms of documentation**

The very complexity of creating an image database for cultural heritage is dictated by the long history of photography and developments in its use as a tool for documentation, research and teaching. In addition to the image content, to which we shall return, it is important to recognize the various types of media encountered and the different "life phases" that an originally analog image experiences (fig. 1), each with its own level of information and set of required metadata. An image, such as that in fig. 2, begins with its content, the real, three-dimensional object [1] – in this case the church building – which is then captured on a physical object, a glass-plate or film negative [2] and developed onto another physical object, a print on paper or a slide [3], before being scanned to create the digital asset [4]. Each of these stages requires core metadata pertaining to title/identifier, material, medium, dimensions, geolocation, and date, etc.

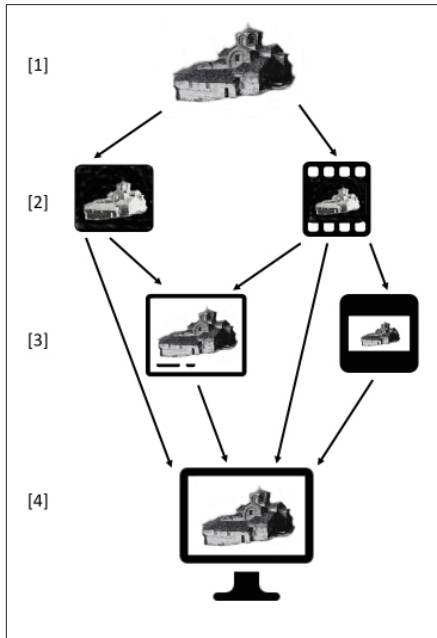


Fig. 1: Diagram of the "life phases" of an image: the real, three-dimensional object [1]; a glass-plate or film negative, a physical object [2]; a print on paper or a slide, a physical object [3]; the digital asset [4]. Icons made by Dave Gandy (film), Google (slide), freepik (photographic print) and Situ Herrera (computer) from [www.flaticon.com](http://www.flaticon.com), [CC 3.0 BY license](https://creativecommons.org/licenses/by/3.0/); modified by authors.



Fig. 2: Church of St. Clement, Ohrid, FYROM. View from the southeast, unknown photographer, undated (early 20<sup>th</sup> century), collection of the Photographic Archive of the Department of Art History, University of Vienna. Digital image available at <http://phaidra.univie.ac.at/o:526660> (accessed 19.5.2017), [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

In general, two types of digital image repositories have grown out of two different approaches: a curated, web-based archive and resource, of which Artstor<sup>5</sup> is a popular example for art history; and digitized archives, which have been implemented by countless libraries and scientific foundations in recent years for the express purpose of making their analog collections available online. While the web-based collection that draws its metadata from multiple sources relies on a minimum of "tombstone" metadata centered on the image content – object/type, work title, artist, date, location, source and copyright – the digitized archive reflects the traditions of library and information sciences by privileging metadata related not only to title, creator, date and copyright, but also medium, reproduction number, inventory or call number, repository and provenance. As a web-based digital image archive DiFAB attempts to combine these two approaches, focusing on both the image content and the image carrier, and to enhance each digital asset through comprehensive descriptor metadata, in order to ensure retrievability and usability in future research applications.

With this combined approach, each of the up to four life stages mentioned above requires in theory its own set of standardized metadata that supply information, at the least, about the "five Ws" – what, where, who, when and why<sup>6</sup> – of that layer. Currently, the PHAIDRA digital asset management system offers all the necessary metadata fields, many of them expandable to accommodate multiple entries. Images are uploaded and the metadata entered using the Phaidra Bagger interface, a mass upload tool; though all fields from the perspective of photograph archiving are present, their arrangement does not allow for a clear relationship of image content, image carrier(s) and digital asset with their respective metadata,<sup>7</sup> which would be optimal from an art historical and archival standpoint. As we will see, the human user can often intuitively assign metadata regarding object/title, geography and date, etc. to the correct "life phase" though these are not currently in a clearly structured form incorporating machine-readable relationships.

### **3. Thesauri and categorization schemes**

Establishing the appropriate metadata structure is the battle only half won. The next step, controlled entry, is essential to ensure regularity of metadata entries, their searchability and hopefully also their future compatibility. One tactic that DiFAB and many other online databases have adopted is the use of openly accessible thesauri and categorization schemes. At their current



state of development, such efforts are in particular applicable to the factors "what", "where" and to some extent "who"<sup>8</sup> – or object/title, geoinformation and person, respectively. This section will deal with DiFAB's integration of well-known thesauri and categorization tools and highlight both the advantages and drawbacks of applying research tools developed for a Western European context to a different area and period of cultural history.

### **3.1. Historical geography**

Let us return to our example of the Church of St. Clement in Ohrid (FY-ROM). This is a structure that currently stands in the Old Town of the modern city Ohrid, an administrative seat located on a lake in the southwestern part of today's Former Yugoslav Republic of Macedonia. An inscription reveals that the church – originally dedicated to the Celebrated God-Bearer (Mother of God) – was built in 1294/95 by the ethnically Albanian, Byzantine official Progon Zgur (Greek: Progonos Sgouros) and his wife Eudokia. The medieval fortress town had been erected on the site of an ancient settlement known as Lychnidos (Greek: Λύχνιδος, Latin: *Lychnidus*). A cultural center in the Middle Ages, it was the seat of the Bulgarian Patriarchate in the late 10<sup>th</sup> to early 11<sup>th</sup> c., was contested between Byzantine, Bulgarian, Albanian, and Epirote rulers before becoming part of the Serbian empire in 1334, and eventually fell to the Ottomans in 1394, around which time the relics of St. Clement of Ohrid were transferred there and the church rededicated in his name.<sup>9</sup> The town declined during the Ottoman period, became Bulgarian in 1878, was taken by the Kingdom of Serbia in 1912, was occupied by the Kingdom of Bulgaria during World War I, then became a southern outpost of the Kingdom of Yugoslavia in 1929 and later part of the Socialist Federal Republic of Yugoslavia until 1991 when the Republic of Macedonia declared independence.

When dealing with cultural history, the question of location or place and space in regard to their respective designations can be a tricky subject, but it is necessary to address.<sup>10</sup> As the brief history of Ohrid and the church we are investigating demonstrates, one location can be known by many names both in the past and the present. With a view to bringing together recorded historical place names, modern place names in multiple languages and geographic coordinates that allow locations to be plotted in a GIS, DiFAB has integrated into its metadata structure the Getty Thesaurus of Geographic Names (TGN).<sup>11</sup> In its current state the TGN offers many benefits, though its application also has some limitations especially for the historical period and areas under consideration (see table 1, below).

Research

Research Home • Tools • Thesaurus of Geographic Names • Full Record Display

Getty Thesaurus of Geographic Names® Online  
Full Record Display

« Previous Page Help

Vernacular Display | English Display

Click the icon to view the hierarchy.

Semantic View (JSON, JSONLD, RDF, N3/Turtle, N-Triples)

**ID: 7015329** **Record Type: administrative**

**Ohrid (inhabited place)**

Coordinates:  
 Lat: 41 07 00 N degrees minutes    Lat: 41.1167 decimal degrees  
 Long: 020 48 00 E degrees minutes    Long: 20.8000 decimal degrees

**Notes:** Located on Lake Ohrid; site of Greek Colony Lynchnidos during 3rd cen. BC; seized by Romans ca. 100 & rebuilt after earthquake in 518; was seat of Bulgarian patriarchate in 9th cen.; ruled by Byzantine & Ottoman empire.

**Names:**  
**Ohrid (preferred, C,V)**  
**Ochrid (C,V)**  
**Ochrida (C,V)**  
**Orchid (C,V)**  
**Ochrida (C,V)**  
**Lychnidus (H,O)**  
**Lychnidus (H,O)** ..... ancient

**Hierarchical Position:**  
 World (facet)  
 ..... Europe (continent) (P)  
 ..... The former Yugoslav Republic of Macedonia (nation) (P)  
 ..... Ohrid (municipality) (P)  
 ..... Ohrid (inhabited place) (P)

**Place Types:**  
 inhabited place (**preferred, C**) ..... from 3rd cen. BC  
 town (C)  
 transportation center (C) ..... in 2nd cen. was a post on Via Aegnatia to Bitola & Greece, now has rail, road & air links  
 agricultural center (C)  
 fishing community (C)  
 tourist center (C) ..... is chief resort area of Macedonia for Lake Ohrid & for medieval & Turkish buildings & frescoes  
 archaeological site (C)  
 episcopal see (C) ..... established by 4th cen., with 1st Slav bishop being St. Clement (died 916) who translated Scriptures from Greek into Slavonic  
 cultural center (H) ..... under St. Clement who founded 1st Slav school of higher education & wrote earliest Slav literature  
 capital (H) ..... under Bulgarian tsar in late 10th & early 11th cen.

Fig. 3: Truncated screenshot of the TGN entry for "Ohrid (inhabited place)"; last accessed 2 January 2017, <http://www.getty.edu/vow/TGNFullDisplay?find=&place=&nation=&english=Y&subjectid=7015329>.

A search for "Ohrid" within TGN yields three results: a municipality [ID 7017378]; an inhabited place [ID 7015329]; and a lake [ID 1117570]. Each of these is categorized by its record type, defined by geographic coordinates and enhanced through the addition of variant names, a position in a geographic hierarchy, a list of place types with further historical or current details, and a list of sources and contributors (see fig. 3). One basic issue is the reliability of each entry and in particular its completeness and exactness. Another, especially for application to cultural history initiatives, is the TGN's foundation in contemporary political geography.

| Feature                  | Benefits   | Limitations   |
|--------------------------|--|---|
| ID                       | + Certain identification<br>+ Retrievable<br>+ Compatibility   | - Many locations missing  |
| Coordinates              | + Known location<br>+ Easy to map  | Accuracy and precision of given coordinates:<br>- Pinpoint location, not space<br>- Only for large-scale mapping; alone too imprecise for site or urban mapping   |
| Names                    | + Variant spellings of transliterated place name<br>+ Some historical place names<br>+ Increased findability | - "Vernacular" also transliterated (Cyrillic and Greek alphabets, i.a., are supported but rarely entered)<br>- Many historical place names lacking (esp. medieval period)<br>- Not possible to search in (non-Latin) historical or contemporary local languages |
| Hierarchical position    | + Additional geopolitical designations<br>+ Findable within hierarchy  | - Based on modern political geography   |
| Place types              | (currently not considered by DiFAB since not integrated into Phaidra)  |   |
| Sources and contributors | (currently not considered by DiFAB since not integrated into Phaidra)  |   |

Tab. 1: Analysis of the benefits and limitations of applying TGN to the cultural sphere of the former Byzantine empire.

Other valuable resources have developed out of interest in the historical geography of the Ancient World; two prominent examples are the Digital Atlas for Roman and Medieval Civilizations (DARMC) and Pleiades.<sup>12</sup> These have been combined in Pelagios Commons<sup>13</sup>, a community-supplied Linked Open Data initiative, with a current focus on Greek and Roman antiquity and lay-

ered mapping approaches to this period and region. The methods of these projects are interesting for DiFAB though the content is still lacking for the Byzantine/medieval period and for some areas of eastern Europe, the Caucasus region, the Middle East and North Africa. For this reason, we are looking to contribute content based on our documentation of material sources and possibly integrate their geoinformation into our metadata in the future.

### **3.2. Multilingualism**

The issue of multilingualism arises not only in regard to contemporary and historical geographic names, as we have just seen, but also when it comes to the title or designation given to a certain monument or object. The chosen example of St. Clement in Ohrid is especially complex – though not at all unique – in the fact that the structure has been known by different names throughout its long history. Dedicated originally to the Celebrated God-Bearer (Mother of God), in Greek *θεοτόκος Περιβλέπτος* (Theotokos Peribleptos), and in Slavonic *Пресвета Богородица Перивлепτος* (Presveta Bogorodica Perivleptos), it was rededicated to St. Clement of Ohrid at the end of the 14<sup>th</sup> century and is also known as the Church of St. Clement, in Greek *Άγιος Κλήμης της Αχρίδας* (Hagios Klēmēs tēs Achridas), in Macedonian *Свети Климент Охридски* (Sveti Kliment Ohridski), and in German *Klementskirche*. This structure is not to be confused with two other St. Clement churches located in the same city.

To the best of our knowledge no authoritative list of names or index for Byzantine material culture exists, and the designations used in published literature depend on language, scholarly tradition and convention at the time. The task of a digital research archive, however, is to be as easily accessible as possible for all users and must therefore take into consideration different designations, multiple languages and spelling variants.

Though place names are to a certain extent covered by TGN,<sup>14</sup> it is the policy of DiFAB to supply the main title in German in the format of "location, monument type and name" – for example: "Ohrid, Klementskirche (Sveti Kliment)" – and alternative titles in English, Greek, and the current local language, as well as Turkish or other languages of historical relevance. In the case of a monument with multiple historical names, these alternative titles are also supplied in at least three languages and linked to one particular digital asset. Thankfully templates can be created for metadata entry, since this effort is deemed essential – still lacking some type of index or name authority – in order to ensure better findability, ease of use for scholars, and assistance for students and those unfamiliar with these monuments.

### 3.3. Culturally specific terminologies

For cultural history documentation and research, it helps to provide a digital asset with additional metadata, namely descriptors related to the "what" of the image content. Adding such keywords contributes to a more precise identification of the depicted object and allows for more targeted searches, thereby preparing these digital assets for potential future use in other contexts beside the digital archive itself. Again here thesauri can be helpful tools even though the terminologies they contain have their limitations, as two examples – one architectural term and one iconographic term – will show.

Architectural terminology is the common language that art historians and architects use to describe the forms and elements of a man-made structure, and in an ideal world there would be complete consensus about the terms used and their exact meaning. Unfortunately, the terms found in many thesauri, which have derived from classical archaeology and western European monuments, are not always applicable to Byzantine material culture, or can cause considerable confusion when terms carry a different meaning in a different cultural context. A case in point is the word *naos* (Greek: ναός). While the word in its modern Greek usage denotes the function of a building as a house of worship and can denote an ancient temple, a church, a synagogue, a mosque or a pagoda,<sup>15</sup> for classical archaeologists the term describes a particular part of a specific type of building, namely the walled-off interior chamber of a pagan Greek temple where the statue of the deity would have stood. To a Byzantine art historian, *naos* describes the main central space of a Byzantine church where the congregation stands; separated from the enclosed space with the altar, it is usually domed and decorated to become "the earthly embodiment of the Christian universe."<sup>16</sup> This distinction may not seem important until we turn to a well-known thesaurus, in this case the Getty Art and Architecture Thesaurus (AAT), which provides only the first, classical, definition.<sup>17</sup> To include this reference in the metadata of a Byzantine monument would be akin to adding false information, since the given meaning does not fit the other cultural context.

Not only parts of buildings but also devotional and narrative images found in these buildings and elsewhere are described using shorthand terms for their content; art historians refer to conventional subjects with their characteristic imagery and symbols as iconography. Also here thesauri and categorization schemes have positive and negative aspects regarding their use. Let us take as an example a scene that was painted on the interior west wall of the Church of St. Clement in Ohrid (fig. 4). Known in the context of Byzantine art history and Greek Orthodox religion as *koimēsis* (Greek:

*Κοίμησις Θεοτόκου*), this scene above the entrance to the church depicts the peaceful death, literally "falling asleep" or dormition, of the Mother of God who is surrounded by apostles, bishops and angels and whose soul Christ holds above her body. This painting can be located, roughly dated, and even connected to both donors and artists – a rarity in this area of study. The elements in its composition can be described and linked with thesauri such as AAT or classification systems like Iconclass,<sup>18</sup> yet the term for the whole scene, *koimēsis* [*theotokou*], which the artist also inscribed at the top, cannot. This current lack is surprising considering that the Death of the Virgin was also a well-known theme in western medieval art. The classification "the Dormition" is contained in Iconclass [ID 73E74], and while it supplies references to the respective Biblical passages and common actors, these details and the further sub-classifications betray the different characteristics and traditions of the body of evidence upon which these classifications are based.<sup>19</sup> As with architectural terminology, many contributions from historically less prominent areas of the field still need to be made in order to improve available thesauri and classification systems.



Fig. 4: *Koimēsis* scene on the interior west wall of the Church of St. Clement, Ohrid, painted by Michael Astrapas and Eutychios, around 1294/5 or 1305, collection of Horst Hallensleben, inv. nr. 018-099 (originally published in Djurić, *Vizantijske freske u Jugoslaviji*, Belgrade 1974, color plate XV). Digital image available at <http://phaidra.univie.ac.at/o:526659> (accessed 19.5.2017), [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

## 4. Dating

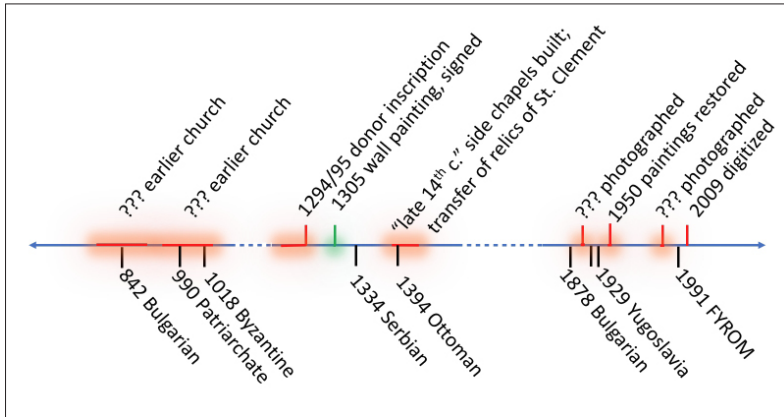


Fig. 5: Timeline for the Church of St. Clement in Ohrid: above, building history and documentation of the structure; below, political history in relation to Ohrid (not to scale). Known absolute dates and unknown precise dates are represented by a short vertical line; uncertain dates and date ranges are shown as a fuzzy horizontal line.

Returning to our "5 Ws", the question of "when" is a key aspect of metadata, and in the case of cultural history archives, dating applies to all layers of information – the original physical object, the image carrier(s) and the digital asset. The creation and manipulation of the digital asset tends to be documented and is easy to express as a specific date in the common format YYYY-MM-DD. Sometimes photographic prints, slides and negatives are dated; sometimes a date range can be determined based on association with a particular individual, technical characteristics of the medium and production, or internal clues derived from the image content. When it comes to the depicted object itself, however, we must consider ourselves lucky to have an absolute or calendar date at all, or just one date at that. Antique and medieval monuments as we have inherited them seldom reflect one period of construction and are instead a composition of multiple building phases, including later additions and subtractions, redecorations and renovations.

A timeline for the Church of St. Clement in Ohrid (fig. 5) attempts to visualize the history of the structure and its documentation (above), juxtaposed with pivotal events in the political history of that location (below). Two general observations can be made. First, few precise, absolute dates directly related to the church can be identified; more often construction



phases and events correspond to a date range that can be more-or-less precise and more-or-less conjectural.<sup>21</sup> Second, for lack of explicit sources, archaeological and art historical research often makes use of relative chronology (establishing a sequence of events) based on the relationships between constituent parts and comparison with other monuments.

The first issue can be resolved by allowing for two metadata entries pertaining to date; the VRA Core integrates as an XML sub-element "earliestDate" and "latestDate". Where beginning and end points are approximate, however, a quantitative value is not ideal, because it gives the initial impression of certainty.<sup>22</sup> Yet the greater challenge proves to be relative chronology, which relies on a web of contingencies and interdependencies, and the ability to capture a complex relationship not readily expressed in Dublin Core. This is an area that requires further development in order to provide comprehensive, object-related metadata that are ready and able to be used for future research applications.

## **5. Closing remarks and future prospects**

It was the goal of this paper to highlight some areas of concern for cultural historians who work with material culture that have until recently received less attention in the development of metadata structures and standards. The conversion of analog documentation of monuments and artifacts poses distinctive challenges and has other requirements than born-digital assets, especially from the point of view of a research archive. Here, we chose to focus on three aspects – "life phases" and different forms of documentation, thesauri and categorization schemes, and different types of dating – to underscore the vital importance of these metadata for research endeavors in cultural and historical sciences.

Ultimately it is the aim of DiFAB, and any digital archive, to compile information related to the digital images that it curates and to prepare these digital assets – both image and metadata – in a way that focuses on accessibility and usability. As the example from the Digital Research Archive for Byzantium – the detailed investigation of the Church of St. Clement and its documentation from an object-based approach – has hopefully shown, continuing work on metadata standards and continuous contributions to Linked Open Data initiatives like the Getty Vocabularies are essential to meet these goals.

Beside compiling data to fill knowledge gaps and to expand existing structures, we are also looking toward future applications and long-life inte-



roperability. For this reason, DiFAB is partnering with the FWF-funded "Byzantine Stone Bridges" project directed by Galina Fingarova<sup>23</sup> to develop a pilot project based on material culture of the Byzantine cultural sphere. This collaboration will work on developing a method for the creation of a GIS-based interface and research tool, and will entail a critical analysis of the metadata structure currently in place. From this collaboration we hope to gain a detailed use analysis and a proof of concept for our metadata, as well as suggestions for future best practice models in order to ensure that the efforts invested now bear fruit for upcoming research and teaching applications.

Nina Rannharter, BA  
Universität Wien, Institut für Kunstgeschichte  
E-Mail: [nina.rannharter@univie.ac.at](mailto:nina.rannharter@univie.ac.at)

Sarah Teetor, Bakk. BA MA  
Universität Wien, Institut für Kunstgeschichte  
E-Mail: [sarah.teetor@univie.ac.at](mailto:sarah.teetor@univie.ac.at)

## Notes

- 1 Information on this work package of the e-Infrastructures Austria project can be found at <http://e-infrastructures.at/das-projekt/work-package-cluster/cluster-i/>.
- 2 For further information, see <http://difab.univie.ac.at> and for access to the database, visit <https://phaidra.univie.ac.at>.
- 3 An image database jointly initiated by the Faculty of Historical and Cultural Studies and the Faculty of Philological and Cultural Studies, both of the University of Vienna; accessible at <http://unidam.univie.ac.at>.
- 4 PHAIDRA is a digital asset management system geared toward long-term digital archiving that is based on the open source software Fedora and was developed by the IT Center of the University of Vienna in collaboration with the Vienna University Library. For further information (in German) see <http://phaidraservice.univie.ac.at> and for access to the database <https://phaidra.univie.ac.at>.
- 5 <http://artstor.org>.
- 6 Used here with a touch of irony, the "5 Ws" are a standard series of questions used to gather information that is fundamentally necessary to explain an event or to solve a problem. Rooted in the rhetorical tradition of classical antiquity and its somewhat longer list of circumstances,

these five question words plus "how" became staples of early scientific research and in the late 19<sup>th</sup> c. of journalistic investigation. Although "why" seems not to fit as well as a category for metadata, it can be understood as a broader contextualization that *is* directly relevant to the image content and/or the image carrier, and therefore valid metadata.

- 7 This is a well-known caveat for working with metadata schemes based on Dublin Core. The VRA Core of the Visual Resources Association and the Network Development and MARC Standards Office of the Library of Congress makes this distinction by adding in the XML schema an upper-level wrapper denoting either a "work" (a physical object at the level of image content), "collection" (a group of "works") or "image" (a visual surrogate of a physical object, itself either a physical image carrier or a digital image file); see: <http://www.loc.gov/standards/vracore/> and specifically [http://www.loc.gov/standards/vracore/VRA\\_Core4\\_Element\\_Description.pdf](http://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf), p. 2. This places metadata in a relationship with a certain level – either depicted image content or its reproduction – though the distinction between physical image carrier and digital asset seems not to be clearly defined.
- 8 The issue of "who", or participating persons, as related to image content, will not be discussed here, since the main issues of coverage in thesauri and dating are mentioned elsewhere. For a comparison, see the article by Sandor Kopacsi in this issue.
- 9 For general information on the historical and art historical significance of the Church of St. Clement, Ohrid, see: Slobodan Ćurčić, *Architecture in the Balkans: From Diocletian to Süleyman the Magnificent*. New Haven, CT: Yale University Press, 2010, p. 571–577, including references in notes, p. 871–872; and "Ohrid" in: Alexander Kazhdan (ed.), *Oxford Dictionary of Byzantium, Vol. 3*. Oxford: Oxford University Press, 1991, p. 1514–1515, especially for primary sources and older literature.
- 10 Here we are using the terms "location", "place" and "space" in their simplified, common meanings based on traditional geography and cartography in order to denote a fixed locality with measurable geographic coordinates (location and place) or a geographic area that can also be quantified and represented on a map (space). In this article the meaning of these words should not be confused with their use in more recent cognitive approaches in cartography or the use of the terms "place" and "space" in other areas of study such as anthropology, whereby – again overly simplified – "place" designates a perceived or experienced abstract space that holds meaning, and "space" refers to an impersonal natural location.

- 11 This is one of several openly accessible and Linked Open Data Getty Vocabularies developed by The Getty Research Institute; available at <http://www.getty.edu/research/tools/vocabularies/tgn/>.
- 12 DARMC, first based on the Barrington Atlas of the Greek and Roman World (2000), is available at <http://darmc.harvard.edu>; it has expanded into the medieval period though the data presented is richer for western and central Europe. Pleiades is a community-built, map-based index of the Ancient World based on place names found in primary and secondary sources; it also seeks to expand into other regions and periods and is available at <https://pleiades.stoa.org>.
- 13 For more information, visit <http://commons.pelagios.org>.
- 14 The two other churches are the newly erected Church of Sts. Clement and Panteleimon (Црква Св. Климент и Пантелејмон), which stands on the foundations of a 9<sup>th</sup> century building, and the single-nave church "Small" St. Clement of Ohrid (Црква Мал Свети Климент Охридски), both also in the Old Town of Ohrid.
- 15 Locations that do not have an ID in TGN are added to the metadata manually – both the place name and the coordinates; a list of "missing" places and alternative names is regularly contributed to Getty Vocabularies.
- 16 "ναός" in: *The Modern Greek Dictionary / Λεξικό της κοινής νεοελληνικής* (1998), available through the website of the project "Portal for the Greek language and language education", co-financed by the European Commission: [http://www.greek-language.gr/greekLang/modern\\_greek/tools/lexica/triantafyllides/](http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/).
- 17 "Naos" in: Alexander Kazhdan (ed.), *Oxford Dictionary of Byzantium*, Vol. 2. Oxford: Oxford University Press, 1991, p. 1436.
- 18 Since the presentation of this paper in June 2016, the AAT has updated its entry for "naoi" / "naos" [ID: 300004658] to include a note on its usage in Byzantine architectural history; last accessed on 4 January 2017: [http://www.getty.edu/vow/AATFullDisplay?find=naos&logic=AND&note=&english=N&prev\\_page=1&subjectid=300004658](http://www.getty.edu/vow/AATFullDisplay?find=naos&logic=AND&note=&english=N&prev_page=1&subjectid=300004658).
- 19 Iconclass is a classification system for art and iconography that was first developed by the Dutch art historian Henri van de Waal and is currently managed by the Rijksbureau voor Kunsthistorische Documentatie / Netherlands Institute for Art History; for more information, see: <http://www.iconclass.nl>.
- 20 Byzantine iconography is almost notoriously consistent, and although some variation can be seen across time, space, and medium, the strongly symmetrical scene comprises certain core elements: the Moth-

er of God in the foreground reclining on a bed; generally six apostles standing at the head of the bed and six at the foot; Christ in a mandorla holding up in his hands her soul, wrapped as an infant; a pair or pairs of angels at his sides; and an inscription identifying the scene. Leaving aside differences of dogma, the Catholic tradition of this scene became conflated with the Assumption and also the (heavenly) Coronation of the Virgin, which was not known in medieval Orthodox art. An example of the Dormition and Assumption in one frame can be seen on the eastern side of the tabernacle in the Gothic church of Orsanmichele in Florence, which was completed by Orcagna (Andrea di Cione) in 1359. For an image and more information, see: Nancy Rash Fabbri and Nina Rutenberg "The Tabernacle of Orsanmichele in Context," *The Art Bulletin*, Vol. 63, No. 3 (Sept. 1981), p. 385–405, esp. 397 (image), 399–400; available at <http://www.jstor.org/stable/3050142>. In the western European context, the Assumption eventually gained predominance over the depiction of the death of the Virgin.

- 21 Overall it is quite fortuitous to find an absolute date on the monument itself, such as a painted or incised inscription, or a reference in an historical text that can be attributed to a monument with a seeming degree of certainty. Types of evidence that result in a date range are technical and stylistic analysis of the material of the monument itself and its forms – depending on the specific characteristics, this type of dating can be quite precise (within a few years) or very broad (within a couple centuries) – and secondary finds such as coins, which yield a *terminus post quem* or a *terminus ante quem*, as well as early research and historic photographic documentation which provide clues to more recent changes.
- 22 At the same time, qualifiers such as "about", "approximately" or "circa" – which are so prevalent in secondary literature concerned with objects of material cultural history – are also problematic due to the quantitative values that underlie them; these are less transparent and nevertheless affect search results and usability.
- 23 "Byzantine Stone Bridges: The material evidence and their cultural meanings" V 546 Richter Program, funded by the Austrian Science Fund (FWF) from August 2016 to July 2019; the project abstract can be found using the Project Finder feature on the website of the Austrian Science Fund: <https://pf.fwf.ac.at/de/wissenschaft-konkret/project-finder>. The openly accessible web-portal for the Byzantine Stone Bridges project is currently under construction.

# ■ IMPLEMENTATION OF A CLASSIFICATION SERVER TO SUPPORT METADATA ORGANIZATION FOR LONG TERM PRESERVATION SYSTEMS

by *Sándor Kopácsi*<sup>1</sup>, *Rastislav Hudak*, *Raman Ganguly*

**Abstract:** *In this paper we describe the implementation of a classification server that helps in metadata organization for a long term reservation system of digital objects. After a short introduction to classifications and knowledge organization, the requirements of the system to be implemented are summarized. Some Simple Knowledge Organization System (SKOS) management tools we have evaluated are briefly presented. These include Skosmos, the solution we have selected for implementation. Skosmos is an open source, web-based SKOS browser based on the Jena Fuseki SPARQL server. We present the main steps of the installation of the applied tools and some potential problems with the classifications used, as well as possible solutions.*

**Keywords:** *long term preservation; metadata; classification; SKOS; Skosmos; Jena Fuseki*

## IMPLEMENTIERUNG EINES KLASSIFIKATIONSSERVERS FÜR METADATENORGANISATION FÜR LANGZEITARCHIVIERUNGSSYSTEME

**Zusammenfassung:** *In diesem Artikel beschreiben wir die Implementierung eines Klassifikationsservers für Metadatenorganisation in einem Langzeitarchivierungssystem für digitale Objekte. Nach einer kurzen Einführung in Klassifikationen und Wissensorganisationen stellen wir die Anforderungen an das zu implementierende System vor. Wir beschreiben sämtliche Simple Knowledge Organization System (SKOS) Management Tools, die wir untersucht haben, darunter auch Skosmos, die Lösung, die wir für die Implementierung gewählt haben. Skosmos ist ein open source, webbasierter SKOS Browser, basierend auf dem Jena Fuseki SPARQL Server. Wir diskutieren einige entscheidende Schritte während der Installation der ausgewählten Tools und präsentieren sowohl die potentiell auftretenden Probleme mit den verwendeten Klassifikationen als auch mögliche Lösungen.*

**Schlüsselwörter:** *Langzeitarchivierung; Metadaten; Klassifikation; SKOS; Skosmos; Jena Fuseki*



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## **Contents**

1. *Introduction*
2. *Classification and knowledge organization systems*
3. *Application of the classification server*
4. *Requirements*
5. *Tool selection for implementation*
6. *Implementation*
7. *Available services and usage*
8. *Connection to the preservation system*

## **1. Introduction**

Long term preservation of digital objects is a key issue for libraries and research institutes today, because they need to ensure that the digital content of books, documents, pictures, research data, etc. remains accessible and usable within a required period of time [1]. Digital preservation includes the activities of planning, resource allocation, and application of preservation methods and technologies [2].

When we store digital objects in an archiving system, it is fundamental to assign well-defined metadata to make the discoverability of the object easier. Metadata can provide title, authors, keywords, and other important information about a document. Metadata can also store technical details on format and structure, ownership and access rights information, as well as the history of preservation activities on the digital object.

When the data provider of the digital object is allowed to add non-standardized values as metadata, it can be challenging to tell the appropriate keywords, a process that sometimes requires guessing. If we want to avoid ambiguities, misspellings, etc. it is better to select terms from pre-defined controlled vocabularies.

Controlled vocabularies, or rather classifications in multiple topics, are available in several data sources among which we can select. If we want to provide all relevant classifications inside our archiving system and make them available to users so they can select terms during upload or search, a classification server that handles vocabularies and classifications relevant to us seems to be a favourable solution. It is beneficial because we can then ensure that the required classifications are always available within a defined access time in our Classification server.

Classification servers store information according to classification or knowledge organization schemas, usually in the structure of Resource Description Framework (RDF) and/or as Simple Knowledge Organization System (SKOS), and should be organized as Linked Data.

## 2. Classification and knowledge organization systems

### 2.1. Classification

A classification is a form of categorization that collects objects or items according to their subjects usually arranged in a hierarchical tree structure. This knowledge organization technique can take many different forms, such as controlled vocabularies, taxonomies, thesauri, ontologies, and some others (see Fig. 1). There are different interpretations (e.g. [3], [4], [5]) of these types of classifications, from which we can establish the following explanations.

A **controlled vocabulary** is a closed list of words or terms that have been included explicitly, and that can be used for classification. It is controlled because only terms from the list may be used, and because there is control over who can add terms to the list, when and how.

A **taxonomy** is a collection of controlled vocabulary terms organized into a hierarchical structure by applying parent-child (broader/narrower) relationships. Each term in taxonomies is in one or more relationships (e.g. whole/part, type/instance) to other terms in the taxonomy.

A **thesaurus** is a more structured, much richer taxonomy, that uses associative relationships (like "related term") in addition to parent-child relationships.

An **ontology** is a more complex type of thesaurus usually expressed in an ontology representation language that consists of a set of types, properties and relationship types. In an ontology, there are various customized relationship pairs that contain specific meaning, such as "owns" and its reciprocal "is owned by", instead of simple "related term" relationships.

Fig. 1 shows the above mentioned concepts in order of complexity. Controlled Vocabulary is the least complex classification; all other categories are a subset of it. For example Taxonomy is a subset of Controlled Vocabulary with the additional requirement of an hierarchical structure. We can say that every Taxonomy is a Controlled Vocabulary, but not every Controlled Vocabulary is a Taxonomy, and so forth.

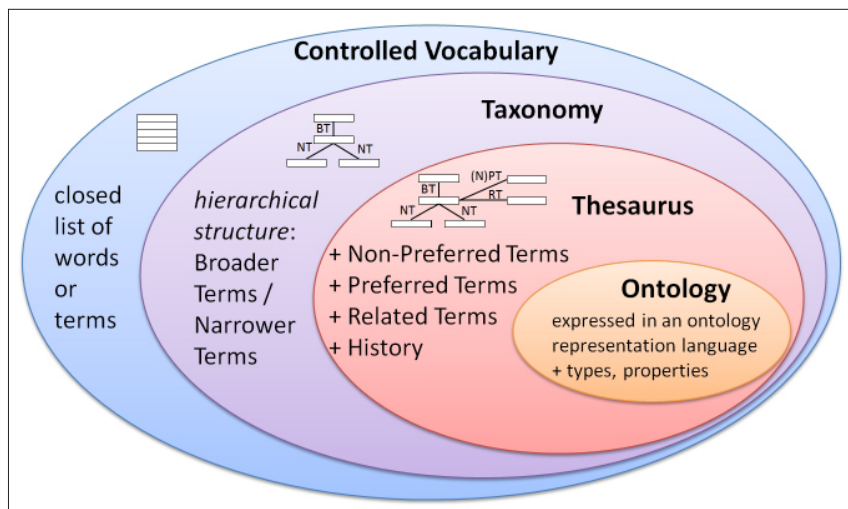


Fig. 1: Categories of classification

## 2.2. Knowledge organization systems and Linked Data

Classifications can be considered as a collection of organized knowledge, therefore the technical background of classification is based on Knowledge Organization Systems (KOS). In knowledge organization systems we usually store knowledge in form of triplets, such as object-predicate-subject, or object-attribute-value.

Classifications can be represented in Simple Knowledge Organization Systems (SKOS<sup>1</sup>) as a Resource Description Framework (RDF) vocabulary. The Simple Knowledge Organization System is a W3C recommendation designed for the representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured and controlled vocabulary.

Using RDF allows knowledge organization systems to be used in distributed, decentralized metadata applications. "Decentralized metadata is becoming a typical scenario, where service providers want to add value to metadata harvested from multiple sources." [6]

Each SKOS concept is defined as an RDF resource, and each concept can have RDF properties attached, which include one or more preferred terms, alternative terms or synonyms, and language specific definitions and notes. Established semantic relationships are expressed in SKOS and intended to emphasize concepts rather than terms/labels. [7]



A special query language, called SPARQL (a recursive acronym for SPARQL Protocol and RDF Query Language)<sup>2</sup>, can be used to query and update data sources stored as RDF. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. [8]

SKOS – as a modern, well established standard – can (potentially) support formal alignments and hierarchical grouping of concepts using different SKOS relations (e.g. `skos:exactMatch`, `skos:closeMatch`, `skos:narrower`, `skos:broader`, `skos:related`), translation of concept labels, and URI-based mapping to similar concepts in other KOS.

### **3. Application of the classification server**

The Classification server that we have integrated from available tools is an independent component of Phaidra<sup>3</sup>, the Digital Asset Management Platform with long-term archiving functionality developed by the University of Vienna.

Phaidra is an acronym for Permanent Hosting, Archiving and Indexing of Digital Resources and Assets. Phaidra is implemented at several local Austrian institutions and also internationally, including universities in Serbia, Montenegro and Italy. Phaidra provides academic, research and management staff the possibility to archive digital objects for an unlimited period of time, to permanently secure them, to supplement them with metadata, as well as to archive objects - and to provide world-wide access to them.

We are going to apply the Classification server both during the ingestion and in the search phase of Phaidra usage. During the ingestion phase the Phaidra user uploads new items to the archiving system and can assign metadata to them from controlled vocabularies. In the search phase the user looks for objects that might have been previously supplied with classification terms from existing vocabularies. We also need the Classification server for resolving the classification terms that were saved together with the objects, when displaying them.

### **4. Requirements of the classification server**

The first step in development was to establish requirements of the Classification server that supports different classifications and controlled voca-

ularies for Phaidra. The requirements were grouped into the categories of General Requirements and Technical Requirements. At this level of development we have not explicitly distinguished functional and non-functional requirements, but there are constraints that are either more functional or more of an administrative feature of the system among the General Requirements and Technical Requirements. Each requirement was prioritized between 1 and 3, where 1 meant the highest priority (= most important), while 3 meant the lowest priority (= least important).

#### 4.1. General requirements

The General Requirements (see Table 1) are related to the main goals of the system that we were going to achieve by the implementation of the Classification server. Some of them (GR-1, GR-2, GR-3, GR-4) are functional requirements, but others (GR-5 and GR-7) are rather administrative issues.

|      | Requirement   | Priority<br>1: highest<br>3: lowest |
|------|---|-------------------------------------|
|      | The Classification server   |                                     |
| GR-1 | should resolve the URIs of the different terms.   | 3                                   |
| GR-2 | should support multiple languages .   | 2                                   |
| GR-3 | should support multiple versions of classifications.  | 1                                   |
| GR-4 | should return the list of sub terms (narrower concepts).  | 1                                   |
| GR-5 | should be Phaidra independent.  | 1                                   |
| GR-6 | should have no assumptions about content, which means that the set of classifications can differ on instances that are locally managed. | 2                                   |
| GR-7 | does not require too much development effort and have low costs.  | 2                                   |

Tab. 1: General requirements

## 4.2. Technical requirements

All the Technical Requirements (see Table 2) can be considered functional requirements, and some of them (TR-1, TR-2, TR-3) are related to the input and output format of the system. TR-4 was a rather important requirement, because we definitely wanted our system to provide a SPARQL endpoint through which other systems can access our Classification server by using SPARQL queries<sup>4</sup>.

|      | Requirement  | Priority<br>1: highest<br>3: lowest |
|------|--|-------------------------------------|
|      | The Classification server  |                                     |
| TR-1 | should return the terms in multiple formats (such as XML, JSON, RDF, TTL).   | 2                                   |
| TR-2 | should support standard import formats for vocabularies (e.g. SKOS/RDF, TTL, N-Triples).                                       | 1                                   |
| TR-3 | should support Linked Data (in SKOS/RDF/XML formats).  | 1                                   |
| TR-4 | should provide a SPARQL endpoint.  | 1                                   |
| TR-5 | should provide a comprehensive search needed for Phaidra.  | 1                                   |
| TR-6 | should also support classifications/vocabularies that do not yet support linked data (do not have URIs).                       | 2                                   |
| TR-7 | should be able to use external terminology services, e.g. dewey.info, so that we do not necessarily have to import it locally. | 3                                   |

Tab. 2: Technical requirements

## 5. Tool selection for the implementation

### 5.1. Evaluation of available tools

Before implementation we tested and evaluated some popular tools (ThManager<sup>5</sup>, TemaTres<sup>6</sup>, SKOS Shuttle<sup>7</sup>, PoolParty<sup>8</sup>, Protégé<sup>9</sup>, Skosmos<sup>10</sup> for managing classifications. We also collected information about other tools (like HIVE<sup>11</sup>, iQvoc<sup>12</sup>, CATCH<sup>13</sup>), but they did not fit our requirements, thus they were not investigated further.

All of the evaluated tools had advantages and disadvantages, but the most important selection criterion for us was to find an open source tool that can provide a SPARQL Endpoint. A comparison of the evaluated tools can be seen in Table 3 where we included the open source products only. Another important selection criterion was to find a tool that is based on the stable and widespread Apache Jena technology and which can be accessed via REST API<sup>14</sup>.

|           | Implemented in | Input                | Multi-lingual | Backend | SPARQL Endpoint | REST API  | Last update |
|-----------|----------------|----------------------|---------------|---------|-----------------|-----------|-------------|
| ThManager | Apache Jena    | SKOS RDF             | yes           | SPARQL  | available       | N/A       | 2006        |
| TemaTres  | N/A            | SKOS, tabulated text | yes           | MySQL   | available       | available | 2017        |
| Protégé   | Java Swing     | OWL, Excel, CSV      | yes           | SPARQL  | available       | N/A       | 2016        |
| Skosmos   | Apache Jena    | SKOS Core            | yes           | SPARQL  | available       | available | 2017        |

Tab. 3: Comparison of the evaluated tools

Based on the above criteria, Skosmos with Jena Fuseki seemed to be the best solution, which is why we selected it for our Classification server implementation.

## 5.2. Skosmos with Jena Fuseki

Skosmos<sup>15</sup>, developed by the National Library of Finland, is an open source web application for browsing controlled vocabularies. Skosmos was built on the basis of prior development (ONKI<sup>16</sup>, ONKI Light<sup>17</sup>) for developing vocabulary publishing tools in the FinnONTO (2003–2012) research initiative from the Semantic Computing Research Group.

Skosmos is a web-based tool for accessing controlled vocabularies used by indexers describing documents, and by users searching for suitable keywords. Vocabularies are accessed via SPARQL endpoints containing SKOS vocabularies.

Skosmos provides a multilingual user interface for browsing vocabularies. The languages currently supported in the user interface are English, Finnish, German, Norwegian, and Swedish. However, vocabularies in any language can be searched, browsed and visualized, as long as proper language tags for labels and documentation properties have been provided in the data.

Skosmos provides an easy to use REST API for read only access to the vocabulary data. The return format is mostly JSON-LD, but some methods return RDF/XML, Turtle, RDF/JSON with the appropriate MIME type. These methods can be used to publish the vocabulary data as Linked Data. The API can also be used to integrate vocabularies into third party software. For example, the search method can be used to provide autocomplete support and the lookup method can be used to convert term references to concept URIs. [9]

The developers of Skosmos recommend using the Jena Fuseki<sup>18</sup> SPARQL server and triple store with the Jena text index for large vocabularies. The Jena text extension can be used for faster text search. In addition to using a text index, caching of requests to the SPARQL endpoint with a standard HTTP proxy cache such as Varnish<sup>19</sup> can be used to achieve better performance for repeated queries, such as those used to generate index view.

## 6. Implementation of the Classification server

The classification server was implemented using Skosmos as a frontend for handling SKOS vocabularies, and Jena Fuseki as a SPARQL RDF store containing SKOS vocabulary data (see Fig. 2). The input of the system and the possible connections to the users or directly to Phaidra are discussed in the sections below.

Alternatively, instead of Fuseki, we could use other SPARQL 1.1 compliant RDF stores, but the performance of other tools did not seem to be sufficient with large vocabularies since there is no text index support for generic SPARQL 1.1.

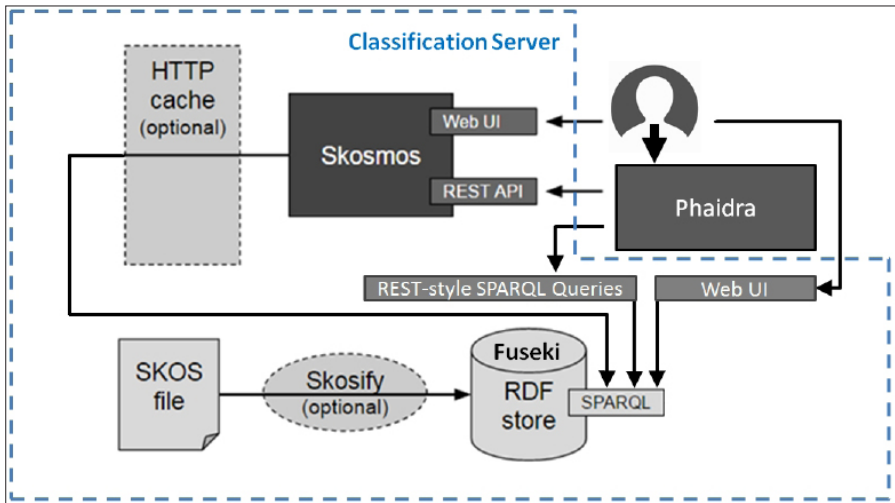


Fig. 2: System architecture (original source: [10])

### 6.1. Installation of Skosmos and Jena Fuseki

Skosmos and Fuseki require Apache and PHP running on the server. We have installed them on a Windows 7 environment (Professional 64 bit, Service Pack 1, Intel Core i7-56000 CPU, 2.6 GHz, 16 GB RAM) using Java 1.8 (jre1.8.0\_40), with XAMPP (xampp-win32-1-8-3-4-VC11), as well as on a CENTOS 6.5 and on a Ubuntu 16.04 virtual machine (Intel Xeon CPU E5-2670 0 @ 2.60GHz).

A detailed installation guide can be found on GitHub<sup>20</sup> for the Linux version, but there are some deviations on the Windows version, as well as some important issues that are worth highlighting.

### 6.2. Configuration of Skosmos and Fuseki

**Configuration of Skosmos.** Skosmos can basically be configured in two files, config.inc for setting some general parameters, and vocabularies.ttl to configure the vocabularies shown in Skosmos.

In `config.inc` one can set the name of the vocabularies file, change the timeout settings, set interface languages, set the default SPARQL endpoint, and set the SPARQL dialect if the Jena text index is needed.

Vocabularies are managed in the RDF store accessed by Skosmos via SPARQL. The available vocabularies are configured in the `vocabularies.ttl` file that is an RDF file in Turtle syntax.

Each vocabulary is expressed as a `skosmos:Vocabulary` instance (subclass of `void:Dataset`). The local name of the instance determines the vocabulary identifier used within Skosmos (e.g. as part of URLs). The vocabulary instance has the following properties: title of vocabulary (in different languages), the URI namespace for vocabulary objects, language(s) and the default language that the vocabulary supports, the URI of the SPARQL endpoint containing the vocabulary, and the name of the graph within the SPARQL endpoint containing the data of the individual vocabulary.

In addition to vocabularies, the `vocabularies.ttl` file also contains a classification for the vocabularies expressed as SKOS. The categorization is used to group the vocabularies shown in the front page of Skosmos. You can also set the content of the About page in `about.inc`, and add additional boxes to the left and to the right of the front page in `left.inc` and `right.inc`.

**Configuration of Fuseki.** Fuseki stores data in files. It is also possible to configure Fuseki for in-memory use only, but with a large dataset, this requires a lot of memory. The in-memory use of Fuseki is usually faster.

The Jena text enabled configuration file specifies the directories where Fuseki stores its data. The default locations are `/tmp/tdb` and `/tmp/lucene`. To flush the data from Fuseki, simply clear or remove these directories.

The Jena text extension can be used for faster text search, and Skosmos needs to have a text index to work with vocabularies of medium to large size. The limit is a few thousand concepts, depending on the performance of the SPARQL endpoint and on how much latency is acceptable for users.

If Fuseki is started in the TDB with `./fuseki-server --config config.ttl` it runs using text indexing. To use Fuseki in TDB, the TDB location for Jena text index should be set, and the Lucene text directory in `config.ttl`. If Fuseki is run in memory with `./fuseki-server --update --mem /ds`, then there is no text indexing by default.

It is also possible to use in-memory TDB and text indexing, but it requires a Fuseki configuration file (`config.ttl`) with special "file names" that are actually in-memory (for TDB: `tdb:location "--mem--"`; and for Jena text: `text:directory "mem"`).

**Timeout settings.** If there is more data than Skosmos is able to handle, some queries can take a very long time. The slow queries are usually the statistical queries (number of concepts per type, number of labels per language) as well as the alphabetical index.

Short execution timeout for PHP scripts can trigger Runtime IO Exceptions. To change the timeout values, check PHP and Apache's time out settings (e.g. in `php.ini` the `max_execution_time`). It is highly recommended to find this setting and change it to a higher value (say to 5 or 10 minutes).

Skosmos also has a `HTTP_TIMEOUT` setting in `config.inc`, that should only be used for external URI requests, not for regular SPARQL queries, because there might be unknown side-effects. The EasyRdf HTTP client has a default timeout of 10 seconds. It is also recommended to change this value.

It is also recommended that users change the timeout value of their browsers from which they are planning to access Skosmos on the client-side. This is possible in Firefox and Internet Explorer, but not in Google Chrome.

### **6.3. Getting and setting vocabularies**

The basic usage of our Classification server is to store the classifications locally (if its access time is acceptable), and we also provide the links to the remote SPARQL endpoints of the classifications if they are available.

If certain vocabularies are planned for local use, they have to be in SKOS format, and should be uploaded to the local SPARQL server, that is to Jena Fuseki.

**Downloading and converting vocabularies.** Vocabularies can be downloaded from the original dataset provider (e.g. from Getty, COAR, Statistics Austria, etc.), or in case of a small dataset, they can be created manually. The vocabularies need to be expressed using SKOS Core representation in order to publish them via Skosmos directly, but SKOS-XL representations or even files in Excel can also be easily converted to SKOS Core. For the SKOS-XL to SKOS Core conversion we use the owlart converter<sup>21</sup>. SKOS-XL labels can be converted to SKOS Core labels by executing SPARQL Update queries as well. If the classification is available in Excel or CVS, then VBA macros can convert it to SKOS Core structures. The format of the file that is accepted by Fuseki can be `rdf/xml` (`.rdf` or `.xml`), `turtle` (`.ttl`) or `N-Triples` (`.nt`). For SKOS files from external resources or files converted from other formats it is recommended to pre-process the vocabularies using a SKOS proofing tool, like Skosify<sup>22</sup>. This ensures, e.g., that the broader/narrower relations work in both directions, and that related relationships are sym-



metric. Skosify reports and tries to correct a lot of potential problems in SKOS vocabularies. It can also be used to convert non-SKOS RDF data into SKOS. An online version of the Skosify tool is also available, where the default options can be used after selecting the vocabulary to be checked.

**Uploading files to Fuseki.** If Skosmos is used for accessing classifications in the local SPARQL triple store, then the datasets have to be uploaded to Fuseki. First, it has to be considered if Fuseki will run either in memory or in a predefined folder, usually called TDB. If Fuseki runs in memory, then all uploads and updates (if allowed) will be temporary. If Fuseki runs in TDB, then uploads and updates will remain there even if we exit Fuseki and restart it.

In a SPARQL triple store there is always a default (unnamed) graph, and there can also be multiple named graphs. In other words, there is only one default graph (with no name), but there can be any number of named graphs in a SPARQL endpoint/dataset. The URI namespaces can be used as graph names (e.g. <http://vocab.getty.edu/tgn/> would store Getty's TGN data).

The datasets can be uploaded to Fuseki online, when Fuseki is running, or offline, when Fuseki is not running. To upload the dataset online the control panel of the web interface of Fuseki or command line instructions can be used. For offline upload the datasets can be directly loaded to TDB.

On uploading datasets online to Fuseki through its control panel, one can set the Graph to "default" or to a graph name provided. If a graph name is used, it should be the name of its dataset in `skosmos:sparqlGraph` (e.g. <http://vocab.getty.edu/tgn/>) in `vocabularies.ttl`.

The Fuseki file upload handling is not very good at processing large files. It loads the dataset into memory first, and to the on-disk TDB database (and also the Lucene/Jena text index) only afterwards. It can run out of memory on the first step ("OutOfMemoryError: java heap space" is a typical error message when this happens). If we give several GB of memory to Fuseki (for example by setting JVM heap to 8 GB: `export JVM_ARGS=-Xmx8000M`) it should be possible to upload large (several hundreds of MB) files, although it might take a while and it is recommended to restart Fuseki afterwards to free some memory.

#### ***6.4. Some examples and problems of adding individual vocabularies***

In this section we are going to describe some examples for individual vocabularies that we are using in our classification server that show typical problems and solutions.

Getty vocabularies<sup>23</sup> contain structured terminology for art and other cultural, archival and bibliographic materials. They provide authoritative information for cataloguers and researchers, and can be used to enhance access to databases and web sites.

Getty has its own SPARQL endpoint, but it is not responding correctly to our Classification server. There seems to be some incompatibility between Skosmos (in practice, the EasyRdf library which is used to perform SPARQL queries) and the Getty SPARQL endpoint.

Even if we could access the Getty SPARQL endpoint, it would most likely be extremely slow to use with Skosmos, since it does not have a text index that Skosmos could use. The lack of text index prevents any actual use of Skosmos with the Getty endpoint.

Therefore, we tried to upload Getty vocabularies to our own local Fuseki SPARQL endpoint with the Jena text index. But unfortunately Getty vocabularies do not work well in Skosmos due to their very large size.

There are two sets of each Getty vocabulary, the „explicit“ set and the „full“ set (Total Exports). With the „explicit“ set, which is smaller, we had to configure Fuseki to use inference so that the data store can infer the missing triples. With the full set this is not needed, but the data set is much larger so we had difficulties loading it. We could finally upload the full set of Getty's vocabularies using the tdbloader utility of Jena Fuseki.

The downloaded export file of the full set includes all statements (explicit and inferred) of all independent entities. It is a concatenation of the Per-Entity Exports in N-Triples format. Because it includes all required Inference, it can be loaded to any repository (even one without RDFS reasoning).

We had to download the External Ontologies (SKOS, SKOS-XL, ISO 25964), from [http://vocab.getty.edu/doc/#External\\_Ontologies](http://vocab.getty.edu/doc/#External_Ontologies) to get descriptions of properties, associative relations, etc. We downloaded the GVP Ontology from <http://vocab.getty.edu/ontology.rdf>. Finally we loaded the full.zip export files (.aat, .tgn and .ulan) from <http://vocab.getty.edu/dataset/>. This way we were able to make some Getty vocabularies available in our Classification server, but due to their huge size, they are rather slow.

COAR Resource Type Vocabulary<sup>24</sup> defines concepts to identify the genre of a resource. Such resources, like publications, research data, audio and video objects, are typically deposited in institutional and subject repositories or published in journals.

The main problem with COAR is that it only represents labels using SKOS XL properties. Skosmos currently does not support SKOS-XL. Unfortunately, the remote endpoint of COAR<sup>25</sup> cannot be used either, because the COAR endpoint data currently is not in SKOS Core, but in SKOS-XL. Since we wanted to use COAR data in our Classification server, we had to convert it to SKOS Core labels using owlart<sup>26</sup>.

ÖFOS<sup>27</sup> is the Austrian version of the Field of Science and Technology Classification (FOS 2007<sup>28</sup>), maintained by Statistik Austria. The Austrian classification scheme for branches of science (1-character and 2-character) is a further development modified for Austrian data.

The ÖFOS can be downloaded in PDF and CSV format, but no SKOS structure (in RDF/XML, Turtle or N-Triples) or Linked Open Data representation through a SPARQL Endpoint is available.

Since we received it directly from Statistik Austria<sup>29</sup> in Excel format, the simplest way of converting it to SKOS was using VBA macros. These macros simply read the content of the Excel file, extend it with the appropriate RDF and SKOS labels, and write it to the desired RDF/XML or turtle format.

## 7. Available services and usage of the classification server

Currently our Classification server makes four general on-line classifications from external triple stores (AGROVOC, Eurovoc, STW, UNESCO), some other general local classifications (e.g. COAR Resource Type Vocabulary, GND, ÖFOS, etc.) and two local, Phaidra specific classifications available. The local classifications have been uploaded to our local triple store in order to make them accessible from Skosmos.

The Classification server is available at <http://vocab.phaidra.org> from any web browser. The operation of the server is quite simple: on the opening page of Skosmos (see Fig. 3) the user simply has to click on one of the classifications, and then the selected classification will be opened. First it shows the vocabulary information, like Title, Creator, Date Issued, Rights, etc. Alphabetical and hierarchical view is available for browsing the classifications. Depending on the configuration in some cases the Change History of the vocabulary or the Group of Concepts can also be seen. The user can search for specific contents either in the remote or local triple store servers or simply in a selected classification.

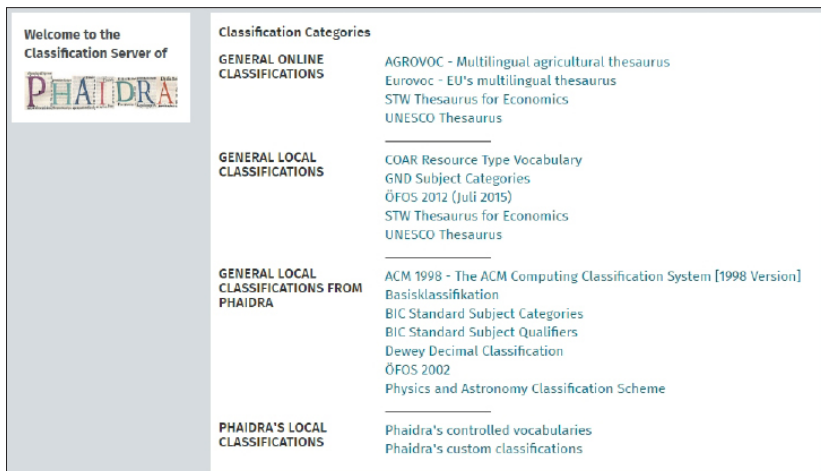


Fig. 3: Opening page of the classification server

## 8. Connecting the preservation system to the classification server

Our Preservation System Phaidra requires the Classification server when the user ingests new items and wants to add metadata from a controlled vocabulary. Another scenario is when the user searches for some documents classified with terms from a controlled vocabulary and wants to display or resolve them.

The connection between Phaidra and the Classification server was realized using the REST API of Skosmos and/or the REST-style SPARQL Queries of Jena Fuseki (see Fig. 1). These are read-only interfaces over HTTP to the data stored in the Classification server, where requests can be built in the URL. The returned data is in UTF-8 encoded JSON-LD format.

Skosmos provides a REST-style API and Linked Data access to the underlying vocabulary data. The REST URLs must begin with the `/rest/v1` prefix. Most of the methods return the data as UTF-8 encoded JSON-LD, served using the `application/json` MIME type. The data consists of a single JSON object which includes JSON-LD context information (in the `@context` field) and one or more fields which contain the actual data.

Jena Fuseki provides REST-style SPARQL HTTP Update, SPARQL Query, and SPARQL Update using the SPARQL protocol over HTTP. Fuseki implements W3C's SPARQL 1.1 Query, Update, Protocol and Graph Store HTTP Protocol.

## 9. Conclusions and future plans

In the presented work we have successfully completed our research objectives, i.e. to collect some available methods and tools for classification, with which we could implement a Classification server. The selected tools (Skosmos and Jena Fuseki) seemed to be a good choice, despite the difficulties during implementation, as well as with the upload of certain classifications.

The classification server has fulfilled the requirements that we have set up. The current stable version contains 14 internal and 4 external classifications at the moment. There have been hundreds of visits since the official launch of the Classification server, and we are receiving positive feedbacks from users continuously.

The general online external classifications (AGROVOC, Eurovoc, STW and UNESCO) are currently using the SPARQL server of the National Library of Finland, because this server is operated by Skosmos, and this way we can guarantee that these vocabularies work well from our Classification server. In the near future we are going to redirect these external requests to their original data source. The connection between the Classification server and Phaidra is still in development at the moment.

Sándor Kopácsi†, PhD  
ehem. Universität Wien, Zentraler Informatikdienst

Mag. Rastislav Hudak  
Universität Wien, Zentraler Informatikdienst  
E-Mail: [rastislav.hudak@univie.ac.at](mailto:rastislav.hudak@univie.ac.at)

Dipl.-Ing. (FH) Raman Ganguly  
ORCID: <http://orcid.org/0000-0002-9837-0047>  
Universität Wien, Zentraler Informatikdienst  
E-Mail: [raman.ganguly@univie.ac.at](mailto:raman.ganguly@univie.ac.at)

† Sándor Kopácsi regrettably died in August 2017, just before the publication of this issue.

## References

1. Digital Preservation Coalition: Introduction: Definitions and Concepts. Digital Preservation Handbook. York, UK. (2008). <http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>
2. Day, Michael: The long-term preservation of Web content. Web archiving, Springer, pp. 177–199. (2006).
3. Lars Marius Garshol: Metadata? Thesauri? Taxonomies? Topic Maps! <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html#N429>
4. Hedden Information Management: Taxonomies, Thesauri, and Controlled Vocabularies. <https://www.hedden-information.com/taxonomies.htm>
5. The Web Graph Database: What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model? <http://infogrid.org/trac/wiki/Reference/PidcockArticle>
6. W3C Semantic Web: Introduction to SKOS. <http://www.w3.org/2004/02/skos/intro>
7. Zeng, M. L., & Chan, L.M.: Semantic Interoperability. Encyclopedia of Library and Information Sciences 4th Edition, p. 8. (2015).
8. Bob DuCharme: Learning SPARQL, 2nd Edition, Querying and Updating with SPARQL 1.1, O'Reilly Media (2013).
9. Suominen, O., Ylikotila, H., Pessala, S., Lappalainen, M., Frosterus, M., Tuominen, J., Baker, T., Caracciolo, C., Retterath, A.: Publishing SKOS vocabularies with Skosmos. Manuscript submitted for review (2015).
10. Osma Suominen: Publishing SKOS concept schemes with Skosmos. AIMS Webinar 6<sup>th</sup> April 2016, Slide 25. (2016).

## Notes

- 1 <https://www.w3.org/2004/02/skos/intro>
- 2 <https://www.w3.org/TR/rdf-sparql-query/>
- 3 <https://phaidra.univie.ac.at/>
- 4 <https://www.w3.org/TR/rdf-sparql-query/>
- 5 <http://thmanager.sourceforge.net/>
- 6 <http://www.vocabularyserver.com/>
- 7 <https://ch.semweb.ch/leistungen/thesaurus-services/en-thesauri/?ucl=en>
- 8 <https://www.poolparty.biz/>
- 9 <http://protege.stanford.edu/>

- 10 <http://skosmos.org/>
- 11 <http://onlinelibrary.wiley.com/doi/10.1002/bult.2011.1720370407/full>
- 12 <http://iqvoc.net/>
- 13 <http://www.cs.vu.nl/STITCH/repository/>
- 14 Representational State Transfer that relies on stateless, client-server, cacheable HTTP communication.
- 15 <http://skosmos.org/>
- 16 <http://onki.fi/en/>
- 17 <http://light.onki.fi/fi/>
- 18 [https://jena.apache.org/documentation/serving\\_data/](https://jena.apache.org/documentation/serving_data/)
- 19 <https://varnish-cache.org/>
- 20 <https://github.com/NatLibFi/Skosmos/wiki/Installation>
- 21 <https://bitbucket.org/art-uniroma2/owlart/downloads>
- 22 <https://code.google.com/p/skosify/>
- 23 <http://vocab.getty.edu/>
- 24 <https://www.coar-repositories.org/activities/repository-interopability/ig-controlled-vocabularies-for-repository-assets/deliverables/>
- 25 <http://vocabularies.coar-repositories.org/sparql/repositories/coar>
- 26 See the description of "Setting and getting vocabularies" in section 6.3.
- 27 <http://www.statistik.at/KDBWeb/kdb.do?FAM=OESTERR&&NAV=EN&&KDBtoken=null>
- 28 <https://www.oecd.org/science/inno/38235147.pdf>
- 29 [https://www.statistik.at/web\\_de/statistiken/index.html](https://www.statistik.at/web_de/statistiken/index.html)

## ■ DATA SCIENCE AND ANALYTICS IN LIBRARIES

by José Luis Preza

**Abstract:** Libraries have the virtue of managing vast amounts of information. Data Science and Analytics techniques and methodologies allow libraries to fully exploit the content they hold with the goal of providing better information to their users: better search, recommendations, etc.

**Keywords:** Data Science; Analytics; Libraries; Machine Learning; e-Infrastructures Austria; Metadata

## DATA SCIENCE UND DATENANALYTIK IN BIBLIOTHEKEN

**Zusammenfassung:** Bibliotheken sind in einer privilegierten Situation: Sie verwalten riesige Mengen von Daten und Informationen. Data Science und Analytics-Methoden ermöglichen es Bibliotheken, den Inhalt, den sie verwalten, voll auszunutzen, um den Nutzern bessere Informationen, Suche und Empfehlungen zu bieten.

**Schlüsselwörter:** Data Science; Datenanalytik; Bibliotheken; Machine Learning; e-Infrastructures Austria; Metadaten

This document was prepared as part of the Deliverables of Cluster I of e-Infrastructures Austria, a nationwide project regarding the design and management of digital infrastructures for research data ([www.e-infrastructures.at](http://www.e-infrastructures.at)). See also the preprint version: José Luis Preza. (2017, March 8). Data Science and Analytics in Libraries. Zenodo. DOI: <http://doi.org/10.5281/zenodo.375809>.



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)



## **Contents**

1. *What is Data Science?*
2. *Data Science application in libraries*
3. *Use Case for Analytics in Libraries: Analytical platform for an institutional repository*

### **1. What is Data Science?**

Data Science is simply a discipline that combines data with programming languages, algorithms, statistics, machine learning, artificial intelligence, reporting, and data visualization, all to make sense out of data. Data Science is a very important part of Cognitive Computing that enables Artificial Intelligence.

Public, School and University Libraries are in a very advantageous position: they sit on a lot of data. The data stored in such libraries are very diverse. There are books, documents, charts, datasets, experiments, software, tables, images, videos, audio, dissertations, magazines, newspapers, processes, usage, user data, financial data, to mention a few. The challenge for libraries is not only to digitize all their content (digital objects), but also to classify, organize, link and publish all digital objects.

Up until now, most content (digital objects) has been organized and classified manually. However, manual processes are not sustainable, certainly not when you have to process millions of digital objects in a short period of time and with high accuracy. Here is where Data Science comes to the rescue.

### **2. Data Science application in libraries**

The techniques and methods used in Data Science allow libraries to lighten the workload and get results faster than with manual processes. Concrete areas where Data Science can assist libraries include:

- **Digital Object Classification/Semantics/Search:** Automatic classification of digital objects (keywords, entities, concepts).<sup>1</sup>
- **Picture Recognition and Classification:** automatic classification and tagging of pictures (also extracted from video).
- **Content Clustering and Segmentation:** Automatic clustering and segmentation of digital objects based on content.

- **Reporting:** Generate reports from your content.
- **Predictive Analytics:** Answer the question of who is going to read/use what?
- **Machine Translation:** Automatic translation of digital objects, including e.g. Braille.
- **Speech to Text:** Extraction of audio speech to convert it into text.
- **Text to Speech:** Convert text into audible speech.
- **Plagiarism Detection:** Machine learning enables the development of advanced techniques to prevent plagiarism.
- **Analytical Platform for Institutional Repository:** Advanced reporting and analysis of digital content in institutional repositories.

### **3. Use Case for Analytics in Libraries: Analytical platform for an institutional repository**

Most repository software applications lack a module to analyze in detail – and visually – the usage, storage and other key indicators within the repository. This is true for any open source package<sup>2</sup>. Commercial repository applications like Mendeley might have an analytical module. At best, system administrators create e.g. PERL scripts to extract particular information out of their systems. This information, while useful, is limited and is sysadmin oriented.

Some repositories show how often a particular digital object has been downloaded. This is normally done to show the end user how popular a particular digital object is. Naturally, having the information of how many times a digital object has been downloaded might be a useful indicator, but is not really sufficient for a Repository Manager. This information should be aggregated; displaying it next to the digital object might not be ideal to analyze downloads as a whole.

To have a good idea of what is going on in the repository, the Repository Manager requires a good overview of the content and activities within a repository.

#### **3.1. What information should be analyzed?**

1. **Searches:** it would be important to know what the users are searching for within the repository, what the needs are, what the top searches are. Are users searching for inappropriate content?
2. **Bandwidth:** another good indicator to keep track of.

3. **Storage:** self-explanatory. Storage and bandwidth data can assist the Repository Manager and the Institution to justify additional funds, plan growth and usage, estimate costs.
4. **Users and usage:** what are users doing in the repository? Who owns what?
5. **Traffic:** logging all web traffic is always a good idea. Applications like Piwik might be a good option for those repositories that do not include a web traffic management module. Things to track include visits, duration of visits, referring URLs, events (upload, download, etc.), browsers used, etc.
6. **Digital Objects:** what is inside your repository?
7. **Classifications:** usually, when a digital object is uploaded to a repository, the system will ask for a "tag" or "classification" of the object. An object can have more than one tag or classification.
8. **Audio, Video, Text, Image analysis:** what is inside the digital objects? This task can be done automatically using cognitive services.
9. **Top Ten:** the top ten biggest files, top ten most downloaded files, top ten uploaders, top ten searches, etc.
10. **Additional information** such as: files that have never been downloaded or seen, users who have never logged in, etc.

All these data should be aggregated on the fly by user, object, year, month, day, content model, etc. The analytical application should be multisite, multitenant, multiuser, web-based, and easy to use.

### **3.2. Phaidra Statistics**

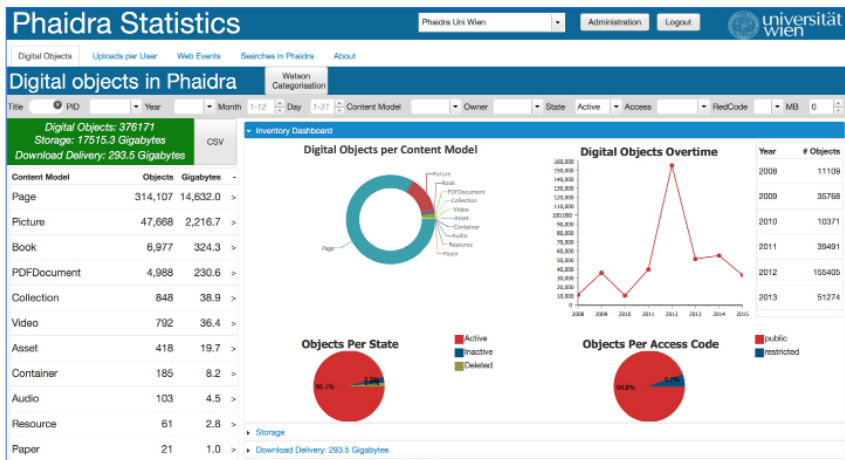
I developed such an analytical platform for Phaidra, the repository at the University of Vienna. This platform manages large amounts of metadata belonging to all the digital objects stored in Phaidra.

I also integrated IBM Watson within it to perform automatic classification of digital objects<sup>3</sup>.

Phaidra's backend uses Fedora Commons. The frontend has been developed by the University of Vienna and uses Piwik to log traffic.

Phaidra University of Vienna: <https://phaidraservice.univie.ac.at>

Phaidra Statistics GitBook: <https://www.gitbook.com/book/jluni/phaidra-statistics/details>



José Luis Preza  
E-Mail: [jl@preza.org](mailto:jl@preza.org)  
Website: <http://www.preza.org>

## Links

1. Wikipedia definition of Recommender System: [https://en.wikipedia.org/wiki/Recommender\\_system](https://en.wikipedia.org/wiki/Recommender_system)
2. Analytical Platform for an Institutional Repository: <https://www.linkedin.com/pulse/analytical-platform-institutional-repository-jos%C3%A9-luis-preza-d%C3%ADaz>

## Notes

- 1 See also: José Luis Preza. (2016, October 31). Automated Information Enrichment for a Better Search. Zenodo. DOI: <http://doi.org/10.5281/zenodo.163933>
- 2 See: José Luis Preza. (2016, September 25). The Best Repository for (Research) Data. LinkedIn. [https://www.linkedin.com/pulse/best-repository-research-data-jos%C3%A9-luis-preza-d%C3%ADaz?trk=pulse\\_spock-articles](https://www.linkedin.com/pulse/best-repository-research-data-jos%C3%A9-luis-preza-d%C3%ADaz?trk=pulse_spock-articles)
- 3 See: José Luis Preza. (2016, April 2). IBM Watson Analytics for Information Enrichment and a Better Search. <https://www.linkedin.com/pulse/ibm-watson-analytics-information-enrichment-better-preza-d%C3%ADaz>

## ■ UMGANG MIT METADATEN IN REPOSITORIEN – EINE ÖSTERREICHWEITE UMFRAGE

von Susanne Blumesberger & Alexander Zartl

**Zusammenfassung:** Dieser Artikel enthält eine Auswertung der Telefonumfrage zum Thema „Metadaten“, die im Sommer 2015 unter den am Projekt e-Infrastructures Austria teilnehmenden Institutionen durchgeführt wurde. Ziel der Befragung war eine Bestandsaufnahme der an den einzelnen Universitäten im Einsatz befindlichen Repositorien, um einen Überblick über die wichtigsten Softwaresysteme und deren spezifische Einsatzzwecke, die am häufigsten verwendeten Metadatenstandards und die Gepflogenheiten bzw. Erwartungen der Nutzer und Nutzerinnen zu gewinnen. Zukünftige Planungen und Entwicklungen waren ebenfalls Gegenstand des Fragebogens.

**Schlüsselwörter:** Metadaten; Repositorien; Österreich

### HANDLING METADATA IN REPOSITORIES – AN AUSTRIA-WIDE SURVEY

**Abstract:** This article discusses the results of a telephone survey on the subject of metadata that was carried out in summer 2015 among the institutions participating in the project e-Infrastructures Austria. The aim of the survey was to provide an overview of the most important software systems and their specific uses, the most commonly used metadata standards and the habits and expectations of the users. Future plannings and developments were also included in the questionnaire.

**Keywords:** Metadata; repositories; Austria



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

#### **Inhalt**

1. Vorbemerkung
2. Liste der befragten Institutionen
3. Ergebnisse

## 1. Vorbemerkung

Ziel der Befragung war, festzustellen, wie man möglichst rasch und effizient herausfinden könnte, wie die Partnerinstitutionen am Projekt *e-Infrastructures Austria* derzeit mit Metadaten in ihren Repositorien umgehen, bzw. wie ihre Planung diesbezüglich aussieht. Da zum Zeitpunkt der Fragebogenerstellung (April 2015) noch nicht alle Partnerinstitutionen über ein Repositoryum verfügten und somit auch keine Erfahrungen mit Metadaten in diesem Bereich hatten, musste die Befragung ziemlich niederschwellig angesetzt werden. Die Gruppe einigte sich auf einen schriftlichen Fragenbogen, der zwar an die AnsprechpartnerInnen geschickt werden sollte, von ihnen jedoch nicht alleine ausgefüllt, sondern in Form eines angekündigten telefonischen Gesprächs abgehandelt werden sollte. Der Vorteil dabei ist, dass im persönlichen Gespräch nachgehakt werden kann und sollte etwas nicht oder falsch verstanden worden sein, die Möglichkeit zur Nachfrage gegeben ist.

In einem ersten Schritt wurden die Fragen überlegt und möglichst klar und deutlich formuliert. Danach wurden die jeweiligen AnsprechpartnerInnen an den einzelnen Institutionen gesucht. In einer vorbereiteten Mail wurde um einen Gesprächstermin ersucht, der Fragebogen wurde bereits beim ersten Kontakt mitgeschickt, um den GesprächspartnerInnen die Möglichkeit zu geben, sich vorab Antworten überlegen zu können, bzw. auch um ihnen die Gelegenheit zu bieten, weitere ExpertInnen an der eigenen Institution in die Befragung miteinzubeziehen.

Während des telefonischen Interviews wurde das Gespräch in Stichworten niedergeschrieben und gleich danach ergänzt. Die personalisierten Fragebögen wurden archiviert, aber – so ist es mit den InterviewpartnerInnen abgesprochen – nicht an die Öffentlichkeit weitergegeben, sondern nur anonym weiter verarbeitet.

Zu den Umfrageergebnissen ist zu ergänzen, dass die Ergebnisse natürlich nur eine Momentaufnahme sein können. Es zeigt sich jedoch bereits deutlich, dass der Umgang mit Metadaten in einem Repositoryum noch keineswegs selbstverständlich bzw. erprobt ist. Für viele Institutionen sind Repositorien eine neue Aufgabe. Metadaten sind zwar aus dem bibliothekarischen Alltag bekannt, die Umsetzung dieses Wissens auf die Beschreibung digitaler Objekte ist jedoch noch relativ ungewohnt und steht ziemlich am Anfang. Hier ist nicht nur technisches Know-How wichtig, sondern vor allem auch das Bewusstsein, wie wichtig Metadaten für ForscherInnen bzw. für künftige NutzerInnen der Objekte sind.

Eine Anmerkung zu den gewählten Diagrammtypen: Balkendiagramme kommen stets dort zum Einsatz, wo Mehrfachantworten möglich waren

und die Summe der Prozentzahlen daher größer als 100 sein kann. Entscheidungsfragen mit nur einer Antwortmöglichkeit werden durch Kreisdiagramme veranschaulicht.

Wir bedanken uns bei allen GesprächspartnerInnen für die freundliche Unterstützung des Projekts.

## 2. Liste der befragten Institutionen

Mit Ausnahme der Österreichischen Bibliothekenverbund und Service Ges.m.b.H., die aufgrund ihrer speziellen Aufgabenstellung mit Bezug auf den gesamten Bibliotheksbereich eine Sonderstellung einnimmt und daher nicht vergleichbar ist, konnten im Zeitraum von Juni bis September 2015 alle am Projekt *e-Infrastructures Austria* teilnehmenden Organisationen in die Befragung einbezogen werden.

- Akademie der bildenden Künste Wien
- Arbeiterkammer Wien
- Institute of Science and Technology Austria (IST)
- Medizinische Universität Graz
- Medizinische Universität Wien
- Montanuniversität Leoben
- Österreichische Akademie der Wissenschaften
- Österreichische Nationalbibliothek
- Technische Universität Graz
- Technische Universität Wien
- Universität für angewandte Kunst Wien
- Universität für Bodenkultur Wien
- Universität für künstlerische und industrielle Gestaltung Linz
- Universität für Musik und darstellende Kunst Graz
- Universität für Musik und darstellende Kunst Wien
- Universität Graz
- Universität Innsbruck
- Universität Klagenfurt
- Universität Linz
- Universität Mozarteum Salzburg
- Universität Salzburg
- Universität Wien
- Veterinärmedizinische Universität Wien
- Wirtschaftsuniversität Wien

## 3. Ergebnisse

### 3.1. Allgemeine Fragen

#### 3.1.1. Welche Arten von Daten werden an Ihrer Einrichtung gesammelt?

Bei dieser Frage gab es kein vorgegebenes Antwortvokabular, sondern es wurden die von den InterviewpartnerInnen gewählten Bezeichnungen übernommen. Daraus ergibt sich eine unscharfe Kategorienbildung mit Überlappungen zwischen den Begriffen „Bilder“ und „Grafiken“ oder „Texte“ und „Publikationen“ etc. Diese Vorgangsweise wurde gewählt, um zu verhindern, dass bestimmte Arten von Daten von vornherein ausgeschlossen werden.

Hier wird ersichtlich, dass an erster Stelle Hochschulschriften archiviert werden (sollen), in weiterer Folge Bilder, Publikationen, Texte, Videos, und an 6. Stelle Forschungsdaten.

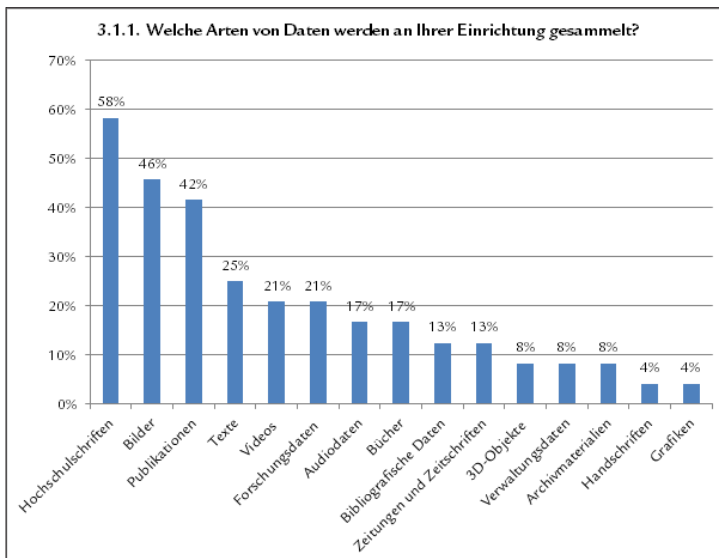
Speziell zum Begriff „Forschungsdaten“ ist anzumerken, dass diese Bezeichnung weiter ausgelegt wurde, als es die üblichen Definition – Messprotokolle, Umfragedaten, sonstige Rohdaten – nahelegt. So wurden z.B. in einem Fall explizit multimediale (Video-)Daten darunter verstanden. Insgesamt gilt damit das Prinzip, dass der Kontext bestimmt, was Forschungsdaten sind und nicht der Datentyp.

Das entspricht auch der Definition, die für die Umfrage, die sich an Forschende an der Universität Wien richtete, verwendet wurde:

„Unter dem Begriff ‚Forschungsdaten‘ sind alle Daten zu verstehen, die im Zuge wissenschaftlicher Forschungs- und künstlerischer Schaffensprozesse entstehen (z.B. Text, Tabellen, Video, Audio, Grafik etc.) und/oder auf deren Grundlage Ihre Forschungsergebnisse und/oder Kunstwerke basieren – z.B. durch Experimente, Quellenforschungen, Messungen, Erhebungen, Digitalisate oder Entwürfe.“

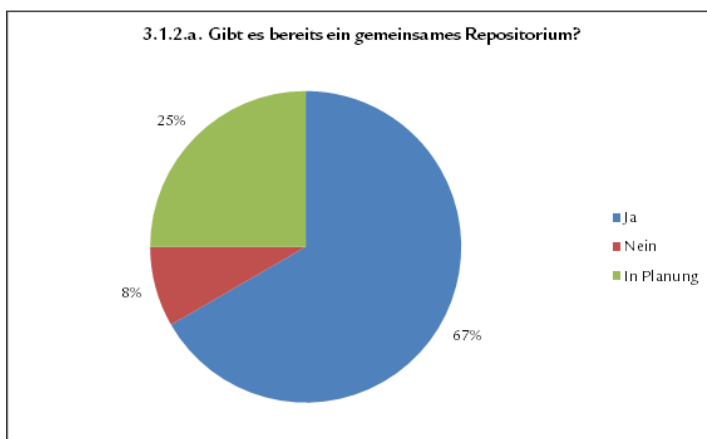
Bei der Interpretation der Grafik ist zu beachten, dass manche Einrichtungen über mehrere Repositorien verfügen und darin unterschiedliche Arten von Daten archiviert werden, wie etwa auch an der Universität Wien, wo der Hochschulschriftenserver wissenschaftliche Abschlussarbeiten enthält, das Repositorium Phaidra für alle Dateiformate offen ist und grundsätzlich alle genannten Arten von Daten archivieren kann und zusätzlich Unidam vor allem, aber nicht nur für Bildmaterial zur Verfügung steht.



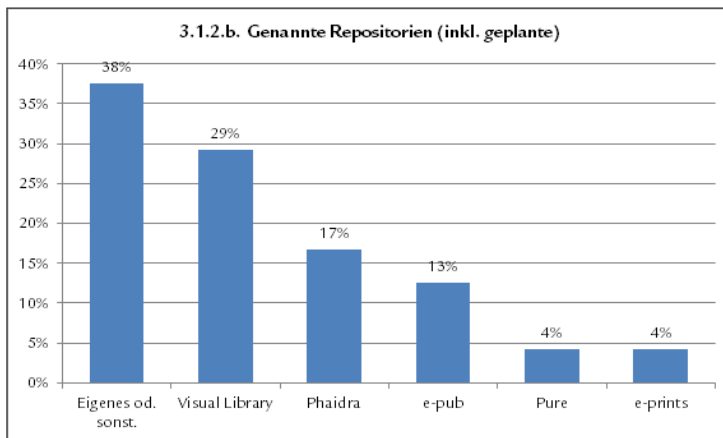


### 3.1.2. Gibt es bereits ein gemeinsames Archiv bzw. Repository für die Angehörigen Ihrer Einrichtung?

Nur zwei der TeilnehmerInnen am Projekt e-Infrastructures Austria planen zum Zeitpunkt der Befragung noch kein Repository; bei den in Planung befindlichen Repositorien war noch nicht überall bekannt, welches System gewählt werden würde.



Bei den bereits vorhandenen oder fix geplanten Systemen sah die Verteilung folgendermaßen aus:



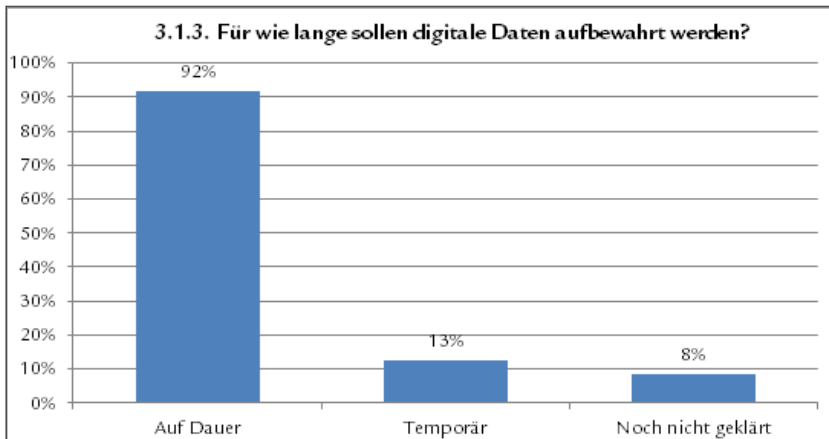
Abgesehen von den Spitzenreitern Visual Library, Phaidra und e-pub, die an mehr als zehn Prozent der Einrichtungen in Gebrauch stehen, zeigt sich eine sehr uneinheitliche Repositorienlandschaft mit einem hohen Anteil an Eigenentwicklungen.

Das verbreitetste Produkt mit einem Anteil von 29% ist Visual Library der semantics Kommunikationsmanagement GmbH. Dem Einsatz dieser Software liegt ein Konsortialmodell der Österreichischen Bibliothekenverbund- und Service-GmbH zu Grunde, sodass die nötige Infrastruktur (Hardware, Software, Datensicherung) zentralisiert ist und nicht von den Einrichtungen selbst getragen werden muss. Visual Library ist zudem in die bestehenden Workflows des Bibliothekenverbundes integriert und kann durch die Suchmaschine Primo erschlossen werden.

Ebenfalls auf Visual Library basieren die Publikationsserver BOKU:ePUB (Universität für Bodenkultur Wien), uni-pub (Universität Graz) etc. Eine Entwicklung der TU Graz ist hingegen das Portal epub.oaew der Österreichischen Akademie der Wissenschaften.

Phaidra (Akronym für: Permanent Hosting, Archiving and Indexing of Digital Resources and Assets) steht bei insgesamt 14 Einrichtungen in Österreich und dem europäischen Ausland im Einsatz. Technisch basiert das System auf der Open-Source-Software Fedora, federführend bei der Entwicklung ist die Universität Wien. Im Gegensatz zu Visual Library bilden die Phaidra-Anwender keinen technischen Verbund, sondern betreiben jeweils eigene (z.T. auch ausgelagerte) Instanzen.

### 3.1.3. Für wie lange sollen digitale Daten aufbewahrt werden? Auf Dauer oder nur temporär?

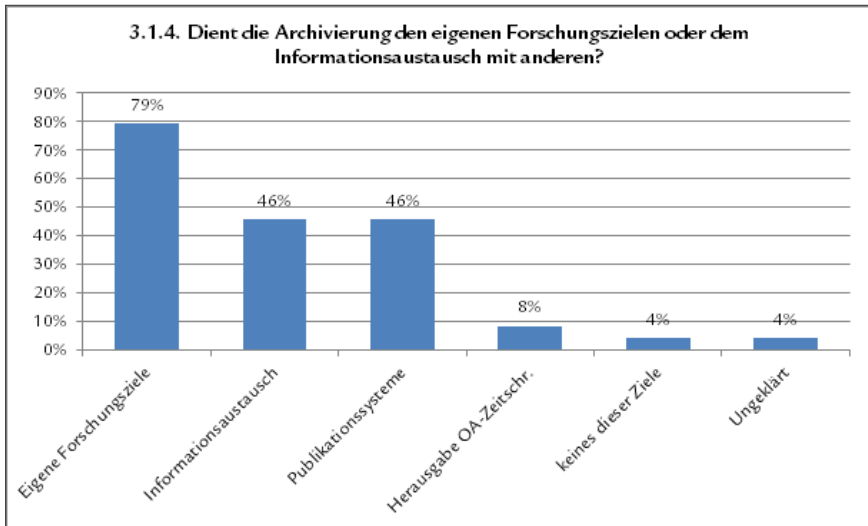


Die Grafik zeigt, dass alle befragten Institutionen, die diese Frage schon geklärt haben, ihre Daten langzeitarchivieren möchten. In drei Fällen ist zusätzlich auch die Möglichkeit der temporären Speicherung wichtig. Der Wunsch nach rein temporärer Archivierung wurde nicht genannt.

Der Wunsch nach zeitlich begrenzter Speicherung ergibt sich vorwiegend aus dem Bedarf nach einer sicheren Ablage für vorläufige Arbeitsdokumente, Rohfassungen und sonstige Zwischenergebnisse mit Ablaufdatum – nicht nur aus dem wissenschaftlichen Bereich, sondern vor allem auch in administrativem Zusammenhang. Ein Feature, welches die Entfernung nicht mehr benötigter Dokumente aus dem System grundsätzlich ermöglicht, wurde von einigen der Interviewpartner als grundlegende Funktion eingestuft.

### 3.1.4. Dient die Archivierung den eigenen Forschungszielen der Einrichtung oder dem Informationsaustausch mit anderen?

Bei dieser Frage waren mehrere Antworten möglich. Die Grafik zeigt, dass die Archivierung hauptsächlich der eigenen Forschung, bzw. der Darstellung des eigenen Forschungoutputs dient (Publikationsserver, z.T. auch mit PR- und Repräsentationsfunktion). An der zweiten Stelle steht der Informationsaustausch.



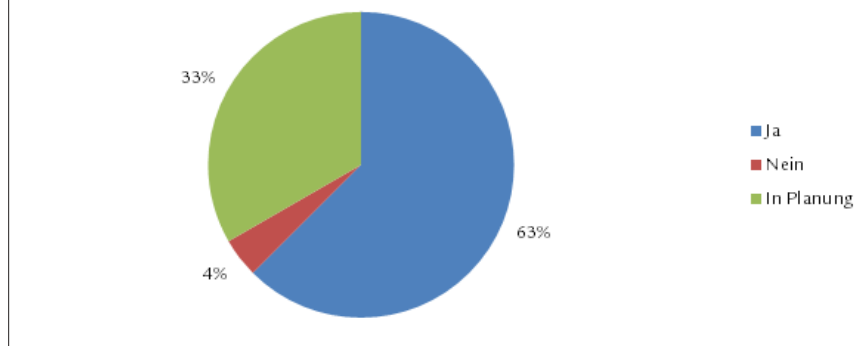
### 3.1.5. Gibt es eine Schnittstelle zwischen Bibliothekskatalog und Repository bzw. ist eine solche in Planung?

Die Schnittstelle zwischen Bibliothekskatalog und Repository ist für die meisten Partner bereits realisiert oder in Planung.

Am besten vernetzt sind die verschiedenen Repositorien nach heutigem Stand mit dem Bibliothekssystem ALEPH, was sich aus dem Anspruch auf eine Erfassung der digitalen Ressourcen in den etablierten Bibliothekskatalogen und -suchmaschinen ergibt: die zentrale Nachweisstelle soll im Normalfall der Katalog bleiben. Eine ausgereifte Lösung in diesem Zusammenhang bietet Visual Library, das einerseits durch die Suchmaschine Primo erschlossen werden kann (vgl. 3.1.2.) und andererseits die Möglichkeit bietet, die in ALEPH erzeugten Katalogdaten automatisch zu übernehmen. In Fällen, wo eine Übernahme von Metadaten aus dem Repository nach ALEPH per Datenschnittstelle bisher nicht möglich ist, spielt die manuelle Katalogisierung (mittels Eintrag in Kategorie 655) noch immer eine wichtige Rolle.

In Einzelfällen realisiert sind auch Schnittstellen zwischen lokal betriebenen Forschungsinformationssystemen und SCOPUS oder Web of Science (z.B. zur Anreicherung des Katalogisats mit Metadaten zu Zeitschriftenveröffentlichungen) oder SAP (als Unterstützung für das Rechnungswesen).

### 3.1.5. Gibt es eine Schnittstelle zwischen Bibliothekskatalog und Repository bzw. ist eine solche in Planung?



## 3.2. Fragen mit Bezug auf Metadaten

### 3.2.1. Gibt es für die Metadaten eine spezifische Policy?

Wie ersichtlich, haben erst sehr wenige Institutionen eine spezifisch ausformulierte Policy für Metadaten. Jene, die bereits eine Policy haben, verwenden i.d.R. internationale Standards.

Die University of Bath richtet sich in ihrer „Research Data Policy Guidance“ an alle Forschenden, egal ob sie intern oder extern finanziert sind und gibt folgendes vor: „Researchers must provide sufficient metadata (descriptive information) and explanatory documentation about their research data to ensure that data are discoverable, understandable and re-usable“.

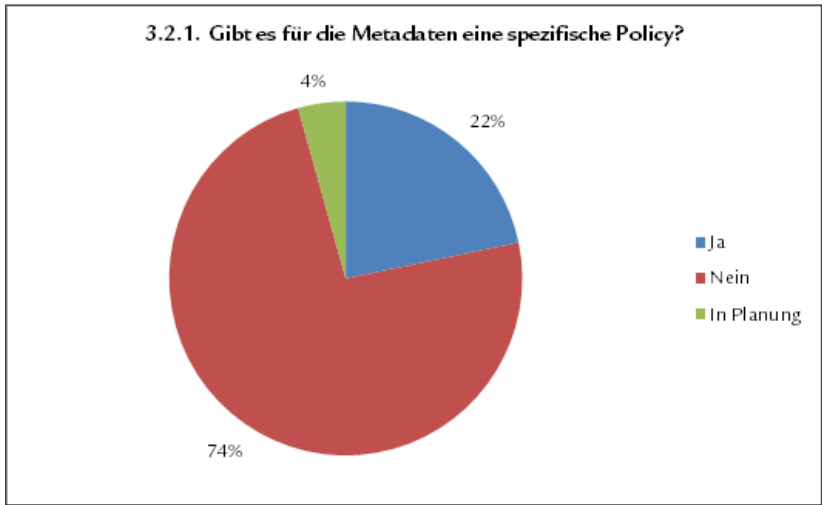
Die Universität Bath stellt folgende Forderungen:

1. Documentation and descriptive information must be recorded as early as possible after data creation.
2. Relevant metadata standards or minimal information sets should be used where applicable.
3. Descriptive information should enable other researchers to understand the research data sufficiently to be able to re-use it for any reasonably foreseeable purpose, such as those outlined in the original project proposal or impact plan.<sup>1</sup>

Europeana und z.B. die Aalto University empfehlen, für die Metadaten die Creative Common Lizenz CC0 1.0<sup>2</sup> zu vergeben.<sup>3</sup> Dies kann in manchen

Ländern, wie zum Beispiel in Österreich, aufgrund des Urheberrechtsgesetzes problematisch sein.

Insgesamt gibt es derzeit zwar wenige Policies, die sich direkt auf Metadaten beziehen, indirekt werden die Metadaten in den Policies aber oft erwähnt.<sup>4</sup>

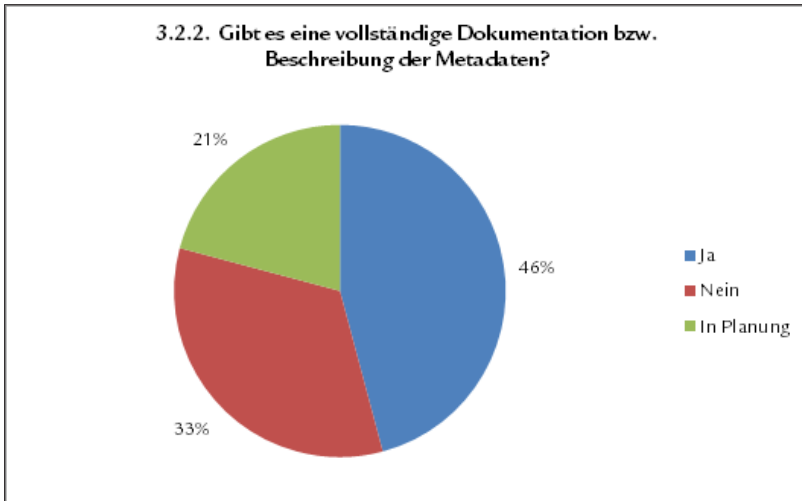


### 3.2.2. Gibt es eine vollständige Dokumentation bzw. Beschreibung der Metadaten?

Bei der Zählung wurde die Antwort „teilweise“ als „Ja“ gewertet, im Aufbau befindliche Dokumentationen wurden dem Punkt „in Planung“ zugeschlagen.

Inwieweit eine Beschreibung der Metadaten notwendig ist, hängt unter anderem davon ab, ob das Repositorium auch Metadatenfelder anbietet, die über Angaben zu AutorInnen, zum Titel, zum Datum oder über eine Beschreibung hinausgehen. In Phaidra, das ja der gesamten Universität Wien, also sehr vielen Disziplinen zur Verfügung steht, werden zum Teil sehr detaillierte Felder angeboten, die nicht selbsterklärend sind. Eine Besonderheit ist, dass man in Phaidra auch Angaben zu analogen Objekten machen kann, beispielsweise kann man die Abmessungen des Originalobjekts eintragen, was für Sammlungen bedeutend sein kann, es können aber auch zwei weitere Daten für analoge Objekte eingetragen werden. So besteht die Möglichkeit z.B. für die Entstehungszeit eines Bauwerkes ein „Datum

von“ und ein „Datum bis“ einzutragen, bzw. eine zeitliche Abdeckung einzutragen. Diese Angaben zur Provenienz, bzw. die kontextuellen Angaben beinhalten auch mehrere freie Felder, die mit forschungsrelevanten Informationen gefüllt werden können. Ohne zusätzliche Angaben würde dies die User verwirren.



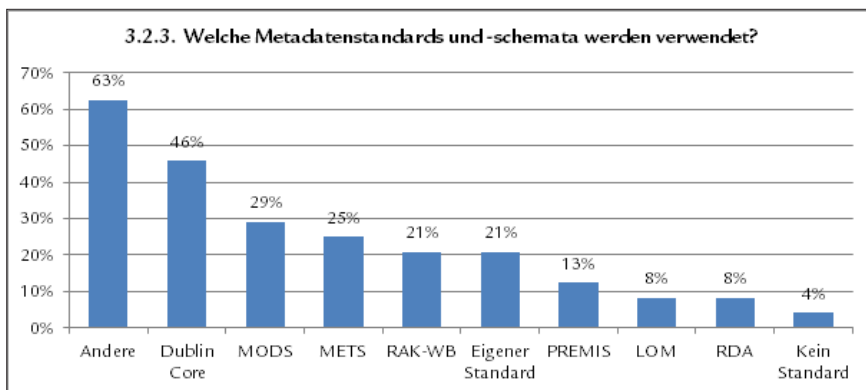
### 3.2.3. Welche Metadatenstandards und -Schemata werden innerhalb Ihrer Institution verwendet?

Hier werden nur die am häufigsten genannten Schemata eigens berücksichtigt, die übrigen werden unter „Andere“ zusammengefasst. Es gab kein vorgegebenes Antwortvokabular, was dazu führt, dass auch Metadatenstandards genannt wurden, die keine Standards für Repositorien sind. Zugelassen wurde auch die Nennung der Regelwerke RAK bzw. RDA statt der inhaltlich besser passenden Datenformate MAB oder MARC. Die Wahl der Standards hängt natürlich stark vom jeweiligen Langzeitarchivierungssystem ab. Dublin Core ist Spitzenreiter.

Kurzbeschreibungen der am häufigsten genannten Standards:

- Dublin Core: ein von auf Initiative des OCLC seit Mitte der 1990er Jahre entwickelter Standard. 15 Kernfelder (Core elements), später erweitert um sogenannte qualified terms. Es gibt keine Pflichtfelder, alle Felder sind mehrfach verwendbar. Ausführliche Dokumentation und Anwendungsbeispiele unter <http://dublincore.org/>.

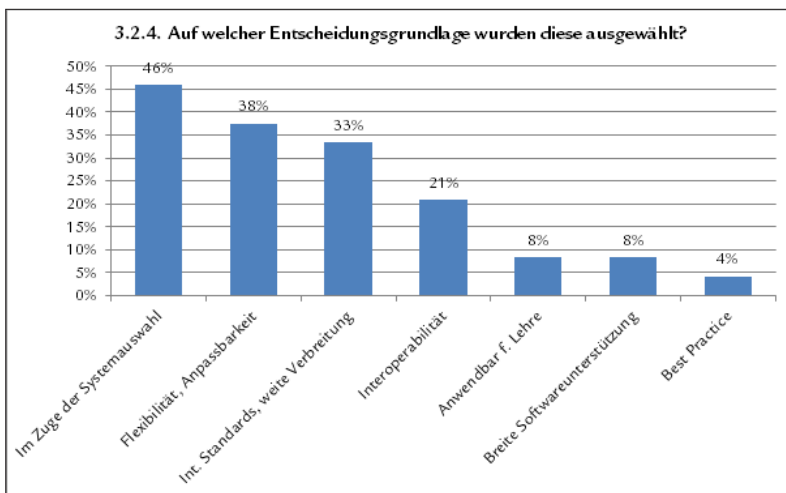
- MODS / METS (kurz für *Metadata Object Description Schema* bzw. *Metadata Encoding & Transmission Standard*): von der Library of Congress entwickelte und betreute XML-Schemata. Während MODS (teilweise in Anlehnung an MARC) für die Erfassung bibliografischer Metadaten gemacht ist, dient METS konkret zur Beschreibung von digitalen Sammlungen. Dokumentation unter <http://www.loc.gov/standards/mods/> bzw. <http://www.loc.gov/standards/mets/>.
- PREMIS (kurz für *PREservation Metadata: Implementation Strategies*): ebenfalls ein von der LoC für die strukturierte Beschreibung langzeitarchivierter digitaler Objekte. PREMIS ist somit eine mögliche Alternative zu METS; Gegenüberstellungen finden sich in der von der LoC angebotenen Dokumentation unter <http://www.loc.gov/standards/premis/>.



### 3.2.4. Auf welcher Entscheidungsgrundlage wurden die verwendeten Schemata ausgewählt?

Die Antworten waren relativ breit gefächert. An erster Stelle steht „Im Zuge der Systemauswahl“, was wohl impliziert, dass bei der Systemauswahl auch andere Kriterien als die dort verwendeten Metadatenstandards im Vordergrund standen und diese gewissermaßen „mitgekauft“ wurden. Auch auf Flexibilität und die Verwendung internationaler Standards (vgl. Frage 3.2.1.) wurde Wert gelegt.

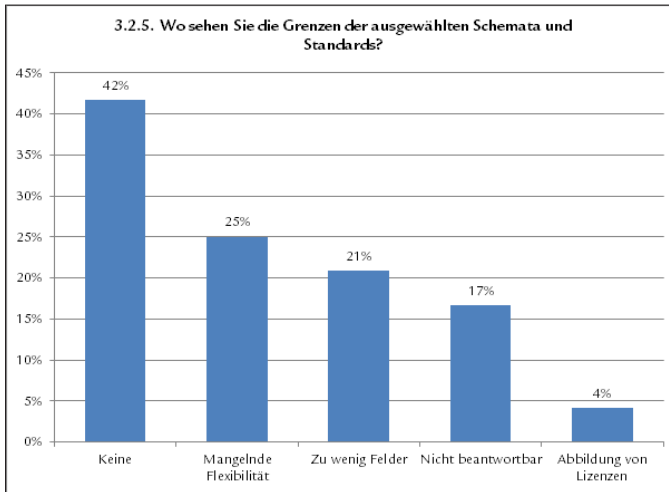




### 3.2.5. Wo sehen Sie die Grenzen der ausgewählten Schemata und Standards?

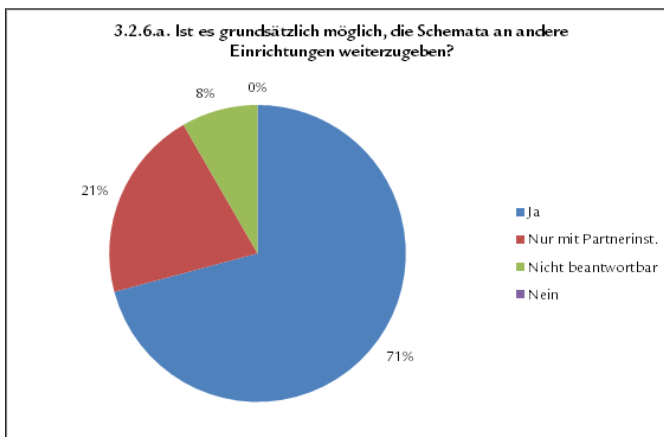
Die meisten InterviewpartnerInnen sahen (noch) keine Grenzen. Beim Beschreiben von sehr institutionsspezifischen Daten tritt gelegentlich der Wunsch nach zusätzlichen Metadatenfeldern auf. Andere Sonderwünsche wurden subsumiert unter dem Begriff „mangelnde Flexibilität“.

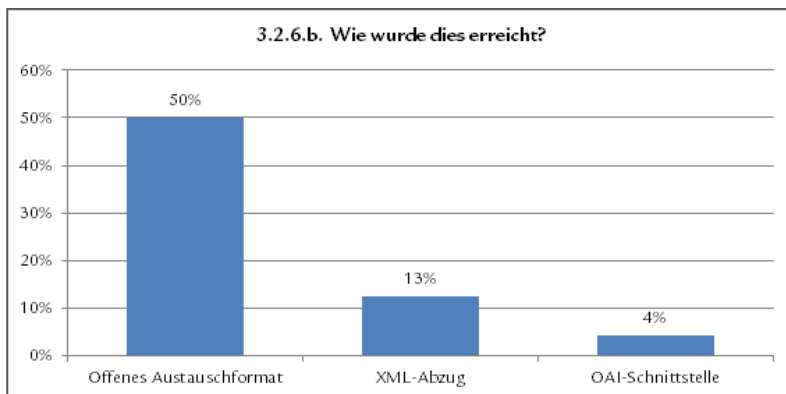
Wo die Grenzen von ausgewählten Standards und Schemata gesehen werden, hängt natürlich sehr stark vom Verwendungszweck des Repositoriums ab. Werden nur Hochschulschriften oder Zeitschriftenartikel archiviert, reichen die Angebote meist aus. Schwierig wird es erst dann, wenn man viele unterschiedliche Wünsche bedienen muss und das zu archivierende Material sehr heterogen ist. Am Beispiel der Universität Wien, wo sämtliche Nachlassmaterialien, historische Bücher, Bilder, Forschungsergebnisse in allen Formaten, audiovisuelle Materialien usw. langzeitarchiviert werden sollen, stößt man bei herkömmlichen Standards und Schemata recht schnell an die Grenzen, wenn man alle Wünsche der User erfüllen möchte. Erfahrungsgemäß genügen in den meisten Fällen eine eingehende Beratung und zum Teil ein kleiner Kompromiss, um nicht Spezialfelder anlegen zu müssen, die nicht kompatibel mit anderen Systemen sind, wodurch beim Austausch mit anderen evtl. Information verloren geht. Damit hängt auch die mangelnde Flexibilität, die von den Befragten genannt worden ist, zusammen. Je flexibler das System ist, desto schwieriger sind die Anpassung an andere Systeme und der Austausch der Daten.



### 3.2.6. Ist es grundsätzlich möglich, die Schemata an andere Einrichtungen weiterzugeben? Wie wurde dies erreicht?

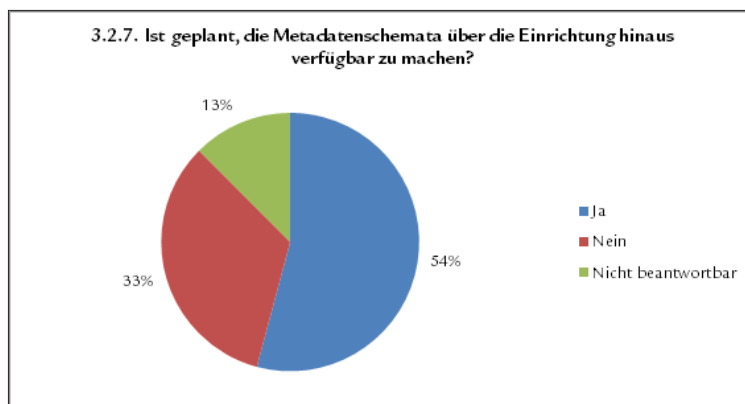
Bei dieser Frage waren sich bis auf jene Institutionen, die noch kein System haben und deshalb auch keine Antwort geben konnten, alle einig, dass die Schemata weitergegeben werden können. Abhängig vom gewählten System – wie z.B. bei Phaidra – geschieht dies durch den Austausch unter Partnerinstitutionen. Da fast alle ein offenes Format verwenden, ist eine prinzipielle Weitergabe zumeist von Vornherein möglich.





### 3.2.7. Ist geplant, die Metadatenschemata über die Einrichtung hinaus verfügbar zu machen?

Diese Frage wurde mehrheitlich mit „ja“ beantwortet, einige Institutionen hatten noch keine Entscheidung darüber getroffen. Das Open-Data-Konzept ist zurzeit noch wenig bekannt und wurde nur in wenigen Fällen explizit angesprochen.



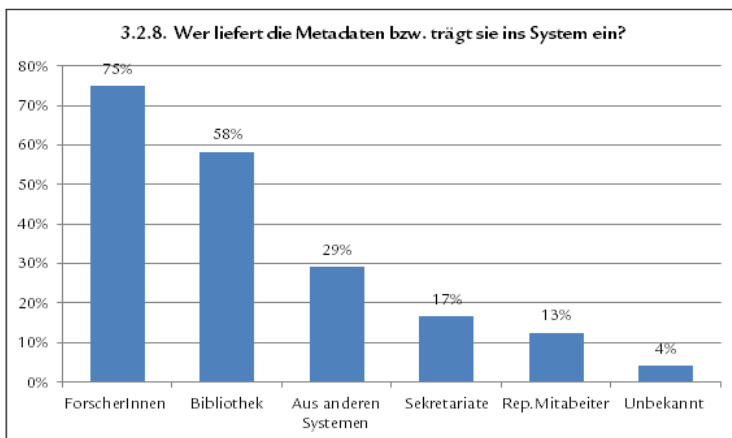
### 3.2.8. Wer trägt die Metadaten ins System ein?

Interessant ist der hohe Anteil jener Einrichtungen, bei denen die ForscherInnen und WissenschaftlerInnen die Metadaten selbst eintragen. Gleich

dahinter folgt das Modell, nach dem die Erfassung der Metadaten von der Bibliothek vorgenommen wird. In einigen Fällen werden die Metadaten (automatisiert) aus anderen Systemen übernommen, weniger oft durch die Verwaltungseinheiten auf Institutsebene (Sekretariate) oder speziell damit betraute RepositoriumsmitarbeiterInnen.

Tendenziell verhält es sich – erwartungsgemäß – so, dass Systeme, die der Forschungsdokumentation dienen, eher vom wissenschaftlichen Personal bzw. den Sekretariaten befüllt werden, während die Betreuung von Systemen im bibliothekarischen Umfeld durch BibliotheksmitarbeiterInnen erfolgt. Hier ergibt sich das Problem, dass beide Personengruppen i.d.R. nur nebenbei für die Repositorien arbeiten und keine spezielle Ausbildung dafür besitzen, was sich in z.T. wechselhafter Datenqualität niederschlägt. Der Beruf des „Content Managers“ mit einschlägigen spezifischen Kenntnissen ist bisher erst wenig verbreitet, wobei zumindest im Bibliotheksbereich ein wachsendes Bewusstsein für dieses Problemfeld zu konstatieren ist. Parallel dazu herrscht gemäß den Aussagen der BefragungsteilnehmerInnen an den Instituten überwiegend die Ansicht, dass (Meta-)Datenerfassung für digitalen Content keine wissenschaftliche, sondern eine bibliothekarische Aufgabe ist. Es handelt sich somit um ein Betätigungsfeld, das die Bibliotheken für sich neu erschließen und professionalisieren könnten.

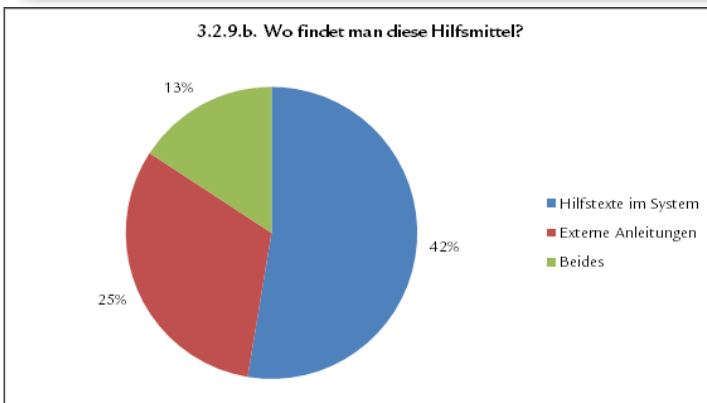
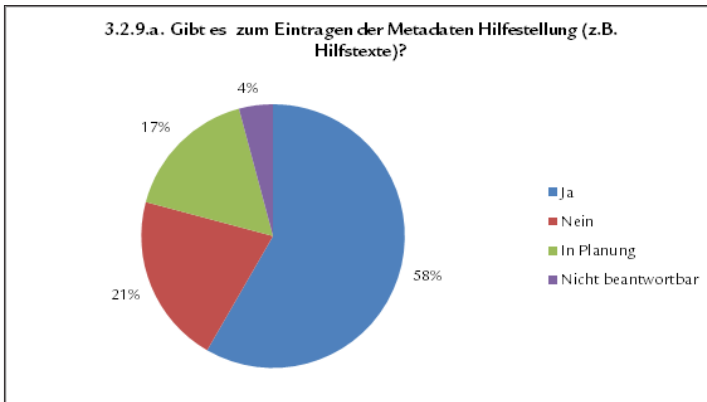
Auch hier hängt die Antwort stark vom Einsatz und der Policy des Repositoriums ab. So kann es z.B. aus rechtlichen Gründen notwendig sein, dass die User die Metadaten selbst ins System eintragen, weil sie für die Richtigkeit haften, zum Beispiel bei Hochschulschriften. Handelt es sich um Scans von historischen Archivbeständen oder Büchern, kann das Eintragen der Metadaten durchaus auch von BibliotheksmitarbeiterInnen übernommen werden.



### 3.2.9. Gibt es zum Eintragen der Metadaten Hilfestellung (z.B. Hilfstexte, wie die Metadatenfelder auszufüllen sind)? Wo findet man diese?

Die überwiegende Mehrheit der Institutionen bietet Hilfestellungen an, und zwar entweder in Form von Hilfstexten, die direkt im System zu finden sind, oder als externe schriftliche Anleitungen. Viele Einrichtungen verfügen über beides.

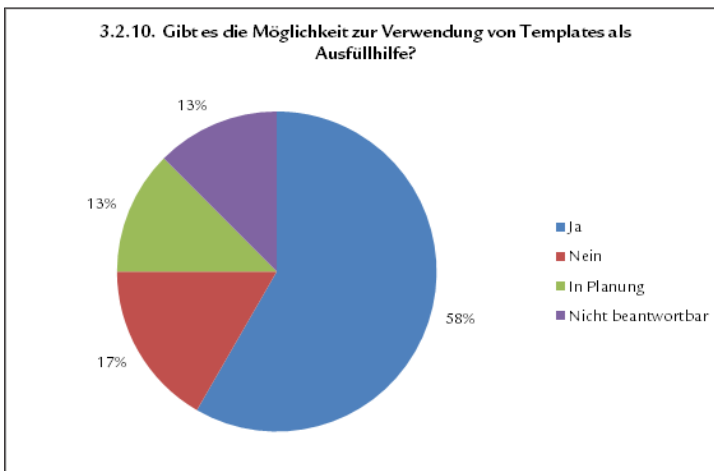
Vor allem bei Feldern, die über Autor, Titel, Datum und Beschreibung hinausgehen, ist es sinnvoll, Hilfstexte anzugeben, um den Usern eine Orientierung anzubieten. Im besten Fall sollten diese Hilfestellungen in mehreren Sprachen angeboten werden und für alle Personen, auch für jene, die über kein Hintergrundwissen verfügen, verständlich sein. Werden nur Standards und keine zusätzlichen Felder angeboten, sollte das Ausfüllen der Felder mehr oder weniger intuitiv erfolgen können.



### 3.2.10. Gibt es die Möglichkeit zur Verwendung von Templates als Ausfüllhilfe?

Das Anbieten von Templates ist ebenfalls überwiegend bereits realisiert bzw. in Planung.

Templates bieten zwei große Vorteile: Erstens kann man den Upload von ähnlichen Objekten wesentlich beschleunigen, wenn man nur diejenigen Eingaben ändern muss, die beim betreffenden Objekt abweichen und zweitens vermeidet man Fehler bzw. Inkonsistenz. Werden zum Beispiel im Rahmen eines Forschungsprojekts von mehreren Personen Daten hochgeladen und beschrieben, hilft ein gemeinsames Template, das gewünschte Vokabular zu verwenden, die passenden Begriffe im Thesaurus oder bei den Stichworten auszuwählen und es verringert sich damit auch die Anzahl der Tippfehler. Besonders komfortabel ist es, wenn man im System die Templates mit anderen Personen teilen kann, damit wird zum Beispiel der Projektleitung die Erstellung und Weitergabe des gewünschten Templates an die MitarbeiterInnen ermöglicht.

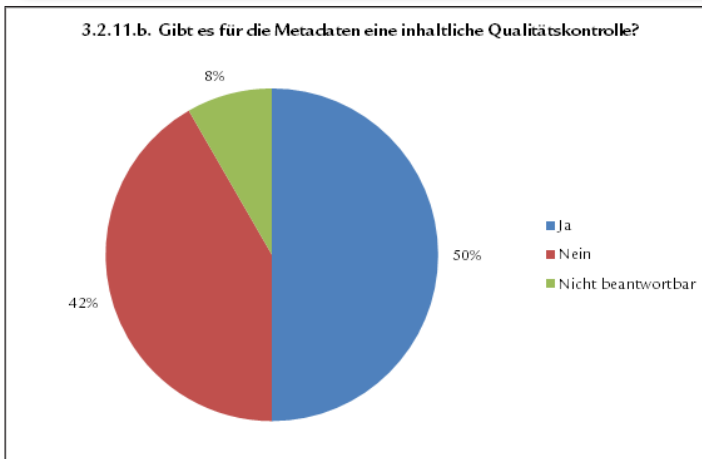
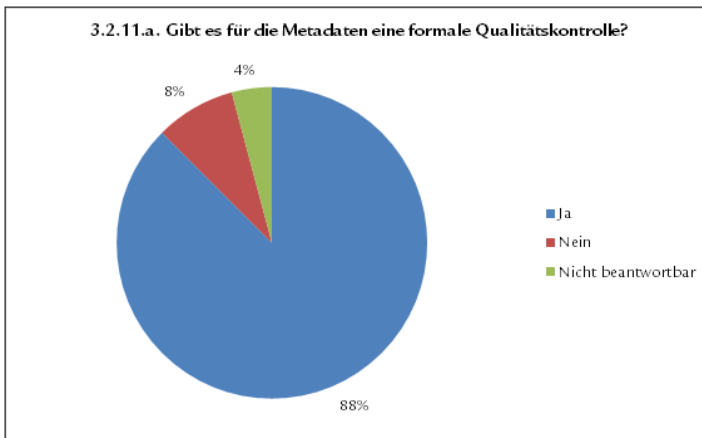


### 3.2.11. Gibt es für die Metadaten eine formale und / oder inhaltliche Qualitätskontrolle?

Nur wenige InterviewpartnerInnen gaben an, keine formale Qualitätskontrolle zu haben oder zu planen. Bei der inhaltlichen Qualitätskontrolle sieht es anders aus. Hier hat oder plant nur eine Minderheit der Institutionen keine oder nur eine sehr marginale Kontrolle. Viele konnten die Fra-

ge zu diesem Zeitpunkt aber auch noch nicht beantworten, da sie noch kein System hatten, bzw. es noch nicht im Regelbetrieb verankert war.

Die bevorzugte Methode zur formalen Kontrolle sind Pflichtfelder in der Eingabemaske, die bestimmte Mindestangaben bei der Datenerfassung erzwingen. Statt freier Textfelder werden – wenn möglich – vorgegebene Listen angeboten, die nur standardisierte Eingaben zulassen, z.B. bei der Verwendung von Thesauri (vgl. 3.2.12.). Weitere Kontrollmöglichkeiten: Datumsangabe per Kalender-Steuer-element; automatische Übernahme von Angaben zum Erfasser aus der Personaldatenbank (verknüpft über die Anmelde-daten – keine Datenerfassung ohne Login im Repositorium); intellektuelle Prüfung bibliografischer Daten auf RAK-Konformität durch Fachkräfte.



### 3.2.12. Verwenden Sie einen Thesaurus oder ein sonstiges kontrolliertes Vokabular? Wenn ja, welches?

Die überwiegende Mehrheit verwendet Thesauri oder ein kontrolliertes Vokabular. Die Grafik 3.2.12.b. zeigt, dass die Auswahl an Thesauri und Klassifikationen recht groß ist, einige nannten auch spezielle Thesauri, die hier unter „andere“ subsumiert sind. Die Nennung freier Schlagwörter wurde trotz der Frage nach kontrolliertem Vokabular im Diagramm berücksichtigt, um diesen möglicherweise auch interessanten Aspekt nicht auszuklamern.

Die im österreichischen Bibliothekenverbund weit verbreitete Basisklassifikation (BK) steht an der Spitze. Überhaupt prägen Thesauri aus dem Bereich der traditionellen bibliografischen Kataloge (= ALEPH) die Antworten sehr stark, wie die weitere Nennung von Dewey (Dezimalklassifikation) und RSWK / Gemeinsamer Normdatei (GND) zeigt. Unter den nicht-universalen, fachspezifischen Klassifikationen ist nur der Getty-Thesaurus (Art & Architecture, Geographic Names) einigermaßen weit verbreitet.

### 3.2.13. Wie viele Felder für die Metadaten sind vorgesehen?

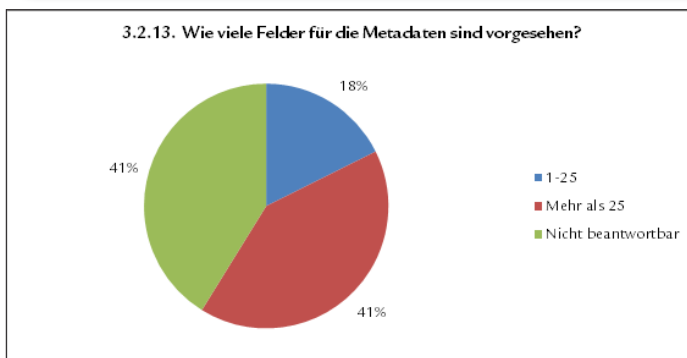
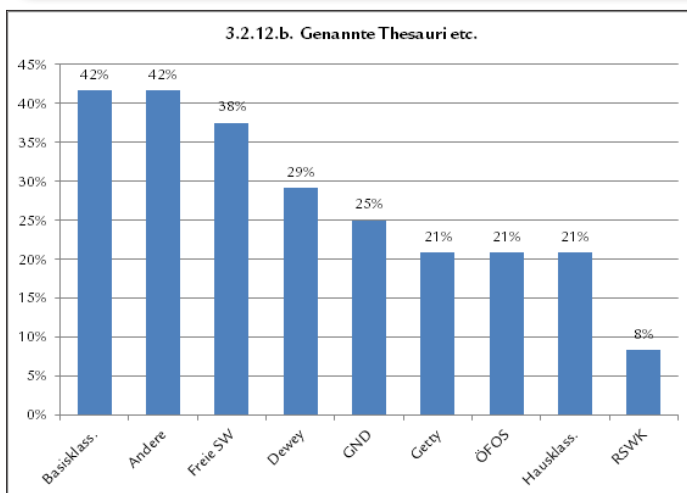
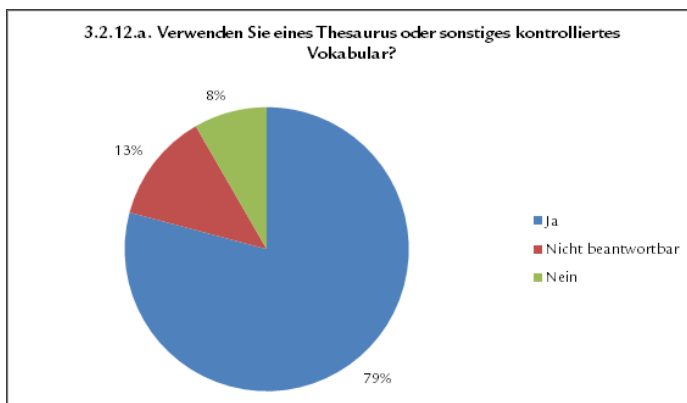
Hier ist ein leichter Trend zu mehr als 25 Metadatenfeldern zu erkennen. Einige InterviewpartnerInnen ergänzten jedoch, dass meist nicht alle dieser Felder verwendet werden. Der hohe Anteil von „nicht beantwortbar“ ergibt sich vor allem aus der Tatsache, dass einige der befragten Einrichtungen mehrere Repositorien (oder auch keines) betreiben und daher eine eindeutige Antwort nicht möglich ist.

### 3.2.14. Werden zurzeit auch Daten ohne Metadaten gespeichert?

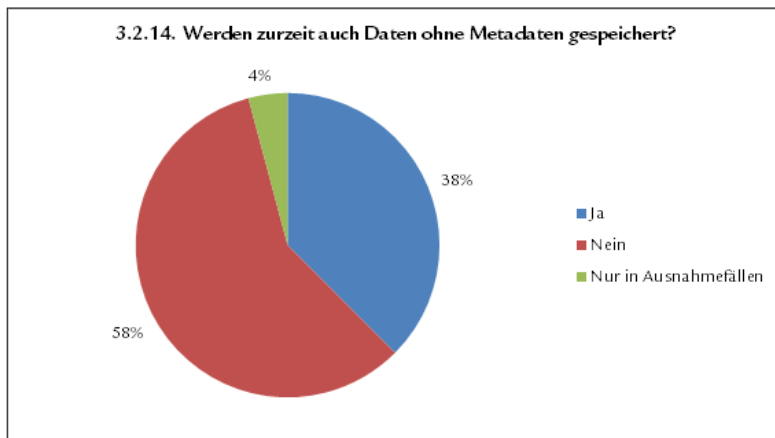
Daten ohne Metadaten werden zumeist außerhalb der offiziellen Repositorien (d.h. lokal auf Institutsebene etc.) gespeichert; die meisten Repositorien selbst lassen dies gar nicht zu. Die Antwort „Nein“ bedeutet i.d.R., dass über Praktiken der Datenerfassung außerhalb der Repositorien keine genauen Informationen vorliegen. Ziel ist in den allermeisten Fällen die Erfassung von Metadaten, die das Auffinden der beschriebenen Objekte erleichtern.

Denn, so wie Jeffrey Pomerantz schreibt: „Metadata is a map, Metadata is a means by which complexity of an object is represented in a simpler form.“<sup>6</sup>





Die Metadaten sollten auch schon vor dem geplanten Upload in ein Repository vorliegen, denn erfahrungsgemäß ist das nachträgliche Erstellen von Metadaten nicht nur sehr zeitaufwändig, sondern auch mit hohen Kosten verbunden. In vielen Fällen ist es auch schwierig, Metadaten nachträglich zu rekonstruieren, wenn beispielsweise nach Projektende die MitarbeiterInnen nicht mehr greifbar sind.



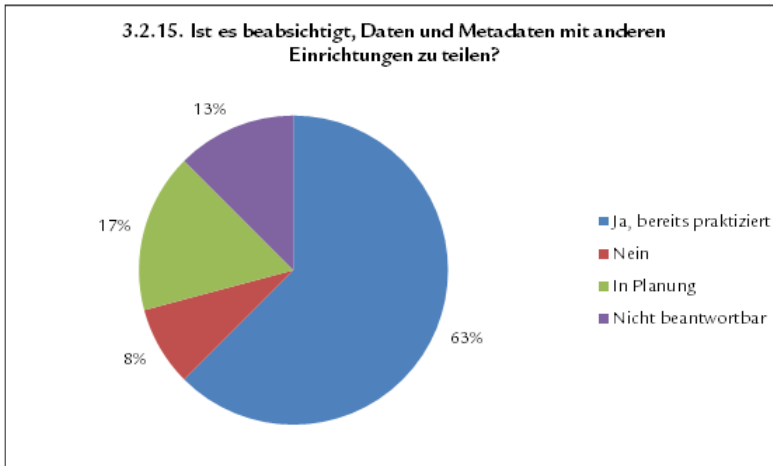
### 3.2.15. Ist es beabsichtigt, Daten und Metadaten mit anderen Einrichtungen zu teilen?

Eine deutliche Mehrheit der Befragten gibt an, die Daten und Metadaten mit anderen Einrichtungen teilen zu wollen und dies auch bereits zu praktizieren. Besonders für jene Einrichtungen, die sich dem Visual Library-Verbund angeschlossen haben, ist der Austauschgedanke ein wichtiger und wurde mehrfach als Grund für die Auswahl dieser Repositoryssoftware genannt.

Die konkreten Formen der Zusammenarbeit sind vielfältig: Datenlieferung an andere Repositorien und Datenbanken wie Europeana, Genios, UFORDAT oder arXiv; Öffnung des Repositoriums für das Harvesting (z.B. durch die Bielefeld Academic Search Engine [BASE]) über eine OAI-Schnittstelle; Lieferung von Forschungsdaten aus geförderten Projekten an den Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Ein Thema für die nahe Zukunft ist der Ausbau von Linked Open Data-Schnittstellen.

In einem Fall wurden Probleme in urheber- und datenschutzrechtlichen Fragen angesprochen, die dem eigentlich gewünschten Umfang des Daten-

austausches Grenzen setzen. Hier wird die Rechtslage als insgesamt zu wenig transparent beurteilt und es besteht Klärungsbedarf bezüglich dessen, was erlaubt ist und was nicht.

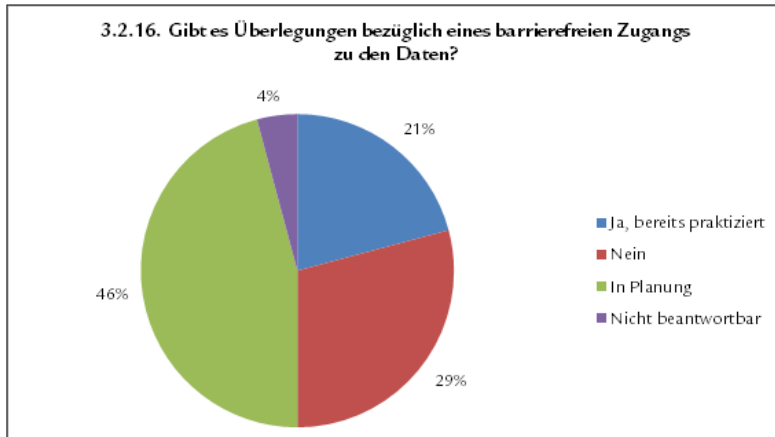


### 3.2.16. Gibt es Überlegungen bezüglich eines barrierefreien Zugangs zu den Daten?

Bemerkenswert ist, dass erst wenige Institutionen diese Frage bejahen konnten, was vorwiegend an den derzeit verwendeten, dafür nicht ausgelegten Systemen liegt. Die überwiegende Mehrheit strebt jedoch nach Barrierefreiheit und es gibt mehrere konkrete Projekte zur Umsetzung dieses Vorhabens.

Die barrierefreie Zurverfügungstellung von Daten in Informationsvermittlungssystemen wird mittlerweile auch im österreichischen Bundesbehindertengleichstellungsgesetz (BGStG) thematisiert<sup>7</sup>. Dennoch kann bisher kaum ein Repository wirklich als barrierefrei gelten. Die Metadaten können hier in zweifacher Weise helfen, die Barrierefreiheit zu unterstützen. Einerseits dienen Beschreibungen dazu, z.B. Bilder für blinde oder sehbehinderte Personen zu erklären und erfahrbar zu machen, bzw. Videos mit Gebärdensprache können gehörlosen Personen, die die Gebärdensprache als Muttersprache erlernt haben und Deutsch als Fremdsprache erlernen mussten, Unterstützung beim Verstehen der Daten geben. Andererseits können aber auch Hilfstexte, die die Metadatenfelder erklären, in einfacher Sprache verfasst werden, sodass nicht nur Personen mit Lern-

schwierigkeiten, sondern auch Menschen, die eine andere Muttersprache haben, den Sinn besser erfassen können.<sup>8</sup>



Mag.<sup>a</sup> Dr.<sup>in</sup> Susanne Blumesberger, MSc  
ORCID: <http://orcid.org/0000-0001-9018-623X>  
Universität Wien, Bibliotheks- und Archivwesen  
E-Mail: [susanne.blumesberger@univie.ac.at](mailto:susanne.blumesberger@univie.ac.at)

Dr. Alexander Zartl  
Universität Wien, Bibliotheks- und Archivwesen  
E-Mail: [alexander.zartl@univie.ac.at](mailto:alexander.zartl@univie.ac.at)

- 1 Bauer, Bruno et al.: Forschende und ihre Daten. Ergebnisse einer österreichweiten Befragung – Report 2015. Version 1.2. DOI: <https://doi.org/10.5281/zenodo.32043>
- 2 <http://www.bath.ac.uk/research/data/policy/research-data-policy-guidance.html#policy-guidance-1> (letzter Aufruf am 31.10.2016)
- 3 <https://creativecommons.org/publicdomain/zero/1.0/deed.de>
- 4 [http://www.aalto.fi/en/midcom-serveattachmentguid-1e6480be-8be5bd0480b11e693ae8b27aad222542254/2016\\_02\\_10\\_datapolicy.pdf](http://www.aalto.fi/en/midcom-serveattachmentguid-1e6480be-8be5bd0480b11e693ae8b27aad222542254/2016_02_10_datapolicy.pdf) (letzter Aufruf am 31.10.2016)
- 5 An dieser Stelle vielen Dank an Imola Dora Riehle-Traub, die sich im Rahmen des H2020 Projekts LEARN (LEaders Activating Research Networks) intensiv mit Policies auseinandersetzte.

- 6 Pomerantz, Jeffrey: Metadata. Cambridge, London: The MIT Press 2015, p. 12.
- 7 Vgl. Bundesbehindertengleichstellungsgesetz (BGStG) § 6 Abs. 5: „Barrierefrei sind bauliche und sonstige Anlagen, Verkehrsmittel, technische Gebrauchsgegenstände, Systeme der Informationsverarbeitung sowie andere gestaltete Lebensbereiche, wenn sie für Menschen mit Behinderungen in der allgemein üblichen Weise, ohne besondere Erschwernis und grundsätzlich ohne fremde Hilfe zugänglich und nutzbar sind.“
- 8 Siehe auch: Blumesberger, Susanne; Andreas Jaitler: Metadata and Accessibility. <http://phaidra.univie.ac.at/o:459817>

## ■ DER SPRINGER COMPACT-DEAL – EIN ERSTER EINBLICK IN DIE EVALUIERUNG EINER OFFSETTING-VEREINBARUNG

von Magdalena Andrae & Márton Villányi

**Zusammenfassung:** Seit Anfang 2016 gilt zwischen 32 wissenschaftlichen Bibliotheken in Österreich und dem Springer-Verlag eine gänzlich neuartige Vereinbarung, die neben der üblichen Lizenzierung von Inhalten auch die Möglichkeit bietet, in den Springer-Zeitschriften Open Access zu publizieren. Mehr als 1.000 Autorinnen und Autoren aus Österreich haben im ersten Jahr diese Möglichkeit genutzt. Die Daten dazu sind nun auf der Plattform OpenAPC einzusehen.<sup>1</sup> Ein guter Anlass, eine Zwischenbilanz zur Evaluierung des Deals zu ziehen.

**Schlüsselwörter:** Österreich; OpenAPC; Publikationskosten; Konsortialvertrag; Transformation zu Open Access

## THE SPRINGER COMPACT DEAL – FIRST INSIGHTS INTO THE EVALUATION OF AN OFFSETTING AGREEMENT

**Abstract:** On January the 1<sup>st</sup>, 2016 a new agreement between 32 Austrian scientific libraries and the publisher Springer took its effect: this deal covers accessing the licensed content on the one hand, and publishing open access on the other hand. More than 1000 papers by Austrian authors were published open access at Springer in the first year alone. The working group "Springer Compact Evaluierung" made the data for these articles available via the platform OpenAPC and would like to use this opportunity to give a short account of what this publishing agreement actually entails and the working group intends to do.

**Keywords:** Austria; OpenAPC; article processing charges (APC); consortial agreement; transition to open access



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## **Inhalt**

1. *Der konsortiale Springer-Compact-Deal und seine Evaluierung*
2. *Die OpenAPC-Initiative und ihr Nutzen für die Evaluierung*
3. *„Offsetting“ und „Read and Publish“ auf der Plattform OpenAPC*
4. *Die Evaluierung des österreichischen Deals – Ein Ausblick*

## **1. Der konsortiale Springer-Compact-Deal und seine Evaluierung**

Im Jahr 2015 vereinbarte die Kooperation E-Medien Österreich (KEMÖ) in Zusammenarbeit mit dem österreichischen Wissenschaftsfonds (FWF) ein neuartiges Lizenzmodell mit dem Springer Verlag. Aufgrund dieses „Read and Publish“-Deals erhalten Wissenschaftlerinnen und Wissenschaftler in den Jahren 2016 bis 2018 Zugriff auf über 2.000 Springer-Zeitschriften sowie gleichzeitig die Möglichkeit, in mehr als 1.600 Subskriptionszeitschriften des Verlags ohne Zusatzkosten Open Access zu publizieren.

Bibliotheken zahlen also dem Verlag eine Gesamtgebühr, die neben der Lizenzierung auch das Publizieren beinhaltet. Dieses neuartige Modell mit dem Namen Springer Compact soll ein weiterer wichtiger Schritt auf dem Weg zur vollständigen Open Access-Transition des wissenschaftlichen Publikationswesens werden.

Inwieweit Springer Compact die Erwartungen der Bibliotheken erfüllen wird, kann im Moment noch nicht beantwortet werden. Nach Ablauf des ersten Vertragsjahres 2016 wurde aber im Rahmen der KEMÖ eine Arbeitsgruppe auf die Beine gestellt, um wesentliche Faktoren wie Publikationsleistung oder Nutzung des Portfolios durch Autorinnen und Autoren genauer unter die Lupe zu nehmen. Die hier gewonnenen Erkenntnisse sollen Vor- und Nachteile eines solchen Deals aufzeigen und dem Konsortium Unterstützung bei zukünftigen Verhandlungen bieten. Ähnliche Evaluierungen wurden auch in anderen Ländern mit vergleichbaren Verträgen initiiert: Großbritannien, Schweden, die Niederlande und die MPDL (Max Planck Digital Library) analysieren ebenso ihre Springer-Compact-Vereinbarungen. Fragen, die im Zuge dessen beantwortet werden sollen, sind beispielsweise:

- Welche Auswirkungen hat der Deal auf das Publikationsaufkommen?
- Wie verändert sich die Nutzung der lizenzierten Inhalte?
- Taugt eine solche Vereinbarung als Motor zur Open Access-Wende?

Erste diesbezügliche Erkenntnisse sollen Anfang des Jahres 2018 vorliegen.

Für eine fundierte Analyse des österreichischen Publikationsaufkommens bedarf es jedoch einer konsistenten Datengrundlage. Um diese möglichst transparent und auch den Evaluierungsteams anderer Länder zugänglich zu machen, beschloss die österreichische Arbeitsgruppe in einem ersten Schritt dem Beispiel anderer Länder zu folgen und die bereinigten Publikationsdaten des Konsortiums über die Plattform OpenAPC zur Verfügung zu stellen.

## **2. Die OpenAPC-Initiative und ihr Nutzen für die Evaluierung**

Die Darstellung des Publikationsvolumens von Open Access-Deals ist nur ein kleiner, wenn auch nicht unwesentlicher Teil der OpenAPC-Plattform. Wie der Name verrät, ist die ursprüngliche Intention der Initiative, Daten zu APC-Zahlungen – also Open Access-Publikationsgebühren in Gold- und Hybrid Open-Access-Zeitschriften – zu sammeln. Eingebettet ist die Initiative in das von der deutschen Forschungsgemeinschaft (DFG) finanzierte Projekt INTACT, das von der UB Bielefeld, dem Institute for Interdisciplinary Studies of Science der Universität Bielefeld (I2SoS) und der Max Planck Digital Library (MPDL) getragen wird und dessen Ziel die Schaffung einer transparenten Infrastruktur für Open Access-Publikationskosten ist. Neben OpenAPC finden sich auch die Initiativen ESAC (Efficiency and Standards for Article Charges) und OA Analytics (Statistics on Open Access Publishing) im Rahmen dieses Projekts. Die Sammlung von empirischen Daten zu den Kosten von Open Access im Zuge von OpenAPC dient damit in einem breiteren Kontext der Unterstützung der Kostenanalyse sowie dem Ziel, standardisierte Berichts- und Geschäftsprozesse zu ermöglichen. Praktisch umgesetzt wird OpenAPC über Github.<sup>2</sup> Github Inc. ist eine Hosting-Plattform, die u.a. für die Software-Weiterentwicklung bei OpenSource-Projekten genutzt werden kann. Im Fall von OpenAPC ist durch Github für die einzelnen beteiligten Institutionen ein offener und transparenter Weg geschaffen worden, um Daten über die Ausgaben für Open Access zentral abliefern und damit in das Gesamtprojekt speisen zu können. Auch die Aufbereitung und Weiterverarbeitung der Daten erfolgt über Github.

Die Plattform bietet also einerseits Rohdaten zu Open Access-Publikationen, für die eine Gebühr entrichtet wurde; diese können heruntergeladen und in späterer Folge analysiert, angereichert und nachgenutzt werden. Andererseits werden die Daten auch direkt auf der Plattform grafisch aufbereitet; diese Visualisierungen geben einen guten ersten Einblick über Publikationsaufkommen und -ausgaben einzelner Einrichtungen, Verlage, Jahre etc.



### 3. „Offsetting“ und „Read and Publish“ auf der Plattform OpenAPC

Innerhalb der Plattform finden sich die Springer Compact-Daten mehrerer Länder im separaten Bereich „Offsetting“. Diese Abgrenzung war notwendig, da in diesem Fall für Publikationen keine einzelnen APCs mehr entrichtet werden, sondern in einem „Read and Publish“-Deal mitinbegriffen sind.

Im Open Access-Kontext decken die Begriffe „Offsetting“ und „Read and Publish“ ähnliche, wenn auch nicht idente Aspekte ab: Beim „Offsetting“-Modell versuchen Lizenznehmende (Bibliotheken, Konsorten) und Lizenzgebende (Verlage) eine doppelte Zahlung, das sog. das „Double dipping“, zu vermeiden. Dies kann auftreten, wenn erstens die Einrichtungen für ihre Zeitschriftenpakete eine Lizenzgebühr entrichten und zweitens Forschende derselben Einrichtung in diesen Zeitschriften gegen Bezahlung Open Access publizieren. Der Verlag kassiert also in solchen Fällen doppelt; einerseits von den Bibliotheken und andererseits von den Autorinnen und Autoren. Um dem entgegenzuwirken, sollen im jeweils folgenden Jahr diese Ausgaben den Lizenzgebühren gutgeschrieben werden.

Dies erfordert allerdings einen sehr hohen administrativen Aufwand, sodass das „Read and Publish“-Modell durchaus auch als Reaktion auf das mit dem „Offsetting“ verbundene umständliche Prozedere verstanden werden kann. Dabei zahlt die Einrichtung einmalig für beide Komponenten. Das hat zur Folge, dass sowohl der Zugriff auf das gesamte Zeitschriftenportfolio gewährleistet wird als auch die Forschenden ohne Mehrkosten und -aufwand Open Access publizieren können. Springer nimmt in diesem Zusammenhang eine Vorreiterrolle ein. Weitere Verlage, die dieses Open Access-Geschäftsmodell ebenfalls bereits verfolgen, sind beispielsweise Emerald und Wiley-Blackwell.

Übereinkünfte zur Gegenverrechnung von Open Access-Gebühren, die Gegenstand beider Modelle sind, stellen für die OpenAPC-Plattform eine Herausforderung dar, da die Bepreisung einzelner Artikel nicht mehr so eindeutig durchgeführt werden kann. Publikationen, die auf solchen Wegen Open Access publiziert wurden, sollten jedoch genauso erfasst werden – stellen doch diese Deals ein wichtiges Mittel zur Transformation hin zu Open Access dar. Nur so lässt sich ein breites Bild des Open Access-Marktes, mit all seinen Facetten, erstellen.

Zur Zeit sind in den „Offsetting“-Datensätzen keine Preisinformationen enthalten, da die entstandenen Kosten noch nicht auf die einzelnen Artikel heruntergebrochen werden können. Ident zu der restlichen Datenbank werden jedoch Verlag, Zeitschrift, DOIs und die Zugehörigkeit der korrespondierenden Autorinnen und Autoren angeführt.<sup>3</sup>

Zunächst beinhaltet der „Offsetting“-Bereich der OpenAPC-Plattform Daten zu Springer Compact-Publikationen aus Deutschland (Max Planck-Gesellschaft), Großbritannien, Niederlande, Österreich und Schweden. Während der Fertigstellung dieses Artikels wurden jedoch bereits die ersten Daten zu „Offsetting“-Deals mit Taylor & Francis und der Royal Society of Chemistry (RSC) durch die Max Planck-Gesellschaft hochgeladen<sup>4</sup>; weitere Datenlieferungen – z.B. durch die Kooperation E-Medien Österreich (KEMÖ) – könnten diesen Datenbestand in näherer Zukunft noch anwachsen lassen.

Da die Springer Compact-Daten im Unterschied zu den eigentlichen APC-Meldungen von institutsübergreifenden Organisationen wie Konsortien oder Arbeitsgruppen zur Verfügung gestellt wurden, ist auch von einer Vollständigkeit auszugehen. Es wird hier nicht ein Sample von kooperierenden Einrichtungen sichtbar, sondern der flächendeckende Publikationsoutput ganzer Konsortien. Dies ermöglicht in späterer Folge nicht nur einen Vergleich über die Jahre, sondern auch zwischen den Verlagen, und dies sowohl auf nationaler als auch internationaler Ebene. Die Attraktivität unterschiedlichster Deals wird dadurch einfacher quantifizierbar.

Über die Plattform lassen sich verschiedenste Aspekte darstellen, wie z.B.:

- Von welchen Einrichtungen kommen die Autorinnen und Autoren?
- In welchen Zeitschriften wird bevorzugt publiziert?
- Wie steht Österreich im internationalen Vergleich da?

Sobald diese Zahlen zu anderen Quellen in Relation gesetzt werden, können außerdem weitere wertvolle Erkenntnisse gewonnen werden.

Die Bereitstellung der Daten eines gesamten Konsortiums stellt für Österreich jedenfalls ein Novum dar. Auf Basis der einzelnen APC-Zahlungen ist die Initiative jedoch auf die Kooperation der einzelnen Einrichtungen angewiesen. Außer dem Springer-Konsortium wurden hierzulande bisher nur die Daten des österreichischen Wissenschaftsfonds (FWF)<sup>5</sup>, der Technischen Universität Wien<sup>6</sup> und des Institute of Science and Technology Austria<sup>7</sup> veröffentlicht. Um aussagekräftige Kostenanalysen und Prognosen durchführen zu können und eine offene Diskussion über alternative Publikationsmodelle anzustoßen, sollten noch viele weitere österreichische Einrichtungen deren Beispiel folgen.

#### **4. Die Evaluierung des österreichischen Deals – Ein Ausblick**

Die Offenlegung der Daten auf OpenAPC stand zunächst nicht im Fokus der Arbeitsgruppe „Evaluierung Springer Compact“. Im Rahmen der ersten

Treffen stellte sich jedoch rasch heraus, dass die Betrachtung des Vertrags in einem breiteren Kontext – konkret in Relation zu anderen Springer Compact-Verträgen – durchaus Sinn machen würde. Als daraufhin klar wurde, dass die AG nicht nur das vorliegende „Read and Publish“-Modell auswerten würde, sondern auch die Erarbeitung von Leitlinien für die Bewertung weiterer Open Access-Verlagsvereinbarungen Ziel sein sollte, wurde auch die Teilnahme an OpenAPC beschlossen.

Die eigentliche Aufgabe der Arbeitsgruppe besteht in der Formulierung von aussagekräftigen Evaluierungskriterien einerseits und der Auswertung der Daten aus dem Jahr 2016 andererseits.

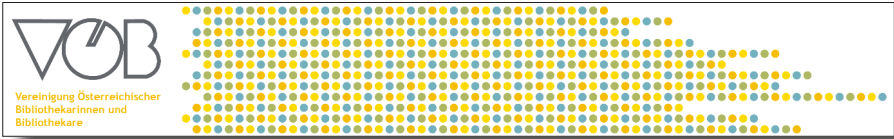
Bei der Formulierung besagter Kriterien konnte zum Teil auf die Vorarbeit anderer Länder zurückgegriffen werden. Spezifische österreichische Bedingungen mussten jedoch berücksichtigt werden. Außerdem sollten die Untersuchungsmethoden auch in späteren Jahren möglichst nachnutzbar sein. Soweit möglich war auch die Übertragbarkeit des Kriterienkatalogs auf zukünftige Deals erstrebenswert.

In einem weiteren Schritt müssen die vorhandenen Daten über das Publikationsaufkommen und/oder die Nutzung der Zeitschriften einer Kontrolle unterzogen werden, um die eigentliche Arbeit, nämlich die Analyse durchführen zu können. Dieser wird gegenwärtig in zwei spezialisierten Arbeitsgruppen nachgegangen. Erste Teilergebnisse werden beim Österreichischen Bibliothekartag im September 2017 präsentiert. Mit einem umfangreichen Bericht ist mit Anfang 2018 zu rechnen. Open Access-Deals zwischen Bibliotheken und Verlagen werden in den nächsten Jahren vermutlich statt der Ausnahme zum Standard werden, und so ist Arbeitsgruppe zur Evaluierung von Springer Compact zuversichtlich, mit ihren Analysen eine Hilfestellung bei dieser Entwicklung bieten zu können.

MMag.<sup>a</sup> Magdalena Andrae  
Technische Universität Wien, Universitätsbibliothek  
E-Mail: [magdalena.andrae@tuwien.ac.at](mailto:magdalena.andrae@tuwien.ac.at)

Mag. Márton Villányi  
Institute of Science and Technology Austria (IST), Library  
E-Mail: [marton.villanyi@ist.ac.at](mailto:marton.villanyi@ist.ac.at)

- 1 <https://treemaps.intact-project.org/> [Stand 27.07.2017]
- 2 <https://github.com/OpenAPC/openapc-de> [Stand 27.07.2017]
- 3 <https://treemaps.intact-project.org/apcdata/offsetting/> [Stand 27.07.2017]
- 4 Einsehbar unter: <https://github.com/OpenAPC/openapc-de/tree/master/data/offsetting/mpg> [Stand 27.07.2017]
- 5 <https://treemaps.intact-project.org/apcdata/fwf/> [Stand 27.07.2017]
- 6 <https://treemaps.intact-project.org/apcdata/tu-wien/> [Stand 27.07.2017]
- 7 <https://treemaps.intact-project.org/apcdata/ist-austria/> [Stand 27.07.2017]



## ■ AUS PRÄSIDIUM UND VORSTAND

Im Mittelpunkt der jüngsten Zusammenkünfte der Vereinsgremien am 19. Juni 2017 in den Räumlichkeiten der Arbeiterkammer Wien stand jeweils ein erster zusammenfassender Bericht von Ute Weiner zu den Ergebnissen der im Auftrag der VÖB vom Gallup Institut durchgeführten Umfrage zu prekären und atypischen Dienstverhältnissen im Bibliotheksbereich. Es ist hier nicht der Platz, die gewonnenen Erkenntnisse über die vielfältigen Beschäftigungsvarianten in unserem Berufsstand zu veröffentlichen, zumal detaillierte Analysen noch ausstehen und eine umfassende Darstellung der Resultate und der daraus zu ziehenden Schlussfolgerungen im Rahmen des 33. Österreichischen Bibliothekartages im September in Linz erfolgen soll. Es kann jedoch schon festgehalten werden, dass mit einem Rücklauf von 616 vollständig ausgefüllten Fragebögen eine bemerkenswert hohe Beteiligung erzielt werden konnte. Diese über allen Erwartungen liegende Quote zeigt einerseits das große Interesse der KollegInnenschaft an berufspolitischen Fragestellungen und rechtfertigt andererseits den hohen finanziellen Aufwand, der für die professionelle Begleitung der Umfrage einschließlich der Ergebnisauswertung notwendig war. Den Mitgliedern der vom Vorstand vor eineinhalb Jahren eingesetzten Arbeitsgruppe „Prekäre Dienstverhältnisse im Bibliothekswesen unter Führung von Ute Weiner sei an dieser Stelle ein großer Dank für das gezeigte Engagement ausgesprochen, das auch dazu Anlass gegeben hat, erste Überlegungen anzustellen, ob und in welcher Form sich die Gruppe – vielleicht erweitert um neue InteressentInnen – auch in Zukunft berufspolitischen Themen widmen könnte. Ebenso gedankt sei auch der Arbeiterkammer Wien, die sich an den Kosten für die Umfrage mit einem ansehnlichen Betrag beteiligt hat.

Auch auf dem 106. Deutschen Bibliothekartag, der heuer von 28. Mai bis 1. Juli in Frankfurt am Main stattfand, zeigte sich, dass sich das Aufgabengebiet besonders in wissenschaftlichen Bibliotheken derzeit überaus dynamisch entwickelt, weil traditionelle Bereiche, vor allem was die Bear-

beitung und Bereitstellung physischer Medien betrifft, stetig schrumpfen, hingegen neue Tätigkeitsfelder, wie zum Beispiel Publikationsunterstützung, Forschungsdatenarchivierung oder Bibliometrie neu hinzukommen. Um diesen Wandel in den Bibliotheken positiv gestalten zu können, erscheint es unerlässlich, rechtzeitig entsprechende Personalentwicklungsmaßnahmen zu setzen, getreu dem Motto „Wer nicht mit der Zeit geht, geht mit der Zeit.“ In diesem Zusammenhang ist es sehr erfreulich, dass in Frankfurt auch das erweiterte Kooperationsabkommen der Bibliotheksverbände in den deutschsprachigen Ländern offiziell vorgestellt wurde.



Abb. 1: Jean-Marie Reding, ALBAD; Christian Jahl, BVÖ; Gerlinde Schmiedhofer, BVS; Sylvia Gladrow, BIB; Konstanze Söllner, VDB; Herbert Staub, BIS; Kristina Lippold, BIB; Werner Schlacher, VÖB (von links nach rechts) (Copyright: VDB)

Der Vereinbarung, die bisher für die Mitglieder der Verbände in Deutschland, Österreich, der Schweiz und Südtirol (DACHS) Gültigkeit hatte, ist nun auch der luxemburgische Verband und der Büchereiverband Österreichs (BVÖ) beigetreten (DACHSL). Außerdem wurde das Übereinkommen, das es bisher schon den Verbandsmitgliedern aus den genannten Länder die Teilnahme an den Bibliothekartagen der Partnerorganisationen zum Mitgliedertarif gestattete, auf alle von den Verbänden durchgeführten Fortbildungsveranstaltungen ausgedehnt, womit eine Intensivierung des grenzüberschreitenden Gedankenaustauschs herbeigeführt werden soll.

Wie in leider so vielen Präsidiums- und Vorstandssitzungen war es auch diesmal unsere traurige Pflicht, der verstorbenen Mitglieder zu gedenken. Im besonderen Maße und in großer Dankbarkeit für ihre Leistungen für unsere Vereinigung, aber auch für das gesamte österreichische Bibliothekswesen galt unsere Trauer diesmal unserer langjährigen Kollegin und vielen von uns auch Freundin Dr. Eveline Pipp von der Universitäts- und Landesbibliothek Tirol, für die Sie einen Nachruf in dieser Ausgabe der Mitteilungen finden, auf den ich Ihre Aufmerksamkeit zum Schluss noch richten möchte.

Mit den besten Wünschen für einen erholsamen Sommer und auf ein Wiedersehen beim Bibliothekartag in Linz

Ihr Werner Schlacher  
(im Namen des Präsidiums und des Vorstandes der VÖB)



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## ■ „LIBRARIES.SOLIDARITY.SOCIETY“ – SCHRITTE ZU EINER INTENSIVIERUNG DER KOOPERATION ZWISCHEN DEN DEUTSCHSPRACHIGEN UND INTERNATIONALEN BIBLIOTHEKSVERBÄNDEN

*von Eva Ramminger, Magdalena Andrae und Andrea Torggler*

Der Trend zur Globalisierung ist für Bibliotheken nicht neu. So ist es bereits seit langem selbstverständlich, mit internationalen Standards zu arbeiten oder mit global agierenden Agenturen, Informations- und Systemanbietern zu verhandeln. Gleichzeitig wächst im Umfeld chronischer Ressourcenverknappung und steigender Medienpreise der Rechtfertigungsdruck für Bibliotheken und somit das Bedürfnis, die Anliegen der Bibliotheken nachdrücklicher in Politik und Gesellschaft zu verankern.

Vor diesem Hintergrund fand am 22. und 23. März ein gemeinsamer Informationsaustausch von Vertreterinnen und Vertretern des Deutschen Bibliotheksverbands DBV, des Verbands Bibliothek Information Schweiz BIS, des Büchereiverbands Österreich und der VÖB statt. Gastgeber dieses Treffens war Mag. Gerald Leitner, ehemaliger Geschäftsführer des BVÖ und seit 2016 Generalsekretär der IFLA. Tagungsort war das IFLA-Generalsekretariat in der Königlichen Bibliothek in Den Haag.<sup>1</sup>



Abb. 1: Tagungsort des Treffens der deutschsprachigen Bibliotheksverbände in der Königlichen Bibliothek in Den Haag (Foto: Eva Ramminger)

Ziel dieses Treffens war es, darüber nachzudenken, welche Möglichkeiten des Lobbyings auf internationaler Ebene genutzt werden können bzw. welches



Potential dabei in einer engeren Abstimmung der einzelnen Aktivitäten im deutschen Sprachraum aber auch mit den Dachverbänden IFLA, EBLIDA und LIBER liegt. Dies ist umso wichtiger als auf Seiten der Verlage sehr professionell mit den gleichen Methoden gearbeitet wird. Um zu diesen Entwicklungen ein entsprechendes Gegengewicht aufzubauen, ist ein verstärktes Engagement insbesondere auf europäischer Ebene gefragt. So braucht es bei den gegenwärtig laufenden Verhandlungen um adäquate (Ausnahme-)Regelungen im Urheberrecht bzw. Copyright oder beim eLending stark verankerte Interessengruppen, um die Durchsetzung bibliothekarischer Anliegen voranzubringen.

Anhand von Best Practice-Beispielen, wie bspw. der internationalen Arbeit bei den deutschen Nachbarverbänden BID und DBV, wurde die Relevanz und Übertragbarkeit dieser Aktivitäten auf den jeweils eigenen Verband – konkret: auch die VÖB – diskutiert. Ganz nebenbei gelang es, praxisbezogene Problemstellungen zu lösen. So wird beispielsweise der Ausbau eines zunächst auf freiwilliger Basis beruhenden und nun professionell weitergeführten Übersetzungsservice vorangetrieben und in Zukunft von den deutschsprachigen Verbänden gemeinsam finanziert. Generell zeigte sich sehr rasch, dass es sinnvoll ist, diese Treffen auf D-A-CH-Ebene in Zukunft entsprechend weiterzuführen.

Vom 19. bis 25. August fand nun die IFLA World Library and Information Congress 2017 (<http://2017.ifla.org>) im polnischen Wrocław/Breslau statt.<sup>2</sup> Unter dem bereits im Titel dieses Beitrags zitierten Generalthema „Libraries.Solidarity.Society“ fanden sich über 3.500 TeilnehmerInnen aus 120 Nationen ein, um sich über aktuelle bibliothekarische Themenstellungen auszutauschen. Einige Highlights daraus:

Bereits am Vorabend zur offiziellen Eröffnung fanden Netzwerktreffen für die teilnehmenden Sprachgruppen bzw. Nationen statt, an dem über Neuigkeiten aus den Ländern berichtet wurde. So wurde hier u.a. eine geplante Fusion der beiden Fachverbände Bibliothek Information Schweiz (BIS) und der Schweizerischen Arbeitsgemeinschaft der öffentlichen Bibliotheken (SAB) zu einem neuen Verband „Bibliosuisse“ bekanntgegeben. Der BVÖ konnte auf erfolgreiche Verhandlungen mit der American Library Association ALA verweisen, wo ab dem kommenden Jahr Stipendien für Kurzpraktika an amerikanischen Bibliotheken vergeben werden können. Außerdem berichtete er von einer Kollegin, die bereits heuer in den schottischen Highlands in einem Bücherbus mitarbeiten durfte.<sup>3</sup>

In der Eröffnungsveranstaltung brachte dann IFLA-Präsidentin Donna Scheeder die Zielsetzung des Kongresses auf den Punkt: „Times are changing – we have to change, too“. Und dies sei eben nur gemeinsam möglich. Nach den Festansprachen der VertreterInnen der Politik und des Biblio-

thekswesens des Gastgeberlandes wurde eine – in ILFA-Kreisen ob ihrer Extravaganz bereits im Vorfeld gerühmte – Showeinlage geboten: gesanglich und mit Nebelmaschine, Feuer und Videomaterial untermalt wurde die Geschichte Polens von AkrobatInnen dargestellt.

Der Aufruf zur Partizipation erfolgte im Verlauf des Kongresses wie auch in der Berichterstattung über die Sozialen Medien gleich mehrfach.<sup>4</sup> Man solle liken, taggen, kommentieren und vor allem teilen – teilen, um eine digitale, globale Bibliotheksgemeinschaft zu bilden. So konnten heuer auch erstmals alle wichtigen Veranstaltungen per Livestream mitverfolgt werden. Die Themen Wandel und Vernetzung zogen sich durch die Sessions und wurden fast zum Mantra.

Mehrfach berichtete auch die IFLA selbst zu ihren Aktivitäten: Im Rahmen des Global Vision Reports 2017 soll erstmals ein Gesamtbild aller Mitglieder für eine zukünftige Strategie der IFLA entstehen. Dazu gipfelte ein über mehrere Monate und auf mehreren Kontinenten geführter Diskussionsprozess nun in einem Online-Voting, das während der Tagung gestartet wurde. Ziel ist es, ein umfassendes Stimmungsbild zu den zukünftigen Schwerpunkten der IFLA-Verbandsarbeit zu erhalten. Jede/Jeder ist aufgerufen, daran teilzunehmen. Die Umfrage läuft noch bis 30.09.2017: <https://globalvision.ifla.org/vote/>.



Abb. 2: VÖB-Präsident Werner Schlacher, Jean-Marie Reding (Vizepräsident der Luxemburgischen Bibliotheksvereinigung ALBAD), Herbert Staub (Präsident des Verbandes Bibliothek Information Schweiz, BIS), Patrick Danowski (IST Austria) und weitere TeilnehmerInnen an einer der vielen Strategiediskussionen (Foto: Eva Ramminger)

Naturgemäß ist die IFLA als Dachorganisation stark daran interessiert, Bibliotheken und ihre Vernetzung nach außen hin sichtbar zu machen. So wurde während des Kongresses auch das Projekt „Library Map of the World“ (<https://librarymap.ifla.org>) gelauncht, welches durch die Erarbeitung einer Datenbasis und die Erhebung statistischer Daten die Verbreitung von Bibliotheken weltweit aufzeigen soll. Auch diese Ergebnisse dienen letztlich dazu, nachdrücklicher und erfolgreicher auf politischer Ebene agieren zu können. In diesem Zusammenhang fand auch die Verlinkung zu den Zielen ihres International Advocacy Programme statt, wo die Positionierung der Bibliotheken im Kontext der UN 2030 Agenda für Nachhaltige Entwicklung im Mittelpunkt steht (<https://www.ifla.org/drupal/libraries-development>).

Zahlreiche weitere Beispiele der Vernetzung wurden in den Vorträgen, Poster-Sessions und Fachgruppensitzungen (Sections) vorgestellt: Beispielsweise das Projekt "Library Publishing Directory" (<https://www.librarypublishing.org/>), in welchem sich Bibliotheken zusammenschließen und Know-How im Bereich Publizieren austauschen, um damit Professionalität in der Zusammenarbeit mit den Fachverlagen zu gewinnen. Auch der Zusammenschluss zehn großer Universitätsbibliotheken in Illinois zu einer Art "Konsortium", welche gemeinsame Archivierungsstrategien, Erwerbungspolitik und ergänzenden Bestandsaufbau praktizieren, kann hier genannt werden (Lynn Norine Wiley, [www.hathitrust.org](http://www.hathitrust.org)).

Change Management und Knowledge Management wurden als ineinandergreifende Felder aufgegriffen und breit diskutiert. Die Themen wurden aus unterschiedlichen Blickwinkeln betrachtet, von praktischen Beispielen, wie partizipatorisch einschneidende Veränderungsprozesse umgesetzt werden können (Vantaa City Library, <http://library.ifla.org/id/eprint/1653> und Uppsala University Library, <http://library.ifla.org/id/eprint/1650>) bis hin zu Überlegungen, welche Rolle Unwissenheit im Knowledge Management spielt (Iain Watt: <http://library.ifla.org/id/eprint/1696>).

Im Rahmen eines Knowledge Cafés – ein für eine internationale Veranstaltung äußerst passendes Format – galt es dann auch für alle TeilnehmerInnen in kleiner Runde zentrale Schlagwörter zu diskutieren und Ideen auszutauschen: Wie hält man Wissen aktuell? Wie kann Wissen intern geteilt werden? Was sind Zutaten einer innovativen Bibliothekskultur? Als Grundtenor zog sich die Idee durch, sich mit spielerischen Ideen neue Zielgruppen zu erschließen (wie etwa durch den Einsatz von Tablets und Robotern) und frei von bisherigen Abläufen konkrete neue Arbeitsfelder (Stichwort: Library Publishing) einzuführen. Wichtig dabei ist die Einbindung des Personals, das direktes Kundenfeedback erhält (bspw. Leihstelle, Informationsstelle). Nicht verschwiegen werden sollte, dass daneben aber

auch immer wieder angesprochen wurde, ob und wie diese neuen Tätigkeiten – vor allem in Zeiten der Ressourcenverknappung – mit dem Aufgeben bisheriger Arbeitsfelder einhergehen könnte.

Einige Vorträge konzentrierten sich auf das Gewinnen neuer Zielgruppen für Bibliotheken. Projekte im Bereich Gamification und Coding z.B. wurden in vielen Sessions als zukunftsweisende Strategien besprochen, um gerade jungen Menschen die Bibliothek und ihre Angebote näher zu bringen (Vantaa City Library: <http://kirjasto.one/pelittaa/hello-gamer/>). Der Umgang mit Digitalisierung, Daten und das Coden selbst findet als neue Form der Informationskompetenz bzw. als Sprache der Zukunft immer mehr Eingang in den Bibliotheksbereich (z.B. Biblioteka Murow: <https://www.youtube.com/watch?v=yDTV481oXcw>). Hier wird erheblicher Kompetenzaufbau in Bibliotheken gefragt sein, um diese neue „Sprache“ auch zu verstehen und vermitteln zu können, was in Polen bereits mit einem EU-geförderten Projekt unterstützt wird (Agnieszka Koszowska: <http://frsi.org.pl/en/>).

Der/Die UserIn steht aber nicht für sich alleine, er/sie ist Teil einer Community. Und diese Community hat das Potenzial sich im Zeitalter der Digitalisierung zu einer ganz neuen Form der "citizenship" zu entwickeln, welche Probleme gemeinsam lösen können. Sehr schön sichtbar wurde dies im Europeana-Projekt "Transcribathon" (<https://transcribathon.com/>), das eine Plattform für das Transkribieren handschriftlicher Dokumente anbietet.

Prominent vertreten waren natürlich auch die Themen Text-, Data-Mining und Semantic Web, welche in den enthusiastischen Beiträgen von Peter Murray-Rust als unerlässliches Mittel in der Forschung und Wissensgesellschaft gefordert wurden. So werden in [www.wikidata.org](http://www.wikidata.org) strukturierte Hintergrunddaten der Wikipedia miteinander verlinkt, um daraus im Sinne des Semantic Web neue, maschinenlesbare Klassifikationen und Vokabulare zu bilden. Mit gemeinsamen offenen Daten können durch einfache Skripts völlig neue Erkenntnisse in sehr kurzer Zeit geschaffen werden – alles, was es dazu braucht ist ein Laptop und Zugang zu wissenschaftlicher Literatur.

In den Ansprachen der scheidenden IFLA-Präsidentin Donna Scheeder und ihrer Nachfolgerin Glòria Pérez-Salmerón in der Abschlussveranstaltung wurde nochmals nachdrücklich auf die Wichtigkeit der Zusammenarbeit zwischen Bibliotheken, Öffentlichkeit und Entscheidungsträgern betont. Nicht passiv, sondern aktiv sollte das zukünftige Tätigkeitsfeld der Bibliotheken mitgeformt werden. Denn, um es in den Worten von Donna Scheeder zu formulieren: "Libraries are the motor of change – and librarians are the gears of the motors of change ... Call everyone to action!"

Ein Hinweis: Im Rahmen des 33. Österreichischen Bibliothekartages in Linz findet am Donnerstag, 14.09.2017, um 14:00 Uhr eine Podiumsdiskussion mit VertreterInnen der deutschsprachigen Bibliotheksverbände, der neuen IFLA-Präsidentin Glòria Pérez-Salmerón sowie dem IFLA-Generalsekretär Gerald Leitner unter dem Thema „Gemeinsam statt einsam – Zur Rolle von Bibliotheksverbänden in einem globalisierten Umfeld“ statt: <http://bibliothekartag2017.univie.ac.at/home/>.

Mag. Eva Ramminger  
Universität Innsbruck, Universitäts- und Landesbibliothek Tirol  
E-Mail: [eva.ramminger@uibk.ac.at](mailto:eva.ramminger@uibk.ac.at)

MMag.<sup>a</sup> Magdalena Andrae  
Technische Universität Wien, Universitätsbibliothek  
E-Mail: [magdalena.andrae@tuwien.ac.at](mailto:magdalena.andrae@tuwien.ac.at)

Mag.<sup>a</sup> Andrea Torggler  
FH Campus Wien, Bibliothek und Mediathek  
E-Mail: [andrea.torggler@fh-campuswien.ac.at](mailto:andrea.torggler@fh-campuswien.ac.at)

- 1 Der Bericht von der Sitzung in Den Haag baut auf dem von Judith Oliiva (BVÖ) verfassten Sitzungsprotokoll auf. Ihr sei an dieser Stelle sehr herzlich für die Möglichkeit der Nutzung des Textes gedankt.
- 2 Ein Großteil der Beiträge kann bereits unter: <http://library.ifla.org/view/conferences/2017/> nachgelesen werden.
- 3 Ihr spannender Reisebericht ist im BVÖ-Blog nachzulesen unter: <http://blog.bvoe.at/blog/die-fahrt-den-norden>.
- 4 Siehe dazu auch die äußerst informativen Beiträge von Markus Jäger (Stadtbücherei Innsbruck) im BVÖ-Blog unter: <http://www.blog.bvoe.at/blog/librarysolidaritysociety-der-ifla-wlic-2017-und-die-globale-signifikanz-von-bibliotheken>.



Dieses Werk ist lizenziert unter einer  
[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## ■ DR. EVELINE PIPP (12.12.1956–05.05.2017): EIN NACHRUF

*von Helmut Hartmann, Eugen Rott, Klaus Niedermair und Eva Ramminger*



Mit dem Tod von Dr. Eveline Pipp verliert die Universitäts- und Landesbibliothek Tirol und die Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare eine jener Persönlichkeiten, die das österreichische Bibliothekswesen der letzten Jahre maßgeblich geprägt haben.

Geboren in Innsbruck, promovierte sie 1987 bei Prof. Dr. Walter Larcher am Institut für Botanik der Universität Innsbruck mit einer Arbeit zum Energiegehalt von Pflanzen „sub auspiciis praesidentis rei publicae“. Nach einigen erfolgreichen Projektjahren am Institut für Botanik, wechselte sie 1993 an die damalige Universitätsbibliothek Innsbruck in das Team von HR Dr. Heinz Hauße.<sup>1</sup> Es war die Zeit der ersten Literaturdatenbanken und der ersten automationsunterstützten Literaturrecherchen. Eveline Pipp erwarb sich in kürzester Zeit ein umfassendes Wissen über Informationsvermittlung durch elektronische Medien, also Datenbanken, E-Zeitschriften und E-Bücher, sowie Nutzungsstatistiken und die daraus zu ziehenden Schlüsse in Bezug auf Wirtschaftlichkeit und Nachhaltigkeit des Erwerbs der betreffenden Ressourcen. In den nachfolgenden Jahren verhandelte sie

die ersten sogenannten „Big Deals“ mit Zeitschriftenverlagen mit und bereitete den Boden für die Integration dieser neuen Angebote in das Portfolio der Universitätsbibliothek.

Sie war daher prädestiniert, im Dezember 2006 die Nachfolge von Dr. Heinz Hauffe in der Abteilungsleitung anzutreten und damit auch die ULB Tirol in der kurz zuvor gegründeten Kooperation E-Medien Österreich (KEMÖ) zu vertreten. Schon in der Vorbereitungsphase der KEMÖ leistete sie durch ihre präzise Analyse von Kostenaufteilungsmodellen für den konsortialen Erwerb grundlegende Aufbauarbeit. Auch wurde eine bis heute gültige paradigmatische Ankaufsentscheidung von ihr mitgeprägt: die Zustimmung der damaligen Mitglieder zu dem nach Durchrechnung extrem komplexer Varianten gefundenen Kostenschlüssel für eine der weltgrößten bibliographischen Datenbanken.

Ihrer besonderen Gabe, komplexe statistische Zusammenhänge mathematisch zu erfassen und allgemein verständlich darzustellen, war es denn auch zu verdanken, dass sie Anfang 2007 mit der Gründung des Arbeitskreises Kostenschlüssel betraut wurde, den sie bis zu ihrer Erkrankung leitete. Dieser spielt bis heute eine zentrale Rolle in der Organisation der KEMÖ, stellt er doch die Berechnungsgrundlagen für alle wichtigen Entscheidungen zur Verfügung, sei es nun die Aufstockung des Personalstandes der Geschäftsstelle, der Erwerb neuer Ressourcen oder die Berechnung der Beitragsgebühren der Mitglieder.

Als im Zuge der Adaption der Statuten im Jahr 2008 als Lenkungs- und Kontrollorgan der Geschäftsstelle erstmals ein Kooperationsausschuss von den VertreterInnen der Mitgliedseinrichtungen zu wählen war, wurde Eveline Pipp die Funktion der stellvertretenden Vorsitzenden zuteil, die sie nach neuerlicher Wahl bis 2010 ausübte, ehe sie 2011 zur Vorsitzenden gewählt wurde. In dieses Amt wurde sie in den nachfolgenden Jahren mehrfach wiedergewählt.

Als ob diese Aufgaben nicht genug gewesen wären, leitete sie zu verschiedenen Zeiten zusätzlich mehrere Arbeitsgruppen, die zentrale Entscheidungen zur weiteren Entwicklung der KEMÖ vorbereiteten – dies alles unentgeltlich neben ihrer hauptberuflichen Vollzeitbeschäftigung an der ULB Tirol, ihrer Vortragstätigkeit im In- und Ausland, ihrer Lehrtätigkeit im Universitätslehrgang Library and Information Studies und ihrer Mitarbeit im Präsidium und Vorstand der VÖB.

In letztgenanntem Gremium war Eveline Pipp seit 1995 tätig und setzte auch hier zahlreiche wichtige Aktivitäten. So engagierte sie sich insbesondere bei den regelmäßig stattfindenden „Online-Informationstreffen“, die später zur vielzitierten „ODOK“ ausgebaut wurden. Zunächst war sie

hier in der Herausgabe der Tagungsbände involviert, später organisierte sie diese Veranstaltung selbst. Von 2006–2007 war sie für die Herausgabe der „Online-Mitteilungen“ verantwortlich, von 2005 bis 2008 war sie Vorsitzende der VÖB-Arbeitsgruppe „Elektronische Medien“. Darüber hinaus war sie aktives Mitglied in mehreren Kommissionen und Arbeitsgruppen der VÖB.

Sie blieb auch ihrem Studienfach weiterhin verbunden. So arbeitete sie in der AG Hydrobotanik an Ihrem Institut nebenbei als freie Mitarbeiterin und engagierte sich in wissenschaftlichen Projekten. Mit großem persönlichen Einsatz widmete sie sich dabei vor allem der Konzeption von neuen Projekten, der Entwicklung von Optimierungsverfahren für Projektauswertungen sowie der Betreuung von Studierenden (GastwissenschaftlerInnen, Studierenden vom Grundstudium bis zur Dissertation/Habilitation). Die zahllosen, fruchtbaren wissenschaftlichen Diskussionen bei Seminaren und in Zwiegesprächen bewiesen immer wieder ihr klares synoptisches Denkvermögen und außergewöhnliches Interesse an wissenschaftlichen und angewandt-wissenschaftlichen Fragen der Hydrobotanik und Gewässerkunde. Sie hat ganz wesentlich an den europaweit (aber auch in Kanada) angewendeten Indikationssystemen insbesondere für Aufwuchsalgen (Phytobenthos) und an mathematischen Modellen in der Limnologie mitgewirkt.

Lange hoffte sie darauf, ihre geliebte Arbeit nach ihrer schweren Erkrankung im Herbst letzten Jahres wieder aufnehmen zu können. Doch nun ist Eveline Pipp dieser Krankheit leider erlegen. Sie wird uns als Kollegin in Erinnerung bleiben, die durch ihr kaum fassbares Engagement sowie ihre unglaubliche Kompetenz und Vielseitigkeit überzeugte. Gleichzeitig berührte sie uns durch ihre großzügige und keine Grenzen kennende Hilfsbereitschaft sowie ihre unerschütterlich positive Lebenseinstellung. Wir werden sie nicht vergessen!

## **Dissertation**

Pipp, Eveline (1987): Energiegehalte pflanzlicher Substanz, 2 Bände. Innsbruck, Univ., Diss.

## **Herausgeberinnenschaften**

Pipp, Eveline (Hg.) (1997): Online-Datenbanken – CD-ROM-Datenbanken. Tagungsberichte vom 4. und 5. Österreichischen Online-Informationstreffen, 17.–20. September 1991, Schloss Hofen, Lochau



- bei Bregenz, 14.–17. September 1993, Schloss Seggau, Seggau bei Leibnitz; [Trends und Qualitätskriterien]. Innsbruck: Österr. Online-Benutzergruppe (Biblos-Schriften, 158).
- Pipp, Eveline (Hg.) (2000): Dokumente und Datenbanken in elektronischen Netzen. Dokumente und Datenbanken in elektronischen Netzen: Tagungsberichte vom 6. und 7. Österreichischen Online-Informationstreffen bzw. vom 7. und 8. Österreichischen Dokumentartag, Schloss Seggau, Seggau bei Leibnitz, 26.–29. September 1995, Congresszentrum Igls bei Innsbruck, 21.–24. Oktober 1997. Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare; Österreichisches Online-Informationstreffen; Österreichischer Dokumentartag. Innsbruck: Österr. Online-Benutzergruppe (Biblos-Schriften, 161).
- Pipp, Eveline (Hg.) (2002): Drehscheibe E-Mitteuropa. Information Produzenten, Vermittler, Nutzer, die gemeinsame Zukunft; Tagungsberichte vom 9. Österreichischen Online-Informationstreffen und 10. Österreichischen Dokumentartag; 24.–27. April 2001, Universität Graz. Wien: Phoibos (Biblos-Schriften, 173).
- Pipp, Eveline (Hg.) (2004): Ein Jahrzehnt World Wide Web: Rückblick – Standortbestimmung – Ausblick. ODOK '03; Tagungsbericht vom 10. Österreichischen Online-Informationstreffen und 11. Österreichischen Dokumentartag; 23.–26. September 2003, Universität Salzburg, Naturwissenschaftliche Fakultät. Wien: Phoibos (Biblos-Schriften, 179).
- Pipp, Eveline (Hg.) (2006–2007): Redaktion der „Online Mitteilungen“ Dezember 2006 bis November 2007.
- Pipp, Eveline (Hg.) (2007): Zugang zum Fachwissen. ODOK '05, 11. Österreichisches Online-Informationstreffen, 12. Österreichischer Dokumentartag, 13.–16. September 2005, Freie Universität Bozen. ODOK 2005. Bozen, 13.–16. September 2005. ODOK; Universität Bozen. Graz, Feldkirch: Neugebauer (Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare: Schriften der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 1).
- Pipp, Eveline (Hg.) (2008): Informationskonzepte für die Zukunft. ODOK '07; 12. Österreichisches Online-Informationstreffen; 13. Österreichischer Dokumentartag; 17.–21. September 2007, Karl-Franzens-Universität Graz. ODOK '07. Graz, 17.–21. September 2007. ODOK; Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare; Österreichisches Online-Informationstreffen; Österreichischer Dokumentartag. Graz, Feldkirch: Neugebauer (Schriften der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 5).

## Zeitschriftenaufsätze

- Laminger, H.; Schopper, M.; Pipp, Eveline; Hensler, I.; Mantl, P. (1981): Untersuchungen über Nekrozönosen und Taxozönosen der Testacea (Protozoa) im Zirbenwaldmoor (Obergurgl, Tirol/Austria). In: *Hydrobiologia* 77, S. 193–202.
- Pipp, Eveline; Larcher, Walter (1986): Energiegehalte pflanzlicher Substanz. In: *Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, Abteilung 1*, S. 196.
- Pipp, Eveline; Rott, Eugen (1994): Classification of running-water sites in Austria based on benthic algal community structure. In: *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie* 25 (3), S. 1610–1613.
- Pipp, Eveline; Rott, Eugen (1995): A phytoplankton compartment model for a small meromictic lake with special reference to species-specific niches and long-term changes. In: *Ecological Modelling* 78 (1-2), S. 129–148.
- Cantonati, Marco; Rott, Eugen; Pipp, Eveline (1996): Ecology of cyanophytes in mountain springs of the river Sarca catchment (Adamello-Brenta Regional Park, Trentino, Northern Italy). In: *Archiv für Hydrobiologie. Supplementband, Algological studies* 117, S. 145–162.
- Pipp, Eveline (1996): Knight-Ridder Probase : ein Testbericht. In: *Online-Mitteilungen* (56), S. 10–22.
- M. G. Kelly; A. Cazaubon; E. Coring; A. Dell'Uomo; L. Ector; B. Goldsmith et al. (1998): Recommendations for the routine sampling of diatoms for water quality assessments in Europe. In: *Journal of Applied Phycology* 10, S. 215–224.
- Rott, Eugen; Duthie, Hamish C.; Pipp, Eveline (1998): Monitoring organic pollution and eutrophication in the Grand River, Ontario, by means of diatoms. In: *Canadian Journal of Fisheries and Aquatic Sciences* 55 (6), S. 1443–1453.
- Rott, Eugen, Duthie, H.; Pipp, Eveline (1998): Diatoms for monitoring organic pollution and eutrophication in Grand River, Ontario. In: *Canadian Journal of Fisheries and Aquatic Sciences* 55, S. 1443–1453.
- Cantonati, Marco; Pipp, Eveline (2000): Longitudinal and seasonal differentiation of epilithic diatom communities in the uppermost sections of two mountain spring-fed streams. In: *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie* 27, S. 1591–1595.
- Pipp, Eveline (2000): ODOK '99 – Bregenz Online „Informationssysteme und Informationsberufe im Umbruch“. In: *Bibliotheksdienst* 34 (1), S. 30–36.

- Pipp, Eveline (2001): A regional diatom-based trophic state indication system for running water sites in Upper Austria and its overregional applicability. In: *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie* 27, S. 3376–3380.
- Pipp, Eveline (2001): E-Psyche – die umfassendste, kostengünstigste und aktuellste Datenbank im Fachbereich Psychologie. In: *Online-Mitteilungen* (71), S. 3–12.
- Pipp, Eveline (2001): Infobase und ERL-User Meeting 2001. In: *Online-Mitteilungen* (70), S. 15–17.
- Pipp, Eveline (2001): Volltextdatenbanken im Vergleich. In: *Bibliotheksdienst* 35 (9), S. 1061–1078.
- Pipp, Eveline; Fröhlich, G. (2001): ODOK '01: Tagungsbericht. In: *Online-Mitteilungen* (70), S. 5–14.
- Rott, Eugen; Pipp, Eveline; Pfister, Peter (2003): Diatom methods developed for river quality assessment in Austria and a cross-check against numerical trophic indication methods used in Europe. In: *Algological Studies* 110 (1), S. 91–115.
- Pipp, Eveline (2005): Embargos und abgebrochene Volltextangebote verlieren Volltextdatenbanken an Wert? In: *Online-Mitteilungen* (81), S. 19–34.
- Pipp, Eveline (2006): Aller Anfang ist schwer. In: *Online-Mitteilungen* (88), S. 3–5.
- Pipp, Eveline (2006): Vergleich der von Scopus bzw. Web of Science erfassten Zeitschriften. In: *Online-Mitteilungen* (85), S. 3–17.
- Pipp, Eveline (2007): Abschied von den „gelben Seiten“. In: *Online-Mitteilungen* (92), S. 3–4.
- Pipp, Eveline (2007): Rezension: Chan, Leslie & Bob Martens (Hrsg.); Openness in Digital Publishing: Awareness, Discovery and Access. Proceedings of the 11th International Conference on Electronic Publishing. ÖKK-Editions, 2007. In: *Online-Mitteilungen* (91), S. 9–10.
- Pipp, Eveline (2007): Kommen Sie zur ODOK '07? Ein vielseitiges Workshop- und Vortragsprogramm erwartet Sie! In: *Online-Mitteilungen* (91), S. 3–5.
- Pipp, Eveline (2007): Vascoda, virtuelle Fachbibliotheken und Informationsverbünde. In: *Online-Mitteilungen* (89), S. 3–5.
- Pauser, Josef; Pipp, Eveline (2008): 2009 naht! Denken sie an den Bibliothekartag in Graz. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 61 (4), S. 6.
- Pauser, Josef; Pipp, Eveline (2008): „Online-Mitteilungen“ nun integraler Bestandteil der Mitteilungen. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 61 (1), S. 6.

- Pipp, Eveline (2008): Rezension: Marion Kaufer: Erwerbungsprofile in wissenschaftlichen Bibliotheken. Eine Bestandsaufnahme, Graz-Feldkirch 2008. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 61 (4), S. 160–161.
- Pipp, Eveline (2008): Rezension: Rafael Ball (Hrsg.): Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek Forschungszentrum Jülich. 6.–8. November 2007. Beiträge und Poster, Jülich 2007. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 61 (2), S. 79–81.
- Pipp, Eveline (2008): Rezension: Wolfgang G. Stock, Mechtild Stock: Wissensrepräsentation. Informationen auswerten und bereitstellen, München 2008. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 61 (3), S. 137–138.
- Pipp, Eveline (2009): Einladung 30. Österreichischer Bibliothekartag in Graz. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 62 (3), S. 6.
- Pipp, Eveline (2010): ODOK 2010 in Leoben. Eine Nachlese aus der Sicht der Organisationsleitung. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 63 (3/4), S. 139–145.
- Pipp, Eveline; Knitel, Markus; Mayer, Wolfgang; Putz, Michaela (2011): Bericht der AG Primo-Implementierung an die Vollversammlung des österreichischen Bibliothekenverbundes zum möglichen Einsatz des Dienstes Primo Central. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 64 (2), S. 196–214.
- Pipp, Eveline (2012): Call for Papers – ODOK 2012. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 65 (1), S. 133–134.
- Moonsin, Parinya; Peerapornpisal, Yuwadee; Rott, Eugen; Pipp, Eveline; Pimmongkol, Aranya (2013): Diversity of Benthic Diatoms and Water Quality of the Mekong River Passing Through Ubon Ratchathani Province, Thailand. In: *Research Journal of Environmental and Earth Sciences* 5 (2), S. 58–66.
- Pipp, Eveline (2013): Helmut Hartmann zum Abschied. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 66 (3/4), S. 416–421.
- Kirchmair, Susanne; Pipp, Eveline (2015): Schwerpunktthema „ODOK 2014“. In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 68 (1), S. 6–9.
- Blumesberger, Susanne; Gründhammer, Veronika; Koch, Gerda; Pipp, Eveline; Plößnig, Veronika (2016): 6. Tagung „Digitale Bibliothek“ – „Vernetzte Welten – Digitales Kulturerbe als öffentliches Gut“ (Graz, 26. Fe-

bruar 2016). In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 69 (2), S. 319–328.

## Beiträge in Sammelbänden

- Pipp, Eveline (1991): Use of periphyton for monitoring rivers in Austria. In: Brian A. Whitton (Hg.): Use of algae for monitoring rivers. Proceedings of an International Symposium held at the Landesamt für Wasser und Abfall Nordrhein-Westfalen Düsseldorf, Germany, 26–28 May 1991. Innsbruck: Rott, S. 9–16.
- Pipp, Eveline; Rott, Eugen (1996): Recent development on use of benthic algae (excluding diatoms) for monitoring rivers in Austria and Germany. In: Brian A. Whitton und Eugen Rott (Hg.): Use of algae for monitoring rivers II. Proceedings of an International Symposium held at the Volksbildungsheim Grillhof, Vill near Innsbruck; Innsbruck, Austria, 17–19 September 1995. Innsbruck: Rott, S. 160–165.
- Rott, Eugen; Pipp, Eveline (1999): Progress in the use of algae for monitoring rivers in Austria. In: Brian A. Whitton und J. Bukowska (Hg.): Use of algae for monitoring rivers III. Agence de L'Eau: Douai, France, S. 110–112.
- Pipp, Eveline (2000): CD-ROM-Datenbanken im LAN der Universität Innsbruck – technische und lizenzrechtliche Probleme für den Systemoperator. In: Eveline Pipp (Hg.): Dokumente und Datenbanken in elektronischen Netzen. Dokumente und Datenbanken in elektronischen Netzen: Tagungsberichte vom 6. und 7. Österreichischen Online-Informationstreffen bzw. vom 7. und 8. Österreichischen Dokumentartag, Schloß Seggau, Seggau bei Leibnitz, 26.–29. September 1995, Congresszentrum Iglis bei Innsbruck, 21.–24. Oktober 1997. Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare; Österreichisches Online-Informationstreffen; Österreichischer Dokumentartag. Innsbruck: Österr. Online-Benutzergruppe (Biblos-Schriften, 161), S. 63–70.
- Pipp, Eveline (2002): Volltextdatenbanken im Vergleich. In: Eveline Pipp (Hg.): Drehscheibe E-Mittleuropa. Information Produzenten, Vermittler, Nutzer, die gemeinsame Zukunft; Tagungsberichte vom 9. Österreichischen Online-Informationstreffen und 10. Österreichischen Dokumentartag; 24.–27. April 2001, Universität Graz. Wien: Phoibos (Biblos-Schriften, 173), S. 179–196.
- Pipp, Eveline (2004): SFX und weitere Link Resolver – ein Produktvergleich. In: Eveline Pipp (Hg.): Ein Jahrzehnt World Wide Web: Rückblick –

Standortbestimmung – Ausblick. ODOK '03 ; Tagungsbericht vom 10. Österreichischen Online-Informationstreffen und 11. Österreichischen Dokumentartag ; 23.–26. September 2003, Universität Salzburg, Naturwissenschaftliche Fakultät. Wien: Phoibos (Biblos-Schriften, 179), S. 277–288.

- Bauer, Bruno; Androsch, Günther; Dollfuß, Helmut; Hartmann, Helmut; Kastanek, Peter; Pipp, Eveline et al. (2005): Wie beurteilen Nutzer unserer elektronisches Medien-und Dienstleistungsangebot? Ausgewählte Ergebnisse der gemeinsamen Online-Benutzerbefragung 2003 an zehn österreichischen Universitäts- und Zentralbibliotheken. In: Christian Enichlmayr (Hg.): Bibliotheken – Fundament der Bildung. Tagungsband; 21.–25. September, Linz. Österreichischer Bibliothekartag 2004. Linz, 21.–25. September. Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare. [Weitra]: Verl. Publication PN°1 – Bibliothek d. Provinz (Schriftenreihe der Oö. Landesbibliothek), S. 151–189.
- Pipp, Eveline (2013): Das E-Medien-Angebot österreichischer Universitätsbibliotheken. In: Bruno Bauer (Hg.): Universitätsbibliotheken im Fokus. Aufgaben und Perspektiven der Universitätsbibliotheken an öffentlichen Universitäten in Österreich. Graz, Feldkirch: Neugebauer (Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare: Schriften der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare (VÖB), 13), S. 142–153.

## **Studien, Berichte**

- Pechlaner, Robert; Rott, Eugen; Pipp, Eveline (1988): Zwei Hochgebirgsbäche im Einzugsbereich eines Kraftwerkspeichers (Finstertal, Längentalbach). Limnologische Projektstudie, Univ. Innsbruck, Institut für Botanik und Institut für Zoologie, Abt. für Limnologie.
- Rott, Eugen; Pechlaner, Robert; Pipp, Eveline (1990): Zwei Gebirgsbäche im Einzugsbereich der Isar. Limnologische Projektstudie, Univ. Innsbruck, Institut für Botanik und Institut für Zoologie, Abt. für Limnologie.
- Pipp, Eveline; Rott, Eugen (1993): Ökologische Wertigkeit österreichischer Fließgewässer nach dem Algenaufwuchs. durch das Bundesministerium für Umwelt, Jugend und Familie gefördertes Projekt Nr. 3117. 1. Aufl. Wien (Blaue Reihe des Bundesministeriums für Umwelt, Jugend und Familie).
- Pipp, Eveline (1997): Klassifikation oberösterreichischer Fließgewässer anhand der Kieselalgen. Wasserwirtschaftskataster. Wien.

- Rott, Eugen; Hofmann, G.; Pall, K.; Pfister, Peter; Pipp, Eveline (1997): Indikationslisten für Aufwuchsalgen in Fließgewässern in Österreich. Teil 1: Saprobielle Indikation. Wasserwirtschaftskataster. Wien.
- Rott, Eugen; Pfister, Peter; Pipp, Eveline (1997): Use of diatoms for environmental monitoring. Final report of the project study. Nov 2–8, 1997. Innsbruck.
- Rott, Eugen; Pipp, Eveline; Pfister, Peter; van Dam, H.; Ortler, K.; Binder, N.; Pall, K. (1999): Indikationslisten für Aufwuchsalgen in österreichischen Fließgewässern. Teil 2: Trophieindikation (sowie geochemische Präferenzen, taxonomische und toxikologische Anmerkungen). Wasserwirtschaftskataster. Wien.
- Pfister, Peter; Pipp, Eveline (2005): Handlungsanweisungen für die ökologische Bewertung für österreichische Fließgewässer an Hand des Phytobenthos gemäß EU-Wasserrahmenrichtlinie. Studie im Auftrag des Bundesministeriums für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. Wien.
- Pipp, Eveline; Pfister, Peter (2005): Leitbildbezogenes Bewertungsverfahren für österreichische Fließgewässer an Hand des Phytobenthos gemäß EU-Wasserrahmenrichtlinie. Vorschlag für ein Trophie-basiertes Bewertungsverfahren. Studie im Auftrag des Bundesministeriums für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. Wien.
- Pfister, Peter; Pipp, Eveline (2010): Fließgewässer Qualitätselement Phytobenthos: Felderhebung, Probenahme, Probenaufarbeitung und Ergebnisermittlung. Leitfaden zur Erhebung der biologischen Qualitätselemente. Teil A3 – Phytobenthos. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. Wien.

## **Vorträge, Präsentationen**

- Pipp, Eveline; Rott, Eugen (1988): Application of mathematical methods of vegetation analysis to algal communities in running waters. – Poster Abstract, Symposium Progress in Theoretical Vegetation Science, July 4–7.
- Pipp, Eveline; Rott, Eugen; Gesierich, D. (1990): Aktuelle, diatomeenbezogene Untersuchungen an der AG Hydrobotanik, Institut für Botanik, Universität Innsbruck. 13. Deutschsprach. Diatomologentreffen, Schloß Kröchlendorf, Deutschland. 2 S (Posterabstract).
- Rott, Eugen; Pipp, Eveline; Pfister, Peter (1990): Can algal communities be used for the classification of running waters? – Poster Abstract, Symposium: "The conservation and management of rivers", September 10–13, York, U.K.

- Pipp, Eveline (1991): Algenaufwuchs in Fließgewässern (Artengruppen zur Typisierung und ökologischen Bewertung der österreichischen Fließgewässer). – Vortrag-Abstract, 6. Österr. Botanikertreffen, 30. Mai–3. Juni, Graz.
- Pipp, Eveline; Maier, Martin; Pfister, Peter; Rott, Eugen (1991): Phytozoologische Untersuchungen am Algenaufwuchs aus naturnahen Gebirgsbächen. Jahrestagung DGL, Mondsee.
- Pipp, Eveline (1997): The comparison of diatom index systems in view of supraregional standardization and regional applicability. Abstracts. Part III. 3rd European Workshop Use of algae for monitoring rivers. Douai, France.
- Rott, Eugen; Pipp, Eveline (1997): Recent development in Austrian periphyton investigations. Abstracts 3<sup>rd</sup> European Workshop Use of algae for monitoring rivers. Douai, France. Part I. Situation and recent developments in various countries.
- Pernegger, Ludwig; Pipp, Eveline (2000): Seasonality and growth of epilithic phytobenthos in a soft-water high mountain lake. Abstract. Internat. Symposium. High mountain lakes and streams. 4–8 September 2000. Univ. Innsbruck. Poster.
- Pipp, Eveline (2000): Validation of genera based and species based diatom methods for trophic state indication in European rivers. Abstract. 4th Internat. Symposium Use of Algae for Monitoring European Rivers. 15–20 July 2000. Univ. Durham, U.K.
- Rott, Eugen; Pfister, Peter; Pipp, Eveline; Gesierich, D.; Binder, N. (2002): Diatomeen in Forschung und Gewässerüberwachung an der Arbeitsgruppe Hydrobotanik und der ARGE Limnologie Innsbruck. 16. DDT, Kurzfassungen. Institut für Botanik. Innsbruck.
- Pechlaner, Robert; Pfister, Peter; Pipp, Eveline; Rott, Eugen (2003): Threats to freshwater ecosystems from Glyphosate-based herbicide applications on land. Abstract Poster Session. 13<sup>th</sup> SETAC annual European meeting, Hamburg 27.04.2003–01.05.2003.
- Rott, Eugen; Pfister, Peter; Pipp, Eveline (2005): Recent developments in the use of benthic algae for ecological water quality analysis in rivers of Central Europe. Abstract 2.14.5. XVII<sup>th</sup> Botanical Congress 17–24 July 2005, Vienna.



Helmut Hartmann  
E-Mail: [hhart@aon.at](mailto:hhart@aon.at)

Ao.Univ.-Prof. i.R. Dr. Eugen Rott  
Universität Innsbruck, Institut für Botanik  
E-Mail: [eugen.rott@uibk.ac.at](mailto:eugen.rott@uibk.ac.at)

Dr. Klaus Niedermair  
Universität Innsbruck, Universitäts- und Landesbibliothek Tirol  
E-Mail: [klaus.niedermair@uibk.ac.at](mailto:klaus.niedermair@uibk.ac.at)

Mag. Eva Ramminger  
Universität Innsbruck, Universitäts- und Landesbibliothek Tirol  
E-Mail: [eva.ramminger@uibk.ac.at](mailto:eva.ramminger@uibk.ac.at)

- 1 Für die Überlassung von Informationen zu diesem Beitrag sei HR Dr. Heinz Hauffe herzlich gedankt!

**■ TREUHÄNDERISCHE ÜBERNAHME UND VERWAHRUNG – INTERNATIONAL UND INTERDISZIPLINÄR BETRACHTET: TAGUNG ZUR NS-PROVENIENZFORSCHUNG AN DER UNIVERSITÄT WIEN (WIEN, 2.–4. MAI 2017)**

*von Stefan Alker-Windbichler und Bruno Bauer*

Die Eröffnung der Tagung, die ausschließlich dem Thema „Treuhänderische Übernahme und Verwahrung“ im Kontext der NS-Provenienzforschung gewidmet war, erfolgte am Abend des 2. Mai 2017 in der Aula am Campus der Universität Wien. In ihren Grußworten wies Vizerektorin Regina Hitzenberger darauf hin, dass sich die Universität Wien in mehreren Projekten mit den dunklen Jahren ihrer Geschichte aktiv auseinandersetzt. Diese Entwicklung begann mit dem Symposium „Vertriebene Vernunft“ von Friedrich Stadler Ende der 1980er Jahre, wurde fortgesetzt mit der Restaurierung des ehemaligen jüdischen Gebetshauses und dem Gedenkbuch der Universität für die Opfer des Nationalsozialismus und findet ihre Fortsetzung in der aktuellen Kunstinstitution in der Aula im Hauptgebäude der Universität, in der die problematische Rolle der Rektoren während der NS-Zeit thematisiert wird. In diesem Zusammenhang ist auch NS-Provenienzforschung zu nennen, die seit 2004 betrieben wird. In jüngster Zeit hat die Universität mit dem Nationalfonds einen Vertrag über den Umgang mit unrechtmäßig erworbenen Büchern, für die kein Rechtsnachfolger ermittelt werden konnte, abgeschlossen.

Es folgten Grußworte von Hannah M. Lessing, Generalsekretärin des Nationalfonds der Republik Österreich für die Opfer des Nationalsozialismus, die aufgrund einer terminlichen Verhinderung von Olivia Kaiser verlesen wurden. Darin wurden ebenfalls Fälle erbloser Kunstwerke angesprochen, die an den Nationalfonds übergeben und anschließend von der betreffenden Institution rückgekauft worden sind. Dieses Procedere wurde bereits 2009 von der Österreichischen Nationalbibliothek angewendet und kam jüngst auch an der Universität Wien zum Tragen.

Im Eröffnungsvortrag führten Olivia Kaiser, Christina Köstner-Pemsel und Markus Stumpf unter dem Titel „*Treuhänderisch* – *Themenaufriss im Kontext der NS-Provenienzforschung*“ in das Thema der Tagung ein. Zunächst erläuterte Stumpf das Beispiel von zirka 1.800 Bänden der All Peoples



Abb. 1: Eröffnung der Tagung „Treuhandliche Übernahme und Verwahrung“ – (v.l.n.r.) Markus Stumpf, Olivia Kaiser, Maria Seissl, James D. Bindenagl, Vizerektorin Regina Hitzemberger, Christina Köstner-Pemsel (Foto: Josef Krpelan)



Abb. 2: Veranstaltungsort Aula am Campus der Universität Wien (Foto: Josef Krpelan)

Association (APA). Diese Bände waren zunächst als Leihgabe an das Seminar für englische Philologie der Universität Wien gelangt, wurden aber ab 1942 als Eigenbestand der Universität geführt. Weil im Zuge der NS-Provenienzforschung der Universität der Hintergrund zu diesen Beständen ermittelt werden konnte, wurde von der Universität Wien 2010 ein Rückgabebeschluss gefasst. Allerdings konnten keine erbberechtigten Personen ermittelt werden, weshalb die 1.800 Bände als erblose Werke klassifiziert, an den Nationalfonds übergeben und wieder zurückgekauft wurden. Anschließend skizzierte Kaiser die Entwicklung der NS-Provenienzforschung in Österreich im Kontext des Kunstrückgabegesetzes und der Washington Principles. Daran anschließend sprach Köstner-Pemsel über die Sammlung Tanzenberg, die den Beginn der NS-Provenienzforschung an der Universität Wien markiert. 2004 wurde dieser Bestand im Rahmen eines Projektes von Teilnehmerinnen des ULG Library and Information Studies untersucht. Dieser Bestand im Umfang von zirka 38.000 Bänden kam über die Büchersortierungsstelle an die Universitätsbibliothek Wien. 1959/60 erfolgte eine Aufteilung von Teilen dieses Bestandes zwischen den Bibliotheken in Wien und der Universität Jerusalem, dann erfolgte eine Einarbeitung des Bestandes. Die Bestände tragen den Vermerk „Sammlung Tanzenberg“, diese Information ist auch als exemplarspezifischer Hinweis im Bibliothekssystem eingetragen.



Abb. 3: Festredner James D. Bindenagel und Markus Stumpf (Foto: Josef Krpelan)

Keynote Speaker der Tagung war **James Bindenagel**, früherer US-Botschafter für Holocaust Issues, zum Thema „*Artworks Looted during the Holocaust: The Unfinished Story. The Washington Principles on Nazi-confiscated Art: Seeking Just and Fair Solutions*“. Er wies auf die große Bedeutung der Restitution von geraubtem Kulturgut hin. In Österreich war es der Fall Klimt, der den Start der aktuellen Bemühung der NS-Provenienzforschung und Restitution kennzeichnet – das betreffende Bild wurde an Maria Altmann restituiert.

Am Mittwoch, den 3. Mai 2017, startete die Tagung mit einem von Oliver Rathkolb (Wien) moderierten Panel zum Thema „*Österreich*“.

Zunächst sprach **Sebastian Spitra** (Wien) über „*Recht und Metapher: Die treuhänderische Verwaltung von Kulturgut mit NS-Provenienz*“. Für den Terminus „Treuhand“, die er als Metapher des Kulturgüterschutzes bezeichnete, gibt es keine einheitliche Definition. Charakteristisch ist eine starke interdisziplinäre Rezeption.

**Leonhard Weidinger** (Wien) erläuterte die Frage „*Vor Mauerbach – Woher stammen die 967 Kunstwerke, die Österreich 1952 aus München übernahm? Und woher die anderen aus dem Mauerbach-Bestand?*“. Die 1952 übernommenen Kunstwerke wurden mit jenen Beständen zusammengeführt, die von niemandem beansprucht wurden, und in der ehemaligen Kartause Mauerbach eingelagert. 1969 und 1986 erfolgte im Rahmen der Kunst- und Kulturbereinigungsgesetze die Veröffentlichung von Listen mit 8.422 bzw. 8.153 Objekten. Auf Basis einer Novelle (1995) sollte der Rest versteigert werden, was 1996 am MAK auch umgesetzt worden ist. Den Fokus seiner Ausführungen legte Weidinger auf die aus München an das österreichische Bundesdenkmalamt übergebenen 967 Kunstwerke. Zu diesen wurden auch 909 Karteikarten (property cards) mitgeliefert. Von den 967 Kunstwerken wurden fünf Gemälde von erbberechtigten Personen reklamiert, der Rest wurde im Rahmen der Mauerbachauktion versteigert. Im Zusammenhang mit dieser Vorgangsweise ist kritisch anzumerken, dass die Provenienzmerkmale auf den Projekten nicht berücksichtigt und die Erkenntnisse aus diesen Quellen nicht gut dokumentiert worden sind. Im Übrigen stammen viele der 967 Kunstwerke nicht aus Österreich, sondern aus Deutschland. Nur ein sehr geringer Teil der Karteikarten – 35 von 909 – stammt aus Österreich, der überwiegende Teil ist Deutschland zuzuordnen.

**Michael Wladika** (Wien) sprach über „*Die Beanspruchung von Kunst- und Kulturgegenständen durch die Sammelstellen 1959–1972*“. Jüdisches Vermögen nach 1945 blieb oft erblos, sodass jüdische Organisationen die Befürchtung äußerten, dass dieses Vermögen an die Republik fallen oder im Besitz der Profiteure der Arisierung bleiben könnte. Deshalb beinhaltete der Staatsvertrag einen Passus für die Rückstellungen. Im Jänner 1957 wurde

ein Gesetz zur Einrichtung von Sammelstellen beschlossen. Diese sollte jene unterstützen, die bisher keinen Antrag auf Rückstellung machen konnten. Der Schwerpunkt wurde auf Liegenschaften gelegt, weil bis dahin nur sechs Prozent der Fälle auf Liegenschaften gefallen waren. Der Grund für diese geringe Zahl lag im Fehlen einer Provenienzforschung in diesem Bereich. Es kam zu einer Vergleichszahlung im Ausmaß von fünf Millionen Schilling, einer Summe, die in keiner Beziehung zum Wert der Gegenstände oder Bilder stand. 1972 erfolgte die Auflösung der Sammelstelle.

**Anneliese Schallmeiner** und **Alexandra Caruso** (Wien) widmeten sich dem Thema „*Das Bundesdenkmalamt (BDA) und der Bestand der sog. 1960er Jahre-Zuweisungen*“. Dabei handelt es sich um 98 Positionsnummern bzw. 266 Kunstobjekte (Gemälde, Grafiken, Skulpturen), die u.a. dem Kunsthistorischen Museum und der Albertina zugute gekommen sind. Vorangegangen war im September 1952 der Vermögensverfall Adolf Hitler zu Gunsten der Republik Österreich. Im August 1963 erfolgte die Räumung des Depots in der Salzburger Residenz – die Räumlichkeiten wurden der neu gegründeten Universität Salzburg überantwortet – und deren treuhänderische Übernahme durch die Bundesmuseen, denen die Objekte überantwortet worden sind. 1965 erfolgte die definitive Übernahme der betreffenden Objekte in die betreffenden Inventare.

Das zweite Panel, moderiert von **Maria Kesing** (Hamburg), war dem Thema „*Deutschland*“ gewidmet.

**Thomas Rudert** (Dresden) eröffnete das Panel mit einem Vortrag über „*Enteignung – Sicherung – Restitution. Der museumspraktische und juristische Umgang mit Kunstwerken der Bodenreform/Schlossbergung an den Staatlichen Kunstsammlungen Dresden (1945–1994)*“. 1951 erfolgte die Verhaftung von Wolfgang Balzer (1884–1968), der von 1946 bis zu seiner Verhaftung Direktor der staatlichen Kunstsammlung Dresden gewesen war. Wenige Wochen später wurde er aus der Haft entlassen und in den Ruhestand versetzt. Balzer hatte gegen den Abtransport von Beständen opponiert. Bis 1950 waren viele Werke aus den sogenannten Schlossbergungsbeständen verkauft worden; erst in diesem Jahr verfügte der damalige sächsische Innenminister Arthur Hofmann die Übergabe der noch vorhandenen Werke an die Staatlichen Kunstsammlungen Dresden. Ab 1955 wurden von der Sowjetunion Kunstwerke, die als Beutekunst nach Moskau und Kiew abtransportiert worden waren, an die Staatliche Kunstsammlung Dresden zurückgegeben.

Daran anschließend berichtete **Jana Kocourek** (Dresden) über „*Offene Vermögensfragen – von der Suche nach sogenannten Schlossbergungsbeständen in der SLUB Dresden*“. Bis 1945 gab es in Sachsen bedeutende Schlossbibliotheken. An den 170 Bibliotheken befanden sich geschätzte zwei Millionen

Bände. Durch die Bodenreform erfolgte eine Verstreuung dieser Bestände. Zirka 200.000 Bände der Schlossbergungsbestände blieben an der Sächsischen Landesbibliothek. Dubletten wurden an andere Bibliotheken abgegeben; seit den 1970er Jahren erfolgte ein Verkauf von Dubletten zur Devisenbeschaffung. Zwischen 2009 und 2013 erfolgte die Erfassung der Bücher aus den Schlossbergungsbeständen und der entsprechenden Provenienzmerkmale in einer Datenbank, wobei auch eine Verknüpfung mit den Normdaten der GND erfolgte. Die Suche nach berechtigten Besitzern oder Erben geht in zwei Richtungen: NS-entzogene Bücher sowie DDR-entzogene Bücher.

Petra Winter (Berlin) widmete sich dem Thema der *„Überweisungen an die Staatlichen Museen zu Berlin (Ost) durch Ministerien der DDR und andere staatliche Stellen“*. Die Museumsinsel in Berlin wurde im Mai 1945 von der Roten Armee besetzt. Nach Auflösung des Staates Preußen erfolgte eine treuhänderische Verwaltung der Museen. Erst 1951 wurden die Museen wieder zu staatlichen Museen. 1968 erfolgte der Auftrag zur Erhebung, welche Objekte nach dem 1. Mai 1968 an die staatlichen Museen der DDR gelangt sind.

*„Bücher als Danaergeschenk. Nachkriegszugänge der UB Mainz durch die französische Militärregierung“* lautete das Thema des Vortrags von Christian George (Mainz). Weil die Universitätsbibliothek Mainz erst 1946 gegründet worden ist, liegt der Schwerpunkt der gesammelten Literatur auf Titeln mit Erscheinungsjahr nach 1946. Allerdings gelangten Geschenke, auch in Abstimmung mit der französischen Militärregierung, an die Universitätsbibliothek, darunter die 1.223 Titel umfassende sogenannte Bormann-Bibliothek. Diese wurde im Rara-Magazin der Bibliothek aufgestellt. Auch Buchbestände, die in der NS-Zeit beschlagnahmt worden sind, wurden an die Universitätsbibliothek übergeben, weiters die sogenannte Sabotini-Bibliothek – 1991 erfolgte die Rückgabe von 1.253 Bänden an erbberechtigte Personen – sowie die Bibliothek der Kunsthistorischen Forschungsstätte Paris. Die Universitätsbibliothek Mainz stellt somit ein sehr gutes Beispiel für eine Bibliothek dar, die, obgleich erst nach 1945 gegründet, sehr stark vom Thema NS-Provenienzforschung und Restitution betroffen ist, weil wiederholt größere außergewöhnliche Zugänge an Bibliotheksbeständen zu registrieren waren.

Das dritte Panel, als dessen Moderator Reinhard Buchberger (Wien) fungierte, trug den Titel *„Prag“*.

Michal Bušek (Prag) sprach zum Thema *„Provenience research in the book collection of the Jewish Museum in Prague“*. Dem 1906 gegründeten Museum wurden zwischen 1945 und 1950 mehr als 190.000 Bände übergeben, von denen in den Folgejahren zirka 158.000 Bände restituiert, verkauft oder an



jüdische Organisationen überantwortet wurden. Zwischen 2001 und 2015 wurde ein Provenienzforschungsprojekt durchgeführt, in dessen Rahmen knapp 100.000 Bände überprüft und zirka 49.000 Bände in einer Datenbank erfasst wurden. Erfasst wurden auch die rechtmäßigen Besitzer.

„*Looted, stolen, confiscated or just given? About the most common question arising during the provenance signs survey inside the National Library of Czech Republic collections*“ – unter diesem Titel erläuterten **Tomáš Foltýn** und **Marcela Strouhalova** (Prag), wie in der kommunistischen Ära zirka 16.000 Bände von mehr als 2.000 rechtmäßigen Vorbesitzern an die Tschechische Nationalbibliothek im Clementinum gelangt sind. Im Zuge der Recherchen wurde festgestellt, dass etwa viele Bücher aus dem Sudetenland übernommen worden sind.

Über „*Research in the collection of Hebrew prints stored in the National Library of the Czech Republic*“ berichtete **Michael Nosek** (Prag). Das Documentation Centre for Property Transfers of Cultural Assets of WW II Victims unterstützte die Tschechische Nationalbibliothek bei der Identifizierung der an dieser Bibliothek aufbewahrten Sammlung hebräischer Drucke. Der Schwerpunkt der Sammlung liegt auf religiösen Büchern – vom Talmud über Shulchan Aruch bis zu Gebetsbüchern. Aufgrund von Vorbesitzereinträgen in Form von Stempeln oder Etiketten wurde es möglich, die Sammlung zu rekonstruieren. Die Bestände wurden in der Datenbank für Kunstobjekte des Documentation Centre erfasst, wobei die Dokumentation neben der Beschreibung auch Fotos von den Objekten beinhaltet.

**Johana Prouzová** (Prag) führte in die Problematik der „*Sammlung Pollak in Prager Museen*“ ein. Der Kaufmannssohn Josef Pollak (1896–1944) übernahm die Firma seines Vaters und erbt auch eine umfangreiche Kunstsammlung. Er war dreimal verheiratet und hatte eine Tochter. 1942 wurde er nach Theresienstadt gebracht, 1944 nach Auschwitz. 1939 wurden 37 Kunstgegenstände aus der Sammlung Pollak in die Nationalgalerie überstellt, wo sich heute nur mehr ein Bild dieses Bestandes befindet. Ebenfalls im Jahr 1939 wurden weitere 957 Kunstgegenstände an das Kunstgewerbe-Museum überstellt. Während 231 Gegenstände im Jahr 1948 restituiert wurden, verblieben 102 Objekte im Museum, die erst 1952 inventarisiert wurden; das Schicksal der weiteren Gegenstände ist unbekannt. Viele Objekte der Sammlung Pollak sind mit einem kleinen Davidstern gekennzeichnet. Der Restitutionsfall Pollak wurde dadurch verkompliziert, dass sowohl die Tochter als auch seine dritte Frau die zu restituierenden Kunstgegenstände für sich reklamierten.

Das vierte Panel, moderiert von Bertrand Perz (Wien), war dem Thema „*Wien*“ gewidmet.



Zunächst sprach **Monika Mayer** (Wien) über die „*Österreichische Galerie Belvedere*“. 1961 wurden 13 Werke aus dem Besitz des früheren Reichsleiters Bormann an die Galerie übergeben. 1965 wurden weitere 70 Werke in treuhänderische Verwaltung übernommen. 1974, 1980 und 1984 wurden weitere 35 Kunstwerke aus dem Eigentum des Bundes an die Galerie übergeben. In den Folgejahren wurde die aktive Rückgabe zum Prinzip erhoben. In diesem Zusammenhang ist etwa der Mauerbachbestand zu nennen, der im handschriftlichen Inventar der Österreichischen Galerie Belvedere mit M-Signaturen versehen wurde.

Anschließend setzte sich **Monika Löscher** (Wien) mit dem Thema „*Die 1963er Zuweisungen an das Kunsthistorische Museum. Zum Stand der Provenienzforschung*“ auseinander. Diese Objekte wurden 1965 inventarisiert, 1986 erfolgte die Übergabe einiger Objekte an die Österreichische Galerie Belvedere. Erst in jüngster Zeit wird am Kunsthistorischen Museum intensiv NS-Provenienzforschung betrieben, während die Jahre nach 1948 dahingehend argumentiert worden war, dass das Museum nie bei Juden gekauft habe, dass man keine Kriegsnotlage ausgenutzt habe und dass es keine Geschäftsunterlagen geben würde.

„... ich kann Sie versichern, dass ich Ihnen das Paket mit den biogr. Schriften mit dem größten Vergnügen aufhebe“ – ‚*Treuhänderisch*‘ übernommene Sammlungen in der *Wienbibliothek*“ lautete das Thema von **Christian Mertens** (Wien). Die Wienbibliothek startete 1999 aufgrund eines Gemeinderatsbeschlusses mit der NS-Provenienzforschung. Zirka 80.000 Bände wurden untersucht, 2.835 Objekte konnten restituiert werden. In sieben Fällen war die Suche nach Rechtsnachfolgern nicht erfolgreich. Zwei bedenkliche Sammlungen gelangten erst nach 1945 als treuhänderisches Gut an die Wienbibliothek. Charlotte Holzmann (1870–1942), Ehefrau von M. Holzmann (1860–1930), dem Herausgeber des siebenbändigen deutschen Anonymen-Lexikons, wurde nach Theresienstadt und Treblinka deportiert. 1945 wurde der Bestand der Sammlung Holzmann zur treuhänderischen Verwahrung an die damalige Stadtbibliothek übergeben. Zwanzig Jahre später, im Jahr 1965 begann man mit der Inventarisierung. Die Autographen-Sammlung der Schwestern Else Richter (1865–1043) und Helene Richter (1861–1042) wurde unmittelbar vor deren Deportierung nach Theresienstadt von Christine Rohr-Denta, einer Mitarbeiterin der Österreichischen Nationalbibliothek, in Verwahrung genommen. 1947 überantwortete Rohr-Denta die ihr anvertrauten Objekte der Stadt Wien. Zunächst gelangten sie an das Wien-Museum, später in die Wienbibliothek. Ungeklärt ist bis heute, warum Rohr-Denta eine Einrichtung der Stadt Wien und nicht die Österreichische Nationalbibliothek aufgesucht hat. Mit der Katalogisierung der

Objekte der Sammlung Richter wurde um 1950 begonnen, die meisten Objekte wurden allerdings erst um 2000 erfasst.

Abgeschlossen wurde das vierte Panel mit einem Vortrag von **Philipp Mettauer** (Wien) über ein Forschungsprojekt am Institut für Jüdische Geschichte Österreichs zum Thema „*Den neuen Mietern zur treuhändigen Verwaltung übergeben*“. *Die Räumungen von Judenwohnungen im Auftrag von Vugesta und Zentralstelle*“. Für die jüdische Bevölkerung wurden sogenannte Sammelwohnungen zur Verfügung gestellt. Bekanntestes Beispiel hierfür ist die Wohnung von Sigmund Freud in der Berggasse 19, in der 1942 16 Personen untergebracht waren. Für die Vergabe von Lebensmittelkarten an Juden war deren Erfassung in Verzeichnissen Voraussetzung; diese bildeten auch die Basis für die Deportationen. Die Hauslisten werden im Archiv der Israelitischen Kultusgemeinde (IKG) aufbewahrt; die IKG nahm auch die Wohnungsräumungen der Deportierten vor.

Der abschließende Tag der Tagung am Donnerstag, den 4. Mai 2017, begann mit dem von Pia Schölnberger (Wien) moderierten Panel zum Thema „*München*“.

**Meike Hopp** (München) stellte in ihrem Vortrag „*Die Aquarelle und Zeichnungen Rudolf von Alts aus der ehemaligen ‚Sammlung‘ Martin Bormanns an der Staatlichen Graphischen Sammlung München*“ schon durch die Thematisierung des Künstlers Rudolf von Alt (1812–1905), der nicht zuletzt für seine Wien-Ansichten bekannt ist, eine Verbindung zum Tagungsort her. Sie schilderte die Aufarbeitung der sogenannten „Alt-Aktion“, bei der Reichsleiter Martin Bormann über Mittelsmänner unmittelbar nach dem Anschluss 1938 über 800 Blätter des Malers erwerben ließ, die als „ehemaliger Besitz der NSDAP“ an die Staatliche Graphische Sammlung in München gingen. Während einige Werke später als Raubgut restituiert wurden, ist die Identifizierung und Zuordnung anderer oft sehr ähnlicher Werke schwierig.

**Johannes Gramlich** (München) setzte den München-Schwerpunkt mit seinem Vortrag „*Die Übertragung von Kunstgegenständen aus ehemaligem NS-Besitz an den Freistaat Bayern und die Bayerischen Staatsgemäldesammlungen – gesetzliche Grundlagen und individuelle Verantwortung*“ fort und schilderte die Übernahme von Kunstgegenständen aus NSDAP-Vorbesitz durch die Bayerischen Staatsgemäldesammlungen zunächst vom Bayerischen Ministerpräsidenten und später vom Münchner Central Collecting Point. Im Mittelpunkt standen Fragen nach den gesetzlichen Rahmenbedingungen, den leitenden Motiven, dem Einfluss der Alliierten und den Bedingungen treuhänderischer Übernahmen für das weitere Vorgehen zuständiger Stellen.

Zum Abschluss des Panels führte Stephan Kellner (München) in seinem Vortrag „*Überweisung der Alliierten: Der Bestand der NS-Ordensburg Sont-*

hofen in der Bayerischen Staatsbibliothek“ mit der Schilderung der NS-Ordensburgen als Ausbildungsstätten für NS-Kader mit unterschiedlichen Schwerpunkten über die Grenzen von München hinaus, bevor er speziell auf den Bücherbestand der Ordensburg Sonthofen einging. Dieser Bestand war als Zuweisung der Alliierten an die Bayerische Staatsbibliothek gekommen und teilweise restituiert, wobei die Prüfung der angemeldeten Ansprüche mitunter auch NS-Verstrickungen der Antragsteller zutage förderte.

Die weiteren Panels der Tagung widmeten sich Fragestellungen aus verschiedenen Ländern, wobei auch bisher in diesem Zusammenhang kaum beachtete Themen und Gebiete zur Sprache kamen. So gleich im Panel „*International I*“, moderiert von Murray G. Hall (Wien), das sehr unterschiedliche Perspektiven auf den Umgang mit Raubgut in Polen brachte.

Julia Stepnowska (Gdańsk) schilderte in ihrem gemeinsam mit Kamil Zeidler vorbereiteten Vortrag „*The case of Polish museums holding cultural objects ‘in trust’ after WWII*“ die Auseinandersetzung mit in öffentlichen Einrichtung gelangtem Raubgut und seinem Status als kulturelles Erbe, bevor sie auf Fallbeispiele dieser Abwägungen zwischen öffentlichen Ansprüchen und denen von Vorbesitzern wie etwa der Katholischen Kirche in Polen einging. Einige dieser Fälle haben in jüngerer Zeit auch die polnische Öffentlichkeit beschäftigt, sodass der Vortrag eine gute Zeit für die Provenienzforschung, für die Beschäftigung mit dem „nationalen“ Erbe sowie für die Erforschung von aus Polen geraubtem Kulturgut konzedierte.

Andere Erfahrungen und Ausgangslagen schilderte Nawojka Cieślińska-Lobkowicz (München/Warschau) in Ihrem Vortrag zum „*Umgang mit dem sog. ‚postjüdischen‘ Kulturgut in Polen von 1945 bis heute*“. Sie setzte sich mit Kulturgut auseinander, das polnischen Juden geraubt und großteils außer Landes gebracht wurde, aber auch mit in Polen verbliebenen und in öffentlichen Sammlungen und Bibliotheken „nationalisierten“ Beständen, die auch nach dem Ende des Kommunismus nicht restituiert oder in der öffentlichen Wahrnehmung hinterfragt wurden.

Ebenfalls mit Beständen aus jüdischem Vorbesitz beschäftigte sich Lara Lempertienė (Vilnius) im Vortrag „*LOOTED? ABANDONED? SAVED? The provenance and status of Jewish documents in the state document depositories of Lithuania*“. Sie schilderte den Status ehemals jüdischer Buch- und Sammlungsbestände in Bibliotheken und Archiven in Litauen, die trotz zwischenzeitlicher Überlegungen und Vereinbarungen teilweise unverändert an diesen Institutionen geblieben sind, und betonte die Wichtigkeit des internationalen Austauschs für den Fortschritt der Forschungs- und Restitutionsbemühungen.



Abb. 4: Jörn Kreuzer, Monika Schreiber, Ekaterina Oleshkevich (v.l.n.r.) (Foto: Marc Drews)

Das abschließende Panel der Tagung „*International II*“, moderiert von **Monika Schreiber** (Wien), beschäftigte sich weiter mit Fragestellungen aus verschiedenen Ländern und wandte sich nach den sammlungsübergreifenden Überlegungen des vorhergehenden Panels einzelnen Sammlungen zu.

Ekaterina Oleshkevich (Moskau) berichtete über die „*Schneerson collection as a case of looted and lost property: historical aspects and present situation*“. Die von der Rabbinerdynastie Schneerson aufgebaute und nach der Russischen Revolution enteignete Sammlung wurde an der Russischen Staatsbibliothek mit anderen Beständen vermischt, was Rückforderungen der Chabad-Bewegung aus den USA bis heute im Wege steht. Derzeit ist der Bestand im „Museum der Toleranz“ öffentlich zugänglich, während Fragen der rechtlichen Basis einer Restitution wie der Zuordnung der betroffenen Exemplare offen sind.

Einen ganz anders gelagerten, nämlich gewissermaßen positiv abgeschlossenen Fall schilderte **Jörn Kreuzer** (Hamburg) in seinem Vortrag „*Ein historischer „Glücksfall“? Warum Archiv und Bibliothek der Jüdischen Gemeinde Hamburg (wieder) in der Hansestadt sind*“. Während das Archiv der jüdischen Gemeinde früh als Depositum an das Hamburger Staatsarchiv übergeben wurde, war die Gemeindebibliothek nach Berlin beschlagnahmt worden, bevor sie von der Staats- und Universitätsbibliothek beansprucht wieder

nach Hamburg kam und so eine unbeabsichtigte, wenn auch nicht institutionelle Wiederzusammenführung gelang.

Die Tagung, die sowohl was den inhaltlichen und interdisziplinären Fokus als auch die internationale Vernetzung der NS-Provenienzforschung betrifft Neuland betreten hat, wurde von Sitzungen der „VÖB-Arbeitsgruppe NS-Provenienzforschung“ sowie des internationalen „Arbeitskreises Provenienzforschung und Restitution – Bibliotheken“ und nicht zuletzt von einer Ausstellung einiger der vom Nationalfonds wieder erworbenen Bücher der All Peoples Association und Gipsabdrücke aus der Ägyptologischen Sammlung an der Fachbereichsbibliothek Zeitgeschichte begleitet.

Mag. Dr. Stefan Alker-Windbichler, MSc  
Universität Wien, Bibliotheks- und Archivwesen  
E-Mail: [stefan.alker-windbichler@univie.ac.at](mailto:stefan.alker-windbichler@univie.ac.at)

Mag. Bruno Bauer  
ORCID: <http://orcid.org/0000-0002-4729-331X>  
Medizinische Universität Wien, Universitätsbibliothek  
E-Mail: [bruno.bauer@meduniwien.ac.at](mailto:bruno.bauer@meduniwien.ac.at)



Dieses Werk ist lizenziert unter einer  
Creative-Commons-Lizenz Namensnennung 4.0 International

■ **BERTOT, JOHN CARLO [ED. ET AL.]: ACCESSIBILITY FOR PERSONS WITH DISABILITIES AND THE INCLUSIVE FUTURE OF LIBRARIES. EMERALD PUBLISHING GROUP 2015 (ADVANCES IN LIBRARIANSHIP, VOLUME 40). ISBN 978-1785-606533**

Barrierefreiheit in Bibliotheken ist nicht nur essentiell, um Personen mit den verschiedensten Bedürfnissen den Zugang zu digitalen Kompetenzen, (online) Bildung und vielem mehr zu ermöglichen. Sie ist außerdem gesetzlich verankert. Nicht verwunderlich also, dass dieses Thema auch Eingang in eine Bücherreihe – nämlich den *Advances in Librarianship* – fand, um diesbezüglich auf 249 Seiten 1) Best-Practices und innovative Ideen, 2) künftige Herausforderungen sowie 3) soziale Implikationen anhand von Fallbeispielen zu diskutieren. Das Buch gliedert sich außerdem in diese drei Themenbereiche, denen jeweils einige Artikel zugeordnet sind.



Bereits ein Blick ins Inhaltsverzeichnis und auf die einleitenden Seiten jedes Artikels erfreut. Ersteres ist übersichtlich. Letztere enthalten wesentliche Informationen zu den vorliegenden Aufsätzen: AutorInnen werden in Stichworten vorgestellt, was die Kontextualisierung ihrer Beiträge erleichtert. Zudem sind in einem kurzen Abstract das Ziel der Arbeit, Methodologie und Forschungsansatz und wichtige Resultate angegeben. Wer also nicht viel Zeit hat oder nach konkreten Inhalten sucht, wird hier schnell fündig.

Auch jenen LeserInnen, die grundlegendes Interesse an der Thematik und Zeit haben, sei dieser Sammelband ans Herz gelegt. Das Buch ist, wenn an manchen Stellen auch etwas trocken, gut, also verständlich und nachvollziehbar, geschrieben. Obwohl sich die Artikel im Wesentlichen auf Fallbeispiele in den USA beziehen, lohnt es, zu schmökern und sich mit ihnen zu beschäftigen: Wie barrierefrei sind wir tatsächlich? Informieren wir ausreichend über unsere Serviceangebote? Nutzen wir hierzu professi-

onelles Marketing? Welche Technologien erschweren oder erleichtern den allgemeinen Zugang?, etc.

Zwar sind all diese Fragen keineswegs neu, doch werden sie auf konkrete Beispiele bezogen. Beschrieben werden dabei auch Lösungsansätze verschiedener Büchereien. Möglicherweise sind diese nach genauer Lektüre sogar Inspiration genug, um im Hinblick auf Barrierefreiheit in der eigenen Bibliothek wieder einige Schritte weiterzudenken und zu handeln.

Katharina Flicker, Wien



Dieses Werk ist lizenziert unter einer

[Creative-Commons-Lizenz Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

■ **HANNAH FREY: „ZUCKERFREI – DIE 40 TAGE CHALLENGE“.**  
**GU VERLAG 2017. ISBN 9783833859342**

Nicht nur in der Fastenzeit wird uns immer wieder bewusst gemacht, dass Zucker als ungesund, wenn nicht sogar als Gift, angesehen werden sollte. Doch Zucker wirklich strikt aus seiner Ernährung zu streichen ist nicht so leicht wie man annehmen möchte. In vielen, vor allem industriell hergestellten „convenience-Produkten“, stecken Zuckerportionen, wie beispielsweise in Salatdressings oder in Gewürzgurken aus dem Glas.

Hannah Frey, Gesundheitswissenschaftlerin und Ernährungsexpertin aus Hamburg, hat jedoch ein Buch für den Einstieg in das zuckerfreie Leben herausgebracht: „Zuckerfrei – Die 40 Tage Challenge“

erklärt im vorderen Teil des Buchs einfach, aber sehr genau, als was Zucker „verkleidet“ sein kann und was dieser mit uns macht. Ebenso gibt sie Tipps, wie man die 40 Tage gut durchhält und erläutert das 2 Phasen Konzept des Buchs: So sind die ersten 30 Tage die sogenannte „Reset-Phase“ bei der es um einen „richtigen Cut“ der Gewohnheiten geht und man nicht nur auf Zucker und alternative Süßungsmittel verzichten soll, sondern auch auf stärkehaltiges Gemüse wie Kartoffeln, Mais oder Kürbis. Die meisten Getreide sind hierbei auch nicht „erlaubt“, aber bestimmtes Obst ist dafür im Ernährungsplan enthalten. Die letzten 10 Tage sind die „Restart-Phase“ und hier sind unter anderem auch wieder Vollkornprodukte wie Nudeln oder das stärkehaltige Gemüse erlaubt.

Das klingt vielleicht im ersten Moment relativ hart, aber wenn man den Schritt wagt geht es leichter als man denken möchte und nach einer kurzen Entzugsphase fühlt man sich, neben vielen anderen positiven Veränderungen, energiegeladener und wacher.

In dem Buch, welches im Gräfe und Unzer Verlag erschienen ist, findet man eine sehr große Anzahl an schmackhaften und einfachen Rezepten, womit einem der Schritt aus der Komfortzone und hin in ein zuckerfreies Leben genussvoll und leicht gestaltet wird.



Marlene Kersten, Wien



### ■ **FORCE2017 – THE FUTURE OF RESEARCH COMMUNICATIONS AND E-SCHOLARSHIP (BERLIN, 25.–27. OKTOBER 2017)**

Ort: Kalkscheune, Berlin, Deutschland

Website: <https://www.force2017.org/>

### ■ **„LIZENZANGABEN UND RECHTEDOKUMENTATIONEN IM DIALOG – DATENFLÜSSE NACHHALTIG GESTALTEN“ (FRANKFURT AM MAIN, 7. NOVEMBER 2017)**

Ort: Deutsche Nationalbibliothek, Frankfurt am Main

Website: <http://www.dnb.de/DE/Wir/StrategieInnovation/lizenzenImDialog.html>

### ■ **ODOK/INETBIB 2018 (WIEN, 21.–23. FEBRUAR 2018)**

Ort: Juridicum der Universität Wien

Website: <https://www.inetbib.de/>

### ■ **DHd 2018 (KÖLN, 26. FEBRUAR–2. MÄRZ 2018)**

Leitthema: „Kritik der digitalen Vernunft“

Ort: Universität Köln

Website: <http://dhd2018.uni-koeln.de/>

### ■ **D-A-CH-S-TAGUNG DER BIBLIOTHEKARISCHEN VERBÄNDE (MÜNCHEN, 7.–9. MÄRZ 2018)**

Veranstalter: VDB, BIB, VÖB, BVS, BIS, IG WBS

Ort: Ludwig-Maximilians-Universität München

### ■ **107. DEUTSCHER BIBLIOTHEKARTAG (BERLIN, 12.–15. JUNI 2018)**

Ort: Estrel Congress Center, Berlin

Website: <https://bibliothekartag2018.de/>