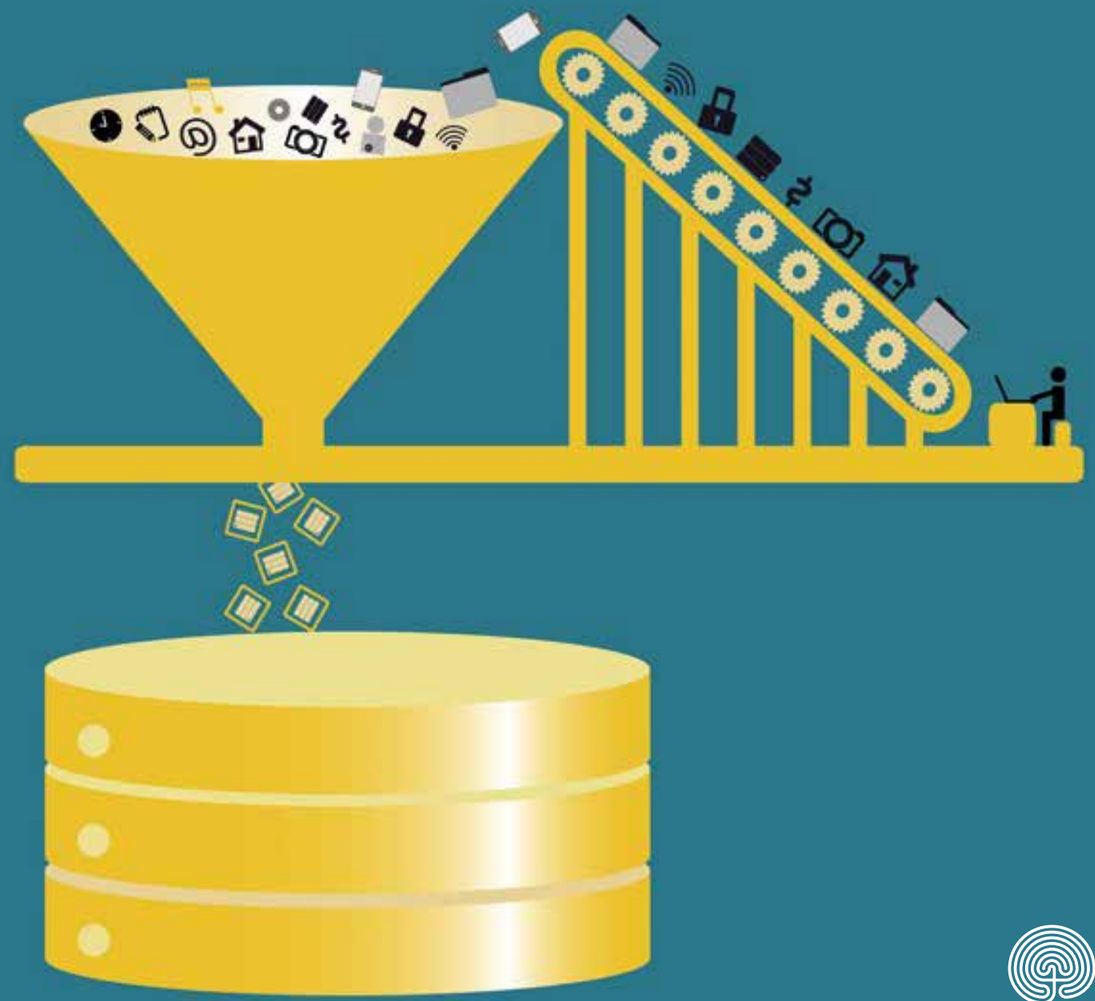
Vol. 10, N° 1, 2015

Akademien der Wissenschaften Schweiz

swiss academies reports

www.akademien-schweiz.ch

Final report for the pilot project “Data and Service Center for the Humanities” (DaSCH)



Swiss Academies Reports, Vol. 10, N° 1, 2015

Final report for the pilot project “Data and Service Center for the Humanities” (DaSCH)

Executive summary

This report describes the main lessons learned from the pilot phase we have carried out with existing real-world humanities research projects. Above all, it has become clear that no single, monolithic virtual research environment (VRE) with a generic user interface can meet the needs of all projects. A significant number of projects require an open, extensible platform that enables them to choose the functionality that is relevant to their needs, to create custom user interfaces and client applications, and to integrate new research tools into their workflow. The new Knora architecture has been conceived from the ground up to meet these requirements. Salsah still exists as one component of Knora, providing a powerful, but optional, virtual research environment with a generic user interface.

The pilot project has enabled us to demonstrate, using real research data, the advantages and disadvantages of using different software technologies to solve the particular problems at hand. The knowledge we have gained from this experience gives us confidence that Knora will be flexible enough to meet the needs of a wide variety of users, and that it will offer the performance and reliability expected of a crucial national infrastructure.

The report emphasizes our links with international research projects and our focus on interoperability and open standards, since Knora aims to interface as much as possible with related initiatives elsewhere. Moreover, fundamental research is closely related to research infrastructure, and all the results of our pilot project sustain the key idea at the heart of the SUC P2 program. Finally, we also discuss the need to train researchers to enable them to take full advantage of Knora. Indeed, a national project in DH infrastructure must be supported by a bottom-up effort and by the dissemination of knowledge, to enable the largest possible number of researchers to use it. In other words, in the digital culture, a VRE centre has to develop close links with research and education challenges.

Die Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW) vermittelt, vernetzt und fördert die geistes- und sozialwissenschaftliche Forschung in der Schweiz. Ihr gehören 60 Fachgesellschaften und rund 20 Kommissionen an und sie leitet mehrere grosse Forschungsunternehmen. Sie versteht sich als Mittlerin zwischen Forschenden und wissenschaftlich interessierten Personen einerseits und politischen EntscheidungsträgerInnen, Behörden und einer breiteren Öffentlichkeit andererseits. Die SAGW verfügt über ein Budget von rund 10 Millionen Franken und wird von einem Vorstand mit 18 Mitgliedern aus Wissenschaft, Politik und Verwaltung geleitet. Im Generalsekretariat arbeiten 13 Mitarbeiterinnen und Mitarbeiter.

Swiss Academies Reports, Vol. 10, N° 1, 2015

Final report for the pilot project "Data and Service Center for the Humanities" (DaSCH)

Schweizerische Akademie der Geistes- und Sozialwissenschaften
Académie suisse des sciences humaines et sociales
Accademia svizzera di scienze umane e sociali
Academia svizra da scienzas umanas e socialas
Swiss Academy of Humanities and Social Sciences

Durch den Vorstand der SAGW am
27. Februar 2015 und durch
die Kommission DDZ am 23. März 2015
verabschiedet und genehmigt.

Bern, 30. März 2015

Herausgeberin
Schweizerische Akademie der Geistes- und Sozialwissenschaften,
Haus der Akademien, Laupenstrasse 7, Postfach, 3001 Bern
Telefon +41 (0)31 306 92 50, sagw@sagw.ch
www.sagw.ch

ISSN 2297-1564

Gestaltung
Druck- und Werbebegleitung, 3098 Köniz

Foto Umschlag
© Koollapan - Dreamstime.com

Autorenschaft
Pilotprojektgruppe DDZ

Druck
Druck- und Werbebegleitung, 3098 Köniz

1. Auflage, 2015 (600 Expl.)
Die Broschüre kann kostenlos bezogen werden bei der SAGW.

Zitiervorschlag:
Schweizerische Akademie der Geistes- und Sozialwissenschaften (2015)
Final report for the pilot project "Data and Service Center for the Humanities" (DaSCH)
Swiss Academies Reports 10 (1).

Table of contents

1. Aim of the pilot project (Data and Service Center for the Humanities, DaSCH) 5
1.1 Starting point / Call for bids 5
1.2 Description of the problem 5
1.2.1 Demand for data curation and support service 7
1.3 Situation in Switzerland and in an international context 7
1.3.1 FORS 7
1.3.2 Swiss Federal Archives (SFA) 8
1.3.3 Relation with SUC P2 program "Scientific information: access, processing and safeguarding" 8
1.3.4 International contacts 8
1.4 Pilot phase objectives 9
1.5 Coverage of the pilot phase 9
1.5.1 Main goals 9
1.5.2 Secondary goals 10

2. Approach of the consortium 12
2.1 Presentation of the consortium 12
2.2 Organizational approach during the pilot phase 13
2.3 Features and technical approach 13
2.3.1 Long-term preservation of digital information 13
2.3.2 Research data in the Humanities, and longevity 16
2.3.3 Data transfer model 16
2.3.4 Data services 17
2.3.5 System design and technology 17

3. Experiences from the pilot phase 20
3.1 Organization 20
3.1.1 Bern 20
3.1.2 Basel 20
3.2 Technical platform 20
3.3 Importing data 21
3.4 Services and support 22
3.4.1 List of services provided 22
3.4.2 Digital editions 23
3.5 Technical platform 23
3.5.1 Data modelling 23
3.5.2 Access 24
3.5.3 Reusing and adding value to existing data 24
3.6 Financial aspects 24
3.7 Further activities during the pilot phase 24
3.7.1 The DaSCH as a national point of contact for full Swiss membership of DARIAH 24
3.7.2 Swiss National Research Infrastructures 24
3.7.3 Collaborations beyond the humanities 25

4. International comparison 26
4.1 Data Center for the Humanities at the University of Cologne (DCH) and Cologne Center for eHumanities (CCeH) 26
4.2 Data Archiving and Networked Services (DANS) 27
4.3 UK Data Archive 28
4.4 TextGrid (D) 28
4.5 Geisteswissenschaftliches Asset Management System GAMS (Graz) 29
4.6 TGIR Huma-Num (France) 30
4.7 OpenEdition 31
4.8 Results of the international comparison 31

→

5. Implications for the implementation of the DaSCH 32
5.1 Organization and governance 32
5.1.1 Organization 32
5.1.2 Governance 33
5.2 Financial aspects 35
5.2.1 Estimate of full cost of national coordination unit including long-term archiving 35
5.2.2 Estimate of full cost of a satellite 36
5.2.3 Further costs 36
5.2.4 Summary of effective costs 36
5.3 Technical infrastructure 36
5.3.1 Knora/SALSAH platform 36
5.3.2 Computer infrastructure 37

6. Overall conclusions 38

7. Annex 41
A. Model for embedding "satellites" (example Lausanne) 41
B. Detailed description of test cases 44
C. Selected further reading on the DaSCH topic from members ot the pilot project (published 2013–2015) 62
D. Legal documents 64

Figures

Figure 1: Scheme of the OAIS reference model 15
Figure 2: Architecture of the Knora/SALSAH platform 19
Figure 3: Steps of a typical ingest process into the Knora platform 22
Figure 4: Proposed organisation of the DaSCH as a network of satellites and a central development unit also providing second level support 33
Figure 5: Governance of the DaSCH 35

Final report for the pilot project "Data and Service Center for the Humanities" (DaSCH)

1. Aim of the pilot project (Data and Service Center for the Humanities, DaSCH)

1.1 Starting point / Call for bids

By a mandate of the State Secretariat for Education, Research and Innovation (SERI), the Swiss Academy of Humanities and Social Sciences (SAHSS) published a call for bids for a pilot project for a "Data and Service Center for research data in the humanities" on 2 April 2013. The deadline for bids was 15 May 2013.

In the description of the pilot project, the primary and secondary goals were described as follows [1]:

Primary goals:
- Preservation of research data in the humanities, and long-term data curation.
- Ensuring permanent access to research data in order to make it available for further research. This facilitates the reuse of existing research data in future research.
- Providing services for researchers to assist them with data life cycle management.

Secondary goals:
- Encouraging the digital networking of databases created in Switzerland or in other countries.
- Carrying out a pilot project in close proximity to humanities research.
- Collaboration and networking with other institutions on digital literacy.
- Preparation of the pilot project to become a national point of contact for the Swiss representation in DARIAH.

A consortium of the Universities of Basel, Bern and Lausanne under the leadership of the DHLab of the University of Basel (Professor L. Rosenthaler) submitted a bid, and learned at the end of June 2015 that its bid had been successful and that the project would start on 1 July 2013. The pilot project was due to last for 2 years, until 30 June 2015. The total budget available from the SERI is CHF 600,000 (300,000 p.a.) The consortium's bid includes a total budget of CHF 1 million, of which the Universities of Basel, Bern and Lausanne contribute a total of CHF 400,000. The consortium officially started work on 1 July 2013. In 2015, an additional sum of CHF 30,000 was granted to UNIL for developing an IT mandate, with a CHF 24,000 contribution from UNIL.

1.2 Description of the problem

The humanities have been transformed by digital methods as the internet and its technologies have become commonplace in society and in research. Since the 1990s, the digitization of manuscripts, photographs, sound recordings and films, and the transcription of text corpora, has made many important sources directly available on the desks of humanities researchers. This development has certainly made for more efficient scholarly practice and created opportunities for entirely new research directions.

Thanks to the availability of large quantities of digitized sources on the internet, and the existence of mature technologies such as the Semantic Web, humanities research is on the verge of a fundamental change in research methods. The emergence of the new label "Digital Humanities" (DH) in the 2000s reflects increasing awareness of this change [2]. However, the use of computer-based methods and online sources in the humanities still faces several challenges, including the difficulty of ensuring the longevity of research data, the lack of common basic services, inadequate standardization of data formats, insufficient training in digital methods and best practices, and weak international Digital Humanities networks. Nevertheless, a great deal of state-of-the-art research in the humanities relies on digital data. Digital documents are accumulated, organized and annotated using electronic databases. However, this infrastructure is most often established in a project-specific way, and is not designed for

1 http://www.sagw.ch/de/dms/sagw/laufende_projekte/DDZ/Ausschreibung_dt/Ausschreibung_Ergaenzungen_def_dt/ Ausschreibung_Erg%C3%A4nzungen_def_dt.pdf, Chapter 3 (page 3).

2 See C. Clivaz, "Common Era 2.0. Mapping the Digital Era from Antiquity and Modernity", in Reading Tomorrow. From Ancient Manuscripts to the Digital Era / Lire Demain. Des manuscrits antiques à l'ère digitale, C. Clivaz – J. Meizoz – F. Vallotton – J. Verheyden (eds.), with Benjamin Bertho, Lausanne: PPUR, 2012, ebook, p. 23–60.

the long-term preservation of data. After the completion of a research project, these digital resources quickly become unavailable if they, and the software and hardware they rely on, are not properly maintained. Keeping digital data accessible after the end of a project is costly in terms of money and labor, and these costs are usually not included in the project funding.

While the digitization of sources produces large numbers of digital documents, these documents usually have a simple structure. There are also institutions with a long-term interest in preserving access to these assets (mainly libraries, archives, universities, and other state-funded institutions that own the original analogue sources). By contrast, the data produced during the research process is much more complex, consisting of interlinked data, such as databases, annotations, and heterogeneous collections of digital information. Because of the complexity of this research data, it is very difficult to make it permanently available. However, there are several reasons for doing so:

Transparency: as research data is the foundation on which published results are based, it becomes necessary to have access to this data in order to evaluate the results.

Reuse: new research projects can reuse existing research data to propose different answers to the same questions, or to ask entirely new questions, especially if the datasets from different projects can be linked.

Citability: digital sources may only be referenced in scientific texts if they can be accessed permanently without modification[3]. At the moment this is only possible for e-publications that are curated by a library or a publisher. The long-term accessibility of arbitrary digital objects (together with permanent links and unique object identifiers) will allow the scientifically correct citation of those objects represented in the platform.

While there are many institutions with an excellent working knowledge of long-term archiving of digital data, e.g. the Swiss Federal Archives[4] (SFA), archives of the cantons, university libraries, etc., most of them are not well prepared for the required task. On one hand, a very close collaboration with researchers from different projects is necessary. This requires direct knowledge and hands-on experiences in humanities research, which none of these institutions have. On the other hand, research often uses advanced technology and methods that have not yet been standardized. The platform must therefore keep up with these advances in methods and technology. This is best achieved when the platform is closely connected to the institutions where research is carried out.

Within universities, it has only recently been acknowledged that humanities research needs a digital infrastructure and specialized IT support, and universities are not yet well prepared for this task. Usually, university IT departments are not sufficiently informed about the needs of humanities researchers. In the natural sciences, where the need for digital research infrastructure and research data repositories has long been recognized, and where researchers are usually more experienced in using computer technology, some local solutions (e.g. the Center for Scientific Computing[5], sciCore, at the University of Basel) have been implemented, some of them as networks (e.g. Vital-IT, a network of the Universities Lausanne, Geneva, Bern and the EPFL)[6].

It is possible that a nationwide network of digital infrastructure for humanities research, built on a shared technological foundation and fulfilling a common objective, will be established. Nevertheless, at least at the end of the life cycle of a research project, the digital data should be transferred to an institution that will keep it accessible for future use. We consider such institutions to be prime examples of *research infrastructures* for the humanities. The European Commission defines research infrastructures as follows:

"The term 'research infrastructures' refers to facilities, resources and related services used by the scientific community to conduct top-level research in their respective fields, ranging from social sciences to astronomy, genomics to nanotechnologies. […] RIs may be 'single-sited' (a single resource at a single location), 'distributed' (a network of distributed resources), or 'virtual' (the service is provided electronically)[7]".

In contrast to the natural sciences, which often use large machines or experimental platforms such as the CERN and/or high-performance computing and data processing capabilities for research infrastructure, the situation is somewhat different in the humanities. Humanities research is most often based on qualitative methods, which require complex digital models of information representation. As noted above, this information should remain accessible well beyond the end of a project. In fact, research data in the humanities often remains useful for decades, and in some cases for much longer, while data in natural sciences usually becomes irrelevant within a few months or years. Therefore, research infrastructure for the humanities

3 In addition, time-dependent media also need the time code identifying the location within the referenced stream (e.g. video of interviews).

4 As the SFA has excellent knowledge in this field, the DaSCH cooperates closely with the SFA for long-term archiving. One of the goals is to implement an interface to the digital archive of the SFA.

5 http://scicore.unibas.ch

6 http://www.vital-it.ch

7 https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=what

must be organized and funded in a way that provides for the indefinite longevity and accessibility of research data.

1.2.1 Demand for data curation and support service

A short survey carried out at the beginning of the pilot phase showed that there is a huge demand for the services provided by the DaSCH. Due to the limited resources available for the survey it is certainly not complete and exhausting, but it pointed out that there were already over 50 data collections in existence. In the meantime, many more research projects have started that will use databases etc. and collect digital information. There are many more support enquiries than we can handle with the available resources. We expect the demand to rise considerably in the near future because it is becoming increasingly necessary to apply digital methods to research in the humanities.

The call for editions by the SNSF for the financing period 2017 to 2020 has also resulted in a peak of inquiries for support and services from the DaSCH. Furthermore, several permanent research infrastructure projects, such as the Historical Dictionary of Switzerland, the Collection of Swiss Law Sources, the Schweizerisches Idiotikon and the Swiss Text Corpus, have expressed a vital need for services or an interest in close cooperation with the DaSCH. It became clear that a dropout of the facilities and services of the pilot project would cause some serious problems within DH projects in Switzerland.

1.3 Situation in Switzerland and in an international context

At the time of writing (early 2015), there is no systematic approach to ensuring long-term access to digital research data in the humanities on a national level in Switzerland. No institution is able to take over this task, and it is up to individual researchers to find a solution. The following institutions represent the options that are currently available:

Universities
As noted above, the universities are not well prepared for this task. IT services or university libraries may provide some support, but often only provide repositories for e-publications[8]. However, libraries that digitize books and documents are well prepared to guarantee long-term access to those types of digital assets.

National Archives
The National Archives also provide an archival service for digital data, primarily for the federal administration. Due to their legal mandate, the National Archives are providing long-term preservation services for all types of documents and data relevant in this context.

National Library
While the National Library[9] is playing a leading role in providing access to digital books and documents, the task of maintaining long-term access to complex research data is not part of its remit.

Since there are neither local nor national institutions that take care of this urgent problem, it is obvious that the most efficient and cost-effective way is to establish a national coordination unit. However, it should be noted that Switzerland is a federalist, multilingual country that is not suitable for a centralized approach. There are 10 full universities that are financed and governed by the cantons (the Universities of Basel, Bern, Fribourg, Genève, Lausanne, Luzern, Neuchâtel, St. Gallen, Zürich, and the Università della Svizzera Italiana), two that are financed by the Swiss Confederation (ETHZ, EPFL), and seven regional universities of applied sciences.

1.3.1 FORS

In the social sciences, FORS, which was founded in 2008, is responsible for long-term access to research data. It can be considered as the model type of institution that is needed in order to fulfil the potential of digital research methods. FORS describes its mandate as follows:

FORS is a national center of expertise in the social sciences. Its primary activities consist of:

- Production of survey data, including national and international surveys
- Preservation and dissemination of data for use in secondary analysis
- Research in empirical social sciences, with focus on survey methodology
- Consulting services for researchers in Switzerland and abroad.

FORS collaborates with researchers and research institutes in the social sciences in Switzerland and internationally[10].
Of course, the mandate of a similar institution for the humanities would have to be adapted considerably.

8 In this context, e-publications are usually just PDF files, i.e. a digital simulation of printed text. Besides offering web-oriented hyperlinks and a full-text search, publication in PDF format does not differ from printed publication.

9 In conjunction with the university libraries.

10 See http://forscenter.ch/en/about-us-2/mandate/

1.3.2 Swiss Federal Archives (SFA)

The Swiss Federal Archives are an important partner for the DaSCH. Digital assets managed by the SFA, such as text, videos, images, or databases, etc. can either be born digitally, or they can be digital copies of the original items. We are currently working on an interface to directly transfer selected datasets from the DaSCH to the digital long-term preservation solution provided by the SFA. Thus, relevant research datasets that will not be changed again can be transferred to the SFA's digital long-term repository using an automated and standardized process. The original research data therefore remains unchanged in the archives, whereas a digital copy can be ordered for reuse. In this way, the SFA protects the original digital assets against losses and guarantees that the research project will always be verifiable.

On the other hand, the SFA and the DaSCH will explore the possibility of adapting and using the generic front end of the DaSCH platform to access the SFA archive as well. We consider this collaboration with the SFA to be essential, as the knowledge bases and experiences of the two institutions are complementary and there will be a mutual benefit.

1.3.3 Relation with SUC P2 program "Scientific information: access, processing and safeguarding"

The national strategy adopted by the CRUS in April 2014 is described as follows:

"The CRUS is adopting a new approach to dealing with scientific information. The aim is to make scientific information a domain in which Swiss universities meet requirements together instead of competing with one another. Targeted funding of collaborative projects is designed to help strengthen the Swiss scientific community's position in the face of international competition[11]." The DaSCH is in line with the SUC P2 program presented in the national strategy (see above) and the white paper[12]. This should help the DaSCH to obtain additional external funding and to collaborate with other Swiss IT research infrastructure projects. Moreover, the DaSCH seeks making use of the coordinating task and aim of the program, as this is the place where many related activities come together. However, a stable source of long-term funding has to be secured as the DaSCH has to guarantee access to research data with a time frame of more than 10 years.

So far, the following collaboration with SUC P2 projects are established or planned:

ORD@CH, led by FORS (UNIL) (funding secured until the end of 2016). The DHLab as head office of the DaSCH makes available the data hosted in the repository, in the context of the pilot phase of the "Open Research Data" platform (as far as legally possible).

Data Life Cycle Management (DLCM) (to be submitted in February 2015, for 2015–2018):
It is planned that the DaSCH (represented by the Universities of Lausanne and Basel) will collaborate closely and be part of the DLCM project that will be submitted to the SUC in February 2015. Seven institutions including UNIL and Unibas support the DLCM project. A close cooperation between the DaSCH and the DLCM is planned at each location (in Lausanne with VITAL-IT, in Basel with sciCore). The DLCM project has already received CHF 400,000 for its preparation, with a 50% position at UNIL for six months, working on the fields of DH, Life Sciences and Bioinformatics.

Other external funding has to be obtained: members of the team apply for SNSF projects, but we also need to collaborate with international projects to support basic research and to address challenges in the research infrastructure (see par. 1.3.4).

1.3.4 International contacts

Since the field of Digital Humanities is quite new in Switzerland – at least under this label – collaborations between Swiss researchers and international researchers on digital research infrastructures for the humanities have yet to be developed in Switzerland. The team working on the pilot project has already achieved some encouraging results in this area (see below, notably EU and US projects). This point is particularly important in order to secure Swiss DH research interoperability and DH research developments, as research infrastructure and research are interrelated in DH. The following formal contacts have been established:

ERIC DARIAH, ERASMUS+, RI Horizon 20/20
Contacts are running at different levels, and four Swiss universities have become cooperating partners of DARIAH. A workshop was co-organized at the DH 2014 (Austria, Switzerland, Serbia) on digital pedagogical innovations[13]. Professor Lukas Rosenthaler and the team presented an extensive paper on the DaSCH at the DH2014.

11 http://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/EN/UH/SUK_P-2/SUK_P-2_NationaleStrategie_20140403_EN.pdf

12 http://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/EN/UH/SUK_P-2/WhitePaper_V1.0-EN.pdf

13 http://informationsmodellierung.uni-graz.at/de/aktuelles/dariah-workshop-dh-2014/

In order to expand our international research impact (and to get external funding), we have received an ERASMUS+ DH grant (Professor Claire Clivaz, UNIL with seven other countries) to build DH modules of reference of teaching project (30 months, start Jan. 2015). On 14 January, we (i.e. the three consortium partners UNIL, Unibe and Unibas) jointly submitted a research infrastructure Horizon 20/20 project with UNIL taking the lead within the Swiss partners in textual scholarship for the development of a Virtual Research Environment (Professor Claire Clivaz, Professor Tara Andrews and Dr Ivan Subotic, €435,000 has been requested for the Swiss team).

Digital Humanities im deutschsprachigen Raum (DHd)
The DaSCH is very actively participating in the working group "data centres" (Arbeitsgruppe "Aufbau von Datenzentren") and has hosted one meeting in Basel in December 2014.

Humanistica and EADH (European Association of Digital Humanities)
The team has been involved since the beginning of the creation of the French-speaking association, founded in summer 2014. Professor Claire Clivaz is elected member of the steering committee. Professor Claire Clivaz is also an elected member of the EADH executive committee.

Harvard Library, Boston
We have a regular exchange with the "Preservation and Digital Imaging Services" of the Harvard Library, Boston.

Center for Hellenic Studies (CHS), Washington (Harvard University)
Collaboration in context with the LIMC[14] project. It is planned that in 2015 the CHS will implement a platform node using the platform provided by the DaSCH. The CHS has a powerful software development team and will bring new methods and tools.

ADHO (Association of DH organizations)
UNIL welcomed with the EPFL the DH2014 in summer 2014. A strong relationship has been developed with ADHO. Professor Claire Clivaz is vice-chair of the Conference Coordinating Committee of ADHO[15].

1.4 Pilot phase objectives

In the response to the call by the consortium of the Universities of Lausanne, Bern and Basel (led by the DHLab at the University of Basel), a major point was to base the pilot phase not only on theoretical and conceptual work, but on a full implementation of a platform using a limited set of real test cases. The work with real data from selected test cases is important for several reasons:
- To present a proof of concept
- To estimate the effort required and determine specific requirements
- To check and, if necessary, adapt the organizational structure
- To test, adapt, and further develop the technical platform
- To test and develop procedures.

The pilot phase has a duration of two years. It started in July 2013 and is due for completion at the end of June 2015. The reader should therefore keep in mind that this report describes the current state of a project that is still in development. In particular, many of the test cases, which we process in parallel in order to have a broad base of experience, are not yet fully completed. Nevertheless, this report provides an initial assessment of experiences that can be used as basis for a decision concerning further action.

1.5 Coverage of the pilot phase

This section will give a brief description of what has been done and what could not be done during the pilot phase. It will follow the outline of the goals given in the call for proposals:

1.5.1 Main goals

Long-term curation of research data
In order to prove the viability of the proposed concept for long-term access to research data, it is clear that a time span of 1.5 years from the beginning of the pilot project until now is not enough to prove longevity. However, we are demonstrating that a major change in the technical base from a MySQL/PHP-based RDF model to a true RDF triple store using Java/Scala is feasible without major hurdles.
The use of JPEG2000 as a common format for digital images and facsimiles has been extremely successful. The on-the-fly conversion to the desired format has far exceeded our expectations and is very efficient. Other groups (e.g. SUC P2 DLCM, Basel) have shown interest in using our software to implement the same concept.

Permanent access and reuse
The aim of permanent access has been achieved. On one hand, all of the digital research information integrated into the platform to date is accessible through a generic, web-based interface. On the other hand, a RESTful Application Programming Interface (API) facilitates the integration of research data into new research projects and into the

14 Lexicon Iconographicum Mythologiæ Classicæ.
15 http://adho.org/administration/conference-coordinating

presentation of results to the general public. Several projects for data reuse have materialized within the short period during which this information has been accessible:

Documentation Library of St. Moritz / Institute of Landscape Architecture ETHZ
Status: in progress
For a research project called "4D Sites", the Institute of Landscape Architecture of ETHZ uses the RESTful API to get images and metadata from the digital photo library of the Documentation Library of St. Moritz, which has been fully integrated into the platform.

Lexicon Iconographicum Mythologiæ Classicæ / Center for Hellenic Studies, Harvard University
Status: planned
The Lexicon Iconographicum Mythologiæ Classicæ (LIMC) is a large and extremely complex database, which has been developed over the last two decades. The project has officially come to an end, and the database – while still in use – has no longer been maintained for many years [16] and is not documented at all. We are currently working hard to import this data into the DaSCH platform. The Center for Hellenic Studies of Harvard University has expressed a very strong interest in accessing this data for reuse in a project focusing on Homer commentaries (Professor G. Nagy).

VitroCentre Romont / International community of stained glass inventories
Status: in progress
The VitroCentre in Romont decided to transfer and merge their existing FileMaker databases and digital images into the DaSCH platform. On one hand, this unification of several distinct, single-user FileMaker databases into a shared, web-based platform adds value to the data for the institution. At the same time, it makes it possible to share this data with the international community of stained glass inventories, where a very close collaboration has existed for many years. We aim to determine whether others may adopt certain procedures and formats used by the VitroCentre in Romont.

Services for researchers to support data life-cycle management
Support, service and trust are extremely important in order to convince researchers to consign their research data to a platform as provided by the DaSCH. The experience at all sites (Lausanne, Bern and Basel) has shown that it may take some time to build this trust. This difficulty is not specific to the Swiss pilot project. As the reviewers of the DH2014 paper on the DaSCH pilot noted, such difficulties have been seen in most international projects. One of the most convincing arguments is that the fine-grained access control offered by the platform means that the researchers are still in control of their data, but relieved of the burden of long-term storage.

The available resources limited the service offered by the DaSCH in the test cases used in the pilot phase. However, the dedicated staff of the DaSCH were able to provide valuable support in most cases.
Since not all researchers in the humanities are very proficient in the use of digital tools, a great deal of consulting and help is required when accessing and using the platform (especially in the beginning). However, researchers' reactions to the possibilities and prospects of using the platform and its tools have been very positive. Their constant feedback regarding improvements and new features helps the DaSCH to develop the platform according to their needs. In fact, close contact with researchers is a basic necessity. The example of the UK Arts and Humanities Data Service (AHDS) [17], which shut down in 2008, is a very striking one. It focused on technical challenges, without any concern for training researchers to use it, nor for obtaining diverse sources of funding. Our first experiments have shown the need for staff who can advise and train researchers, and who therefore must have skills both in the humanities (e.g. via a traditional humanities degree) and in IT.
We have begun to organize training days to train researchers to construct their database in the most efficient way. The first such event will take place at UNIL on 27 February 2015, under the leadership of Marion Rivoal and Claire Clivaz (LADHUL), in collaboration with Davide Picca (Arts and Humanities Faculty), Andréas Perret (FORS) and Frédéric Schütz (Wikimedia, copyright questions). We also are providing training in additional skills for researchers, such as data visualization, in collaboration with the SIB (on 13 March 2015, with Claire Clivaz and Martin Grandjean [LADHUL], and Frédéric Schütz, SIB [18]). Our capacity to provide this sort of training will be strengthened by our ERASMUS+ DH project, the purpose of which is to build digital modules to train researchers in DH from 2017 to 2020.

1.5.2 Secondary goals

The current status in relation to secondary goals is as follows:

Promoting the digital networking of databases created in Switzerland or in other countries
The project has received considerable attention in Switz-

16 The company that developed the database and software went bankrupt many years ago.

17 http://en.wikipedia.org/wiki/Arts_and_Humanities_Data_Service

18 http://edu.isb-sib.ch/course/view.php?id=195

erland and elsewhere. On an international level, a number of institutions in Europe and the US are keen to work with the DaSCH to develop such connections (e.g. University of Graz; Austrian Academy of Humanities, Wien; German Academy of Humanities, Berlin; Harvard University, Boston/Washington DC, etc.).
On a national level, the DaSCH has been very well received within the partner universities (Lausanne, Bern and Basel). In addition, transparent interfaces to e-manuscripta, e-rara and e-codices have already been developed. Connections to the Swiss Historic Lexicon and the "Sammlung Schweizerischer Rechtsquellen" are in the planning stage. Many more interactions are desirable but could not yet be established due to the limited resources of the DaSCH pilot project.

Carrying out a pilot project in close proximity to humanities research
The DaSCH has been very well received by Swiss humanities researchers. It has established itself as a point of contact for questions about the application of digital methods in humanities research. The Humboldt Edition (Professor Lubrich, MA Sarah Baertschi, University of Bern, Zurich University of Applied Sciences Winterthur, Dept. applied Linguistics, the Bernoulli-Euler Edition, e-codices, to name a few), seek the collaboration and support of the DaSCH. We have also had contacts with the computing center at the University of Geneva – a leader in the DLCM project for SUC P2 – and with the National Library, through the database "Artists and Books" (see 3.4.5 below). Contacts need to be developed further with Memoriav and other Swiss partners.

Collaboration and networking with other institutions on developing digital literacy
The DaSCH has been presented to researchers and interested parties through several talks at conferences, meetings and in interested institutions:
– Memoriav Kolloquium (Oct 2013)
– Workshop at the University of Lausanne (Oct 2013)
– Workshop at the University of Neuchâtel (Dec 2013)
– University of Lausanne, Conseil de Faculté des lettres (Jan 2014)
– Linked Data Workshop, Swiss Federal Archives (Jan 2014)
– Memoriav (Feb 2014)
– SERI (Feb 2014)
– EPFL (Mar 2014)
– EHESS, Paris (Mar 2014)
– University of Leuven, Belgium (Mar 2014)
– University of Zürich (Mar 2014)
– University of Lausanne, open meeting (Apr 2014)
– Université de Genève, meeting (Apr 2014)
– Archiving 2014, Berlin (2 talks, May 2014)
– MSH, Lille (May 2014)
– International AIPU meeting, Mons, Belgium (May 2014)
– Archives de l'Etat de Neuchâtel (June 2014)
– DHd AG Datenzentren, Berlin (June 2014)
– ENS, Lyon (June 2014)
– DH2014, Lausanne (4 talks, one workshop and a SNF round table) (July 2014)
– ZHAW Zürcher Hochschule für Angewandte Wissenschaften, Angewandte Linguistik (July 2014)
– Tagung Musikalische Inventare, Bern (Aug 2014)
– Opendata.ch Conference (Sep 2014)
– Hess Art Foundation, Napa, CA (USA) (Nov 2014)
– Cultural Heritage Imaging, San Francisco, CA (USA) (Nov 2014)
– DARIAH ERIC general assembly, Paris (Nov 2014)
– San Diego, USA (Nov 2014)
– EHESS, Paris (Dec 2014)
– DARIAH-FR Day[19], Paris (Dec 2014)
– Repositorien Workshop der Österreichen Akademie (Dec 2014)
– DHd AG Datenzentren, Basel (Dec 2014)
– Hearing "Infrastructure for Data Science in Switzerland", Board of the Swiss Federal Institutes of Technology (Dec 2014)
– DARIAH general meetings and VCC2 meetings (Berlin July 2013; Copenhagen December 2013; Athens March 2014; Rome September 2014; Paris December 2014).

A member of our team represents UNIL in the SUK P2 pilot committee. The LADHUL has organised several seminars about the databank at UNIL since 2014.

Preparation of the pilot project to become a "National Point of Contact" for the Swiss representation in DARIAH
A significant amount of effort has been put into the relationship with ERIC DARIAH, especially by our partner UNIL (Professor C. Clivaz), in keeping with the secondary objective of the project, and to develop our international relations in research. On 17 November, four Swiss universities (Basel, Bern, Geneva, and Lausanne) became "cooperating partners" of DARIAH, at the ERIC DARIAH first general assembly, on the basis of applications signed by the rectors. In Basel, the representative is Professor Lukas Rosenthaler; in Bern, Professor Tara Andrews; in Geneva, Dr Laure Ognois, and in Lausanne, Professor Claire Clivaz. This is a temporary status for two years, namely 2015 and 2016. During these two years, the DaSCH continues to make preparations to become the national point of contact. The contact with DARIAH has been very fruitful and the DaSCH has gained a great deal from this contact.

19 http://www.dariah.fr/rencontres-dariah

2. Approach of the consortium

2.1 Presentation of the consortium

Our team is composed of several researchers, with humanities (SHS[20]) or/and IT competencies. We are convinced that a research infrastructure is fostered by common research at the interface between DH and IT. The consortium consists of three institutions:

University of Basel, Digital Humanities Lab (DHLab)
The DHLab is a technology-based lab in the Faculty of Humanities with almost 20 years practice in interdisciplinary research. It has its roots in digital imaging (starting in 1982) and long-term preservation of digital data (since approx. 1990). All members of the team are directly involved in supporting researchers with their projects. In addition, they support DaSCH's other partners with technological issues related to the platform. The team working for the DaSCH in Basel[21] consists of the following members:

Professor L. Rosenthaler (head of the DHLab, which is leading the DaSCH consortium) began his career with a background in physics and applied computer science (computer vision). Since approx. 1985 he has been collaborating with the "Scientific Photography Lab", which became the DHLab when it moved from the Faculty of Science to the Faculty of Humanities. From 1992 to 2001, he worked as a software developer in industry (on CAD & GIS systems). He is DARIAH's representative for Basel University[22].
Dr P. Fornaro (deputy head of DHLab, deputy leader of DaSCH) has a background in engineering, physics, photography and business administration. He supports the DaSCH in organisation and financing[23].
Dr I. Subotic (software architecture, long-term archiving) completed a PhD in computer science on the subject of long-term digital archiving. Along with software development, he is also responsible for the IT infrastructure (two physical servers with virtualization, three large NAS with a total of around 70 TB, backup and security strategies)[24].
Dr B. Geer completed a PhD in Middle East Studies and worked as a software developer in industry, mainly in the banking sector. Together with T. Schweizer, he is one of the main software developers.
Dr. des. T. Schweizer has a background as a historian specializing in text-related problems (e.g. TEI), digital editions, and new forms of publication. He completed a PhD on methodological questions in digital editions[25].
A. Kilchenmann (PhD student) has a background in media science and folklore studies. He specializes in the integration of moving images and sound, and supports several research projects[26].
D. Böni has an MA in Law and advises the DaSCH on copyright laws and personal rights.

Besides leading DaSCH, the DHLab is one of the leading research labs in the field of computational photography and digitization of cultural assets (historical photography, moving image, paintings, art, sculpture, etc.) It develops new methods to capture the greatest possible amount of information using visual methods (e.g. materiality of brilliant surfaces, etc.) and to present them using state-of-the-art visualization technologies.

University of Bern
The team in Bern is not affiliated with a special institution, but works as a network of members of the Faculty of Humanities:
Professor C. Urchueguia deals with research at the intersection of musicology and digital culture[27].
Professor T. Andrews is the first person at Unibe to hold a position as Assistant Professor of Digital Humanities, and is DARIAH's representative for Bern University[28].
S. Kaufmann is a software developer who provides local support for research projects in Bern[29].

University of Lausanne
In Lausanne, the project is based in a new laboratory, the LADHUL (Laboratory of Digital Humanities and Cultures of the University of Lausanne). It covers three faculties (Arts and Humanities, Social and Political Sciences, and Theology and Religious Studies), and several institutions are involved:
Professor C. Clivaz (50%) is the first person to hold the position of Visiting Professor in Digital Humanities at UNIL, and belongs to the team of Lausanne researchers that pioneered DH in Switzerland in 2010. She is in charge of the DaSCH team's pilot project at UNIL. She is also responsible for developing research projects (such as DLCM) and for all of the DaSCH's international con-

20 French-language research uses the term SHS (Social Studies and Humanities) more frequently than the English-language research: the term "Digital Humanities" is still largely focused on Arts and Humanities, but is evolving towards sociological concerns. At UNIL, our DH laboratory concerns three faculties, which combine Sociology, Arts and Humanities.

21 Please note that only B. Geer (0.8 FTE) and D. Böni (0.2 FTE) are salaried from DaSCH funding. All other persons are funded by other resources not directly connected to the funding of the pilot.

22 CV: http://www.iml.unibas.ch/index.php/de/team/5-rosenthaler-de

23 CV: http://www.iml.unibas.ch/index.php/de/team/6-fornaro

24 CV: http://www.iml.unibas.ch/index.php/de/team/28-benjamin-geer

25 CV: http://www.iml.unibas.ch/index.php/de/team/14-schweizer-de

26 CV: http://www.iml.unibas.ch/index.php/de/team/16-kilchenmann

27 CV: http://www.musik.unibe.ch/content/team/professorinnen/prof_dr_cristina_urchuegua/index_ger.html

28 CV: http://www.dh.unibe.ch/de/tara-l-andrews/

29 CV: http://www.dh.unibe.ch/de/personen/

tacts. She directs several research DH projects, notably the ERASMUS+ DH for UNIL. She is a member of the SUC P2 pilot committee, and DARIAH's representative for UNIL [30].

Professor D. Vinck (10%) directs LADHUL [31] which is an inter-faculty platform, administratively located in Faculty of Social and Political Sciences, but co-directed by a council composed of the three deans. An expert in science and technology studies, he focusses on the ethnographic study of engineering and research labs in the field of digital culture and humanities [32].

Professor Bela Kapossy (20%) represents the Arts and Humanities Faculty in the project. He is Professor in Modern History at the University of Lausanne specializing in the history of political thought. Over the last eight years he has been active in various activities related to DH at Lausanne. He has notably developed the platform Lumières. Lausanne in collaboration with IT developers and pedagogical advisors. He regularly integrates DH components into his teaching and gives talks on topics related to Lumières.Lausanne. He participated in the setting up of an MA level specialization course in DH and is now part of its organizing committee. He is founding member of the Laboratoire de cultures et humanités digitales (LADUHL) at Lausanne and since 2012 has been member of the board [33].

S. Buerli (10%), project manager at FORS for ORD@CH, is responsible for contacts with FORS on open research data [34].

Dr M. Rivoal (80%, Humanities-IT) has a background in archaeology and is responsible for assisting and advising researchers in data analysis and modelling at UNIL. She plays a central role in the project's education and training component [35].

Dr M. Sankar (60%, IT-Humanities) is responsible for specific IT developments and research in the Lausanne team, and maintains links with VITAL-IT and the DLCM project [36].

Local resources play an important role at all three Universities, as the funding for the pilot alone would be totally insufficient.

30 CV: https://applicationspub.UNIL.ch/interpub/noauth/php/Un/UnPers.php?PerNum=891115&LanCode=37

31 Laboratoire de cultures et humanités digitales.

32 CV: https://applicationspub.unil.ch/interpub/noauth/php/Un/UnPers.php?PerNum=1045031&LanCode=37

33 CV: https://applicationspub.UNIL.ch/interpub/noauth/php/Un/UnPers.php?PerNum=1045031&LanCode=37

34 CV: http://forscenter.ch/de/about-us-3/staff/#Stefan_Buerli

35 CV: https://applicationspub.unil.ch/interpub/noauth/php/Un/UnPers.php?PerNum=1156575&LanCode=37

36 CV: https://applicationspub.unil.ch/interpub/noauth/php/Un/UnPers.php?PerNum=1104491&LanCode=37

2.2 Organizational approach during the pilot phase

The DaSCH pilot takes the form of a network that currently consists of the nodes Basel – Bern – Lausanne. This network can be expanded at any time to include new partners. The individual locations have a great deal of freedom to take local decisions (e.g. which research projects are considered important to be included in the platform), but such decisions are usually taken in consultation with the DHLab in Basel. The same holds true for software development. This local decision-making authority is crucial for the acceptance of the DaSCH, as the DaSCH would be virtually unable to function without the funding and support of the local institutions.

At each location, it is necessary to have both a broad knowledge and experience in humanities research and at least some IT and software development skills in order to provide high-quality support.

Basel is currently the main provider of technology and software development, but this is slowly changing, and we view this as a positive development.

So far, there have only been a few meetings which have been attended by all partners (e.g. at the DH2014 conference). Due to limited resources, most communication takes place online, using video conferencing (Skype) and a shared text-editing platform (Google Docs). In-person meetings of all partners are excessively time-consuming and difficult to organize. Digital communication has proven to be very efficient, targeted and adequate. However, bilateral in-person meetings are organized when necessary, and occur quite frequently.

2.3 Features and technical approach

2.3.1 Long-term preservation of digital information

Our daily experience seems to suggest that digital data is quite volatile and unstable. Everybody who works with computers on any scale has suffered the unfortunate experience of data loss. In a recent interview, Vincent Cerf, often regarded as one of the "fathers of the internet", says he is worried that all the images and documents we have been saving on computers will eventually be lost: "Our life, our memories, our most cherished family photographs increasingly exist as bits of information – on our hard drives or in 'the cloud'. But as technology moves on, they risk being lost in the wake of an accelerating digital revolution." [37]

37 http://www.bbc.com/news/science-environment-31450389 (13 Feb 2015)

Thus, it appears that "long-term archival" and "digital" are diametrically opposed concepts. However, the digital domain offers some unique characteristics that allow the long-term preservation of digital data. However, guaranteeing long-term access to digital information remains a tedious and difficult process. The characteristics that permit the long-term preservation of digital information are:

Reproducibility
Digital data can be considered to be a *text* of 0's and 1's, which means that it can be copied – *cloned* – without any loss of information (which is not possible for analogue information). However, such a copy can only be considered successful if the both versions have been proved to be identical through the additional step of comparison.

Distribution
Digital information is immaterial and can be distributed at the speed of light using digital communication channels and digital networks, without transporting any physical material from point A to point B.

Independence of physical medium
Digital data can be recorded on many media types using physically different recording mechanisms (magnetic, electrical, optical, mechanical, etc.).

The combination of these qualities can make digital data suitable for long-term archiving. Digital data can be stored with high redundancy, as the data can be copied with zero loss and distributed anywhere. Copying the data from one storage medium generation to the next *(migration)* renders the unavoidable degradation of storage media irrelevant, provided that the copying process is carried out when the inevitable deterioration still allows for a correct reading of the digital information. From a theoretical point of view, only digital data has the properties for true long-term archiving – given enough time, all analogue data will deteriorate to the extent that it will ultimately become useless.

However, the root of the problem of long-term archiving of digital information lies in the fact that digital technologies are still in their infancy, and are immature compared to established archiving methods. The rapid advances in technology pose enormous compatibility problems between different generations of hardware and software. This rapid change in technology is usually more limiting to the lifespan of digital information than the actual ageing of the physical recording medium.

As a result, the following fundamental methods for long-term securing in the digital domain are possible:

Preservation of hardware ("computer archive" or "computer museum")
Not only the storage media such as magnetic tapes, disks, etc. are archived, but also the machinery and peripherals to read and handle the data must be preserved in working condition. The "old" programs still run on "old" hardware, thus ensuring that all the digital information remains accessible. This approach is not workable due to the physical ageing of computers and storage device components beyond repair.

Emulation
The software and to some extent the hardware of obsolete computer system can be emulated ("simulated") on modern computers. In very limited cases, emulation can be a viable method of preservation. Emulating the hardware and software of redundant computer systems can preserve software such as games, etc., in which the user interaction is part of the system. However, this method still requires the migration of the data carriers as emulation is limited to software, and peripheral hardware cannot be emulated (e.g. to read old 8" floppy disks).

"Eternal" media
The "eternal" media approach requires the digital data to be recorded onto the most robust and durable media available. In addition, the storage medium should be readable using generic technologies. The DHLab has developed a method for storing digital information on microfilm that will last for at least 500 years. Decoding requires only a generic digital camera. Another advantage of this approach is that analogue and textual information can be stored together with the digital data in a human-readable format. For example, an image might be represented as an analogue thumbnail image and as digital dataset on the same piece of microfilm, together with a human-readable description of how to decode the digital data. However, this method is only suitable for selected data because of the high cost and limited storage capacity of photographic microfilm.

Migration
In the context of long-term archiving, migration is defined as the process of copying digital data onto new, up-to-date storage media and, if required, converting the file formats to new, well-documented standard formats. Migration is a *periodic* task that has to be repeated before the media and formats become obsolete, and before the media displays the effects of ageing. If a reformatting of the file formats is necessary, the new format must be chosen very carefully in order to avoid any loss of information during migration. Migration is therefore a very difficult and costly process that requires a lot of technical knowledge. However, so far it has been the only practical solution for long-term archival of digital information.

In any case, digital information has to be stored in a redundant, distributed way. As digital data can be cloned without any loss, many identical copies can be created and stored in different locations. In case of a catastrophic event (human error, hardware failure, fire, earthquake) a high level of redundancy considerably reduces the risk of loss of information.

Experience has shown that the migration model is the most promising when dealing with large amounts of data. It is the best practice currently used by almost all large digital archives. However, there are two different approaches to a migration-based archive that are described below.

metadata required to identify and find a "document", and the technical metadata required for the management of the migration processes.

Archives usually deal with data or entire dossiers (consisting of a series of thematically related documents) that are preserved as entities. These entities, enriched by standardized descriptive metadata and additional context data and documents, are submitted by the producer to the archive in the form of the Submission Information Package (SIP). There, the SIP is transformed and supplemented by technical and administrative metadata to form the Archival Information Package (AIP), which is archived.

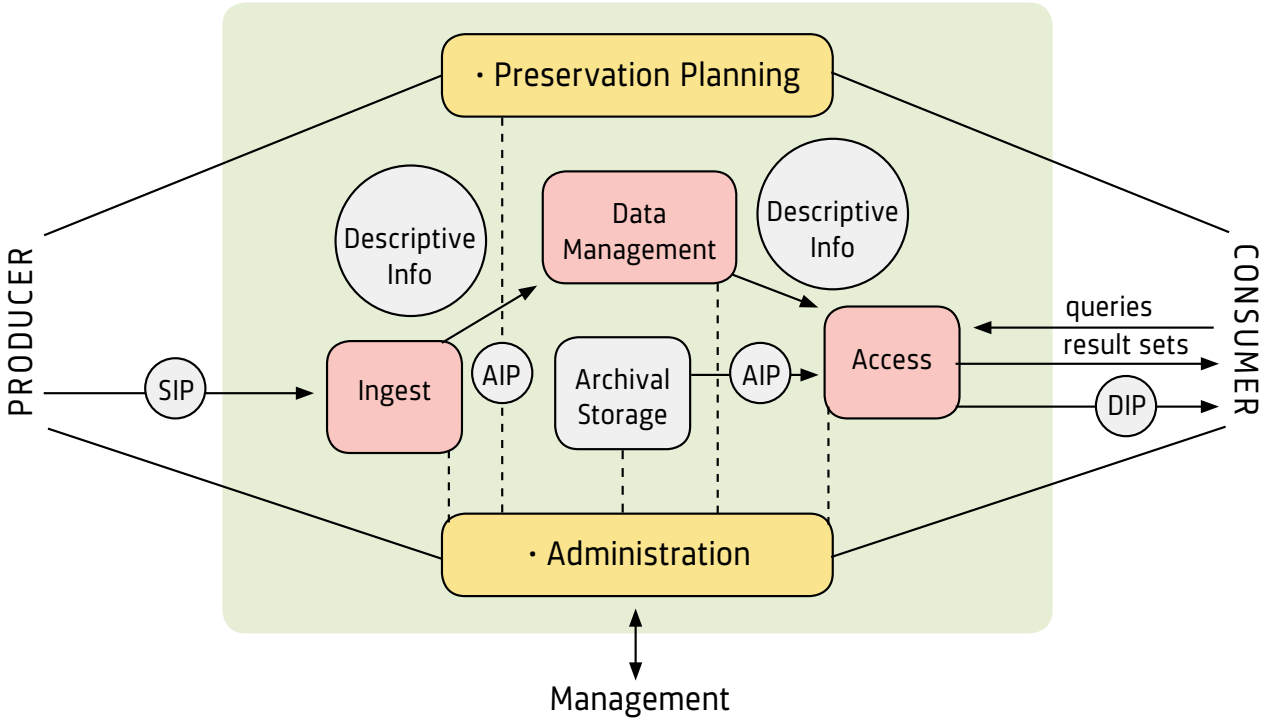


Figure 1: Scheme of the OAIS reference model

2.3.1.1 OAIS reference model

The OAIS reference model for a digital archive is based on the migration model. In addition to a formal process description, it also covers the ingest of data into the archive and the dissemination of archived data to a user. The Management Council of the Consultative Committee for Space Data Systems (CCSDS) has developed the OAIS model that is used by NASA. It is document-oriented, where a document may be a single item or a coherent (regarding the content or topic) collection of items such as text documents, datasets, images, etc. An important aspect of the OAIS reference model is the systematic approach to metadata that is distinguished between the

Finally, if a consumer requests the information, a Dissemination Information Package (DIP) is created with a copy of selected data and transferred to the consumer. Again, the original data remains unchanged.
The preserved data undergo the cycle of periodic migration necessary to overcome technological obsolescence, thus ensuring the continuous use of data and minimizing the risk of losing access.

The OAIS approach can be adapted for complex "objects" such as relational databases or NoSQL databases. The Swiss Federal Archives have developed the SIARD-Suite, a sustainable solution for the archiving of rela-

tional databases (RDBMS[38]). SIARD serializes the tables in a relational database into a standardized[39] XML-based format augmented with metadata, and stores the files in a ZIP container and thus creates a "document" suitable for an OAIS archive. In order to use the archived data, the container has to be extracted from the archive, the files representing the tables have to be retrieved, parsed and then the relational database has to be reconstructed. Currently the SIARD-Suite supports Oracle, MS SQL-Server, MySQL, DB2 and MS Access. One advantage of SIARD conversion is that, in future, preserved data can be uploaded into many different live databases, given that the desired RDBMS is supported by a future version of SIARD. In addition to the preservation of the data container, this procedure obviously requires permanent maintenance of the software components of the SIARD-Suite.

One disadvantage is that, in order to browse or use the data, the whole dataset has to be retrieved from the archive and converted back into a working RDBMS using the SIARD-Suite – a "quick overview" is not possible. Thus, an OAIS-based archive is ideal for static objects that are not accessed frequently. However, if access is required, a whole dossier or database will be retrieved and presented to the user.

2.3.1.2 Keep-alive archiving

Complementary to the OAIS archival process model, *keep-alive archiving* keeps a system of data, data management and access methods online and permanently up-to-date. This means that, whenever the technology evolves (e.g. a new stable version of the data management software or a new version of a file format is released), the whole system is migrated to conform to the new environment. The keep-alive archives are especially well suited to complex data such as databases which are accessed very frequently. Banking systems, land register databases, etc. are therefore usually regarded as keep-alive archives. However, there is one fundamental problem with keep-alive archives: if the data management system does not offer a method to record all changes, the *history* will be lost. It will not be possible to access an earlier system state. One solution is to periodically take snapshots of the system and store them in an OAIS-based archive. If a data system is no longer modified, but immediate online access is still required, a keep-alive archive combined with an OAIS-based archive offers the best solution.

38 Relational Database Management Systems such as Oracle, MySQL, PostgreSQL etc.

39 Adopted by the European PLANETS project as standard. Moreover, in early 2013 the SIARD format has been adopted as a Swiss eGovernment Standard eCH-Standard (eCH-0165).

2.3.2 Research data in the Humanities, and longevity

As our experience during the pilot phase showed[40], research data in the Humanities consists of a wide variety of digital "objects" such as digitized documents, digital texts or text corpora, digital images, films, sound, databases of any kind, etc. In addition, notes, comments, etc. about objects or relations in between those objects play an important role. Most of this data can be subsumed as *qualitative* data.

The DaSCH aims to keep all of this research data, both existing and emerging, directly accessible ("online") for researchers in order to form the base for new research or for validating/disputing previous findings. A keep-alive archive is therefore an obvious choice, independent of the nature of the data (documents, databases, other digital objects). Where suitable, keeping redundant copies of the data in an OAIS-based archive is desirable.

Research data in the Humanities often conforms little to accepted standards for long-term archiving. It comes in a lot of varieties and flavors, because the primary focus is on supporting the research, and not on the longevity of the data. For research projects that enlist DaSCH support from the beginning, we recommend formats and systems that are suitable for long-term archiving, as long as the primary goal of research is not compromised.

2.3.3 Data transfer model

In any case, there are three basic approaches to securing long-term access to digital research data using a keep-alive archive for an institution such as the DaSCH which differ in the way the data has to be delivered:

The data has to be delivered in a precisely predefined standard format and data structure. This places the burden of converting the data into the required structure and format entirely on the party providing the data. However, the advantage is that the repository has to maintain only one technical platform. Access is usually provided through a dedicated portal, and it is not possible to create project-specific applications for accessing the data in different ways. However, open access and linked open data can be easily implemented.

The data is taken as is, without modification. It is the responsibility of the institution running the repository to maintain and migrate each dataset from the corresponding technical platforms. From the point of view of the deliver-

40 See annex with detailed description of the test cases.

ing party, this is the easiest approach. However, the institution responsible for the repository has to permanently maintain and migrate a multitude of technical platforms, which will pose almost insurmountable problems in the long run. Project-specific access portals remain functional, but enabling open access and linked open data is very difficult.

The data delivered to the repository keeps basically its structure and format, but is transformed/adapted so that it can be represented within the technical platform used by the repository. This conversion is carried out through a cooperative effort involving both the repository institution and the delivering party. The disadvantage of this approach is that the necessary transformation of the information into a suitable format can be quite challenging in order to keep the basic data structures and formats intact. The advantages are that the basic structure and formats of the original data can usually be kept, and that there is still only one platform to be maintained. It is possible to create project-specific access portals, but this may require a considerable amount of work. Open access and linked open data can easily be provided for all datasets.

We have decided to use the third method – a middle ground between the first and second approaches – for the following reasons:

The first solution is much too restrictive. The delivering party would be required to transform the data into a predefined structure and format (similar to an ingest process in an archive), which in many cases would not be possible. Either the knowledge and resources are not available, or – since research cannot be standardized – transformation to the standard would cause the loss of significant information.

The second solution is simply not manageable and cannot be realized in a realistic setting. There are so many different platforms used that it would be impossible to maintain them all.

The structure of the data itself is important information and thus has to be preserved. The third solution is able to preserve this structure while still requiring only maintenance of a limited number of technical platforms (ideally only one). Thus this solution offers the best service to researchers while still being efficient for long-term preservation of access. However, this solution requires a highly flexible platform that is able to represent many data models at the same time. Fortunately, the Semantic Web technologies provide a technological framework that allows the "simulation" of many data models (relational databases, XML hierarchies, graph networks, etc.) within one framework. The modelling of the data structures is implemented using open standards such as RDFS and OWL.

The pilot phase has made it clear that project-specific access applications (such as online graphical user interfaces) have to be preserved. While this approach does not make it possible to reuse the original applications, it is relatively easy to re-implement their basic functionality as well as their look and feel.

Using the common platform, it is straightforward to create new tools and applications that reuse existing data by combining information from different datasets. Thus, new research methods can be implemented, e.g. using methods of "big data" analysis.

2.3.4 Data services

Simply presenting the archived research data online – eventually with a possibility to download it – by no means exploits the full potential of the archived data. There is a need to reference and integrate the data into a new research platform and to enrich the data with small pieces of knowledge such as a comment or annotation. It must be possible to integrate such tiny pieces of information with project-specific research tools (e.g., using *Linked Open Data* standards, LOD).
For each of these pieces of information, as well as the primary research data in the archive, it must be possible to create a permanent reference (permalink), which will always represent the data as it was at the moment when the link was created – even if it has been modified or augmented later. The data must be navigable both by humans, using simple interfaces, and by machines/algorithms for data mining and data aggregation.

In addition, appropriate tools and methods that find, access and use the data form a crucial part of a data curation system in the humanities. Often, methods of searching for and presenting the data to the user (i.e. the researcher) are as important as the data itself. A working information system is much more than hardware, software and data; it is also the implementation of a concept that contains structure, methodology and access. These concepts also need to be preserved with a long-term perspective in mind.

2.3.5 System design and technology

Long-term preservation is achieved by actively maintaining the system *and* the data for an indefinite time, while constantly adapting both to technological change[41]. It should be noted that only the *digital representation* but not the semantics (content, meaning, etc.) of the data may

41 Similar to SIARD-Suite which either also has to be maintained for an indefinite time or re-created using the available documentation at some point in the future in order to recreate a working system to access the data.

evolve with the constant evolution of technology. The DaSCH has chosen to use the concept of the *Resource Description Framework (RDF)* as a base for representing the data. The World Wide Web Consortium (W3C) standardizes RDF in a vendor-independent way. In addition, it is a very simple but highly flexible representation of digital information. Thus it is optimal for the long-term preservation of complex, highly networked information.

As RDF can be serialized in a standardized way, RDF-based digital information can be easily converted into a form compatible with the OAIS model. Thus, research data that no longer has to be kept immediately accessible can be transferred to an OAIS archive. In addition, data of high importance should be archived using different archiving strategies. In these cases, using redundant archiving strategies is highly recommended.

In order to offer both of these options, a close collaboration with the National Archives has been established. We are currently developing an automated gateway which allows the bidirectional transfer of digital information between the DaSCH system and the digital archive of the SFA following the OAIS model (DaSCH system → SFA digital archive and SFA → DaSCH) in order to offer the optimum and most cost-effective solution for all types of research data in the humanities.

2.3.5.1 The Knora Platform – Knowledge Organization, Representation and Annotation

Knora is an open, modular, extensible and flexible platform based on industry standards (RDF as data representation, SPARQL 1.1 and a RESTful web service API for access). Salsah is now a generic front end to Knora, but Knora's back-end components make it straightforward to create project-specific front ends with specific application logic and user interfaces.
Knora uses industry-standard tools (Scala/Java for the JVM[42]-based back end, and JQuery plugins for the Salsah front end). Thus, individuals with moderate software development skills can be trained to create simple extensions to the framework. The Knora framework implements the full communication stack required by the "linked open data" (LOD) standard, and thus can communicate with other applications and repositories. Interoperability with special-purpose repositories such as e-codices and e-manuscripta has already been demonstrated. A generic, individually configurable interface to Fedora Commons 4.x will be developed (spring 2015). The modelling of project-specific data structures is implemented by creating standard RDFS/OWL ontologies, which are derived from a specific Knora ontology. This Knora-based ontology guarantees timestamp-based versioning and permalinks that retrieve any digital information object in the state it was in when the link was created. Knora will shortly be available as an open-source project on GitHub (expected Feb/Mar 2015), enabling any interested party to contribute to its development, and to customize and reuse its components in new ways.

2.3.5.2 Long-term aspects of the Knora platform

In relation to long-term archiving, there are three levels of backups that have been or will be implemented in the near future:

A redundant backup of the data has to be made at each location. The backup must be stored at different locations from the location of the live servers, and should involve both binary backup, as well as RDF/XML or Turtle-dumps of the triple store.
A distributed, redundant archiving system based on DISTARNET[43] that cooperates with the Knora platform will be implemented.
An interface to the digital archiving system of the National Archives has been implemented. SIPs[44] can export directly to their archiving system (spring 2015).

However, these archiving strategies are not sufficient to guarantee permanent, long-term access to the digital information on the DaSCH platform. It is even more important for the technical base of the platform to be permanently maintained and updated in accordance with current industry standards, and for satellite installations to be upgraded as the platform software is improved. These are the most important tasks facing the DaSCH's technical group.

42 Java Virtual Machine.

43 DISTARNET stands for DISTributed ARchival NETwork and is a peer-to-peer, self-organizing network for long-term preservation of digital data along the lines of LOCKSS. (See http://www.lockss.org)

44 SIP: Submission Information Package (SIP): In OAIS-speak, this is the information container that is sent from the producer (repository) to the archive (SFA digital archive).

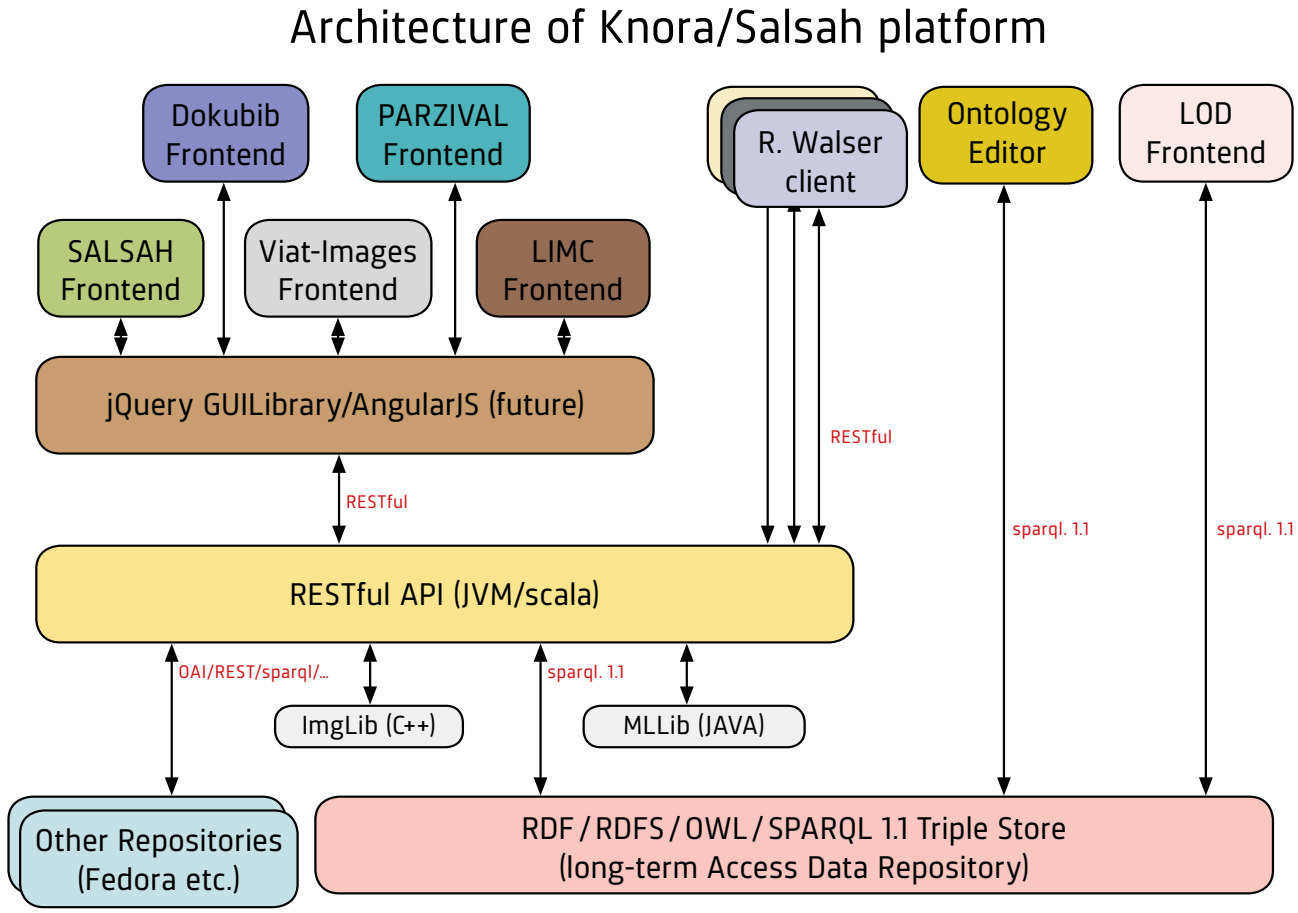


Figure 2: Architecture of the Knora/Salsah platform

As the structured data is stored as RDF, a serialization in a standardized format (e.g. RDF/XML) can be created at any time. Using simple tools, this serialization can then be imported into other repositories, including databases based on other technologies (Big Data stores, relational databases, etc.). Thus, users are not locked into the technologies chosen by the DaSCH. Digital objects such as text, images, sound, video and movies, musical notation, mathematical formulas, etc., are stored in such a way that they can be converted into other standard formats at any time (e.g. JPEG2000 to TIFF, or Standoff markup[45] text to TEI). The necessary tools are either already available or are being developed before an internal representation is used. Thus, if a repository has to be migrated to a different technology in the future, it can be guaranteed that there will be absolutely no loss of information.

45 Standoff markup totally separates the structure information from the actual text. The structure attributes the text by position. This is in contrast to the embedded markup languages such as all XML-based variants, where the structural information is intermingled with the actual text.

3. Experiences from the pilot phase

3.1 Organization

During the pilot phase, the Universities of Basel, Bern, and Lausanne created local points of contact ("satellites") for the DaSCH and provided test cases. In future, we expect that more Swiss universities and universities of applied sciences will participate. Ideally, all existing universities will participate to form a true national network.

The pilot phase has shown that the structure of a national coordination unit with local satellites can be very successful, provided that the following prerequisites are met:

The national coordination unit coordinates and drives the development of the basic software platform following good open-source practices, and guarantees its interoperability with national and international standards. It is responsible for keeping the technology up-to-date and provides migration paths in case of major changes in technology.

The satellites have to be strongly rooted in their home institutions. It is crucial that the DaSCH satellites are well integrated into the structures of each university. They collaborate closely with local structures (such as local research IT and IT services). Along with the local research IT, they support local research projects, and organize and carry out the importation of data into the platform.

At the satellites, there must be local IT knowledge available and the capability of carrying out local software development. At the very least, the development of import scripts for low- to medium-complexity projects should be carried out locally. It is important for the satellites to be able to respond to the immediate needs of researchers. The staff should have good IT knowledge and a humanities background.

Good communication between the satellites has to be developed, enabling them to share experiences, solutions, etc. This communication is encouraged and supported by the national coordination unit.

The national coordination unit provides high-quality second level support.

Procedures for technical coordination, integrating local solutions and extensions, have to be established.

The national coordination unit also provides first level support for researchers and projects that are not located at an institution with a local satellite (smaller universities, research institutions such as the VitroCentre, Historical Lexicon, etc.).

The University of Lausanne is planning quite an elaborate model in which it proposes a flexible, non-hierarchical structure with close collaboration between UNIL technical units and the DaSCH satellite. A more detailed description can be found in the annex.

3.1.1 Bern

The University of Bern has installed an Assistant Professorship in Digital Humanities. The DaSCH will be organized around this professorship and embedded with in faculty. However, no more concrete plans have been developed as yet[46].

3.1.2 Basel

Basel takes on a special role because, on one hand, it currently hosts the national coordination unit of the DaSCH and, on the other hand, the DHLab is performing local research IT support and is therefore also acting as a satellite. For a permanent instalment of the DaSCH, it is likely that the DHLab will be responsible for hosting the DaSCH satellite.

3.2 Technical platform

The test cases have shown that the original Salsah implementation, which was designed as a vertically integrated "Virtual Research Environment", has not been flexible enough to meet all the requirements of the test cases. In particular, the generic web-based user interface has been shown to be inadequate for some tasks, such as simple browsing of a given project's data, as well as for project-specific actions. However, in other cases it is a powerful working environment, enabling users to relate and annotate disparate information. It has become clear that, in some cases, the project-specific user interface must be preserved, in the sense that their basic functionality, along with their look and feel, have to be re-implemented. The new Knora architecture has been devised to simplify this task by modularizing and simplifying the previous

46 Which is also due to the fact that, financially, Bern has the smallest share in the pilot project. Bern will be equivalent to Lausanne and Basel in a permanent installment of the DaSCH.

architecture. It now offers an open, flexible, and extensible platform that follows industry standards. This development is an ongoing effort based on the requirements of the incoming projects. Using the tools that Knora and its libraries provide, it is possible to re-implement project-specific user interfaces without a great deal of effort.

Often, a project-specific user interface is used to consult the repository or perform quick searches, while the Salsah interface is used to add, annotate, and link information, or to visualize the knowledge network using the integrated graphical visualization tools.

Thanks to its modular and extensible architecture, the Knora platform is ideally suited to the integration of existing digital research tools. This enables these tools to become interoperable with other tools and with many different datasets. Moreover, new tools can be developed directly within the Knora platform. For example, a new tool for non-linear publication, as well as tools for critical editions, will be developed in separate research projects that build on the Knora platform. The DaSCH's (limited) support for these activities helps to add value to research data in the DaSCH platform and facilitate reuse of the data in new research. The DaSCH actively encourages the release of such tools as open-source software and their integration into the Knora codebase, in order to make them available to all researchers.

3.3 Importing data

The importation of data into Knora usually follows a pattern that can be separated into a preliminary clarification step and five operative steps:

Preliminary clarifications
A few preliminary clarifications have to take place before the data import can start:

a) Is there a general interest in making the given data permanently available to the research community[47]?
b) Clarification of legal issues, in particular copyright issues and personal rights.
c) Clarifications of access levels (blocking period, restricted access, creative commons, etc.).
d) Is there enough metadata available to describe the objectives and scope of the research project and its digital data?
e) Establishing a legally binding contract between the person(s) responsible for the research data (data owner) and the DaSCH defining the details of the data transfer (see model contract in the annex).

Analysis of the data model, data structure and data formats
The first step is an in-depth analysis of the existing data. For the later steps, the structure and goals of the existing data model have to be well understood. This is also the phase in which we must gain some domain-specific knowledge about the subject of the research data. It is important to understand the concepts that the researchers wanted the data to represent. Experience with many projects has shown that this step often reveals problems or deficiencies in the existing data model. Even simple databases made with software such as FileMaker may have structures that are not completely adequate for the information they are intended to represent.
Who: DaSCH staff + researchers
Results: a complete understanding of the structure of the existing data model and a complete understanding of the domain-specific objects and the relations between them.

Modelling of an ontology to represent the data
Using the results of step 1, an RDF-based model expressed in RDF, RDFS and OWL is created. As the concepts of RDFS and OWL are rather complex, a GUI is provided to simplify this process[48]. The complexity of this step depends on the complexity of the domain-specific objects and object relations.
Who: DaSCH staff
Result: an RDF/RDFS/OWL data model (ontology) implemented in the platform.

Developing transfer programs
In the next step, short programs or scripts must be developed to import the data into the platform. Since the platform software provides a RESTful API for this task, any programming language can be used that supports the industry-standard HTTP protocol. Such programs have been written in Python, PHP, JavaScript, and C#. The import script will have to deal with all inconsistencies in the original data and correct them automatically as far as possible. For testing purposes, the DaSCH is able to provide several (virtual) test servers that are exact copies of the live server.
Who: DaSCH staff and/or IT staff of the research project
Result: all information is imported into the platform and can be viewed in the generic Salsah user interface.

Developing a project-specific access application (GUI)
Salsah, the generic research environment, is often not optimal for a quick assessment of the content in a specific project. It is designed as a generic research tool for working with all the data from all projects in the repository.

47 In cases where it is not obvious, this decision may be presented to a sub-committee of the executive board.

48 As this GUI is primarily meant to be used by the staff of the DaSCH, it has still numerous limitations and lacks user-friendliness. We are aware that this process has to be reviewed and improved. However, despite these drawbacks, an RDF-based data model can be constructed with relative ease.

However, using the widgets provided, the development of a simple web application is straightforward and efficient. This user interface can be given a project-specific look and feel using standard technologies such as CSS.
Who: DaSCH staff and/or IT staff of research project
Result: a project-specific web application for presenting the data (and, optionally, a specialized work environment).

Conformance tests, quality assessment
The last step is to test the completeness and conformance of the imported data.
Who: project staff
Result: verification of the correctness and consistency of the data in the platform.

During the second half of the pilot phase, DaSCH has been over overwhelmed by the demand for advice and support from ongoing and new projects. There seems to be not only a lack of knowledge about digital research methods in the humanities, but there are obviously almost no institutions or facilities that support the researchers not only with advice but with the concrete implementation of new digital research methods. Cutting-edge research in the humanities very often requires existing programs to be modified or new programs to be developed. It has been very exciting to observe that the DaSCH satellites became local catalysts to build such facilities at the participating universities. As an example, the rectorate of the University of Basel approved a new position of "IT Navigator" which will be attached to the DHLab.

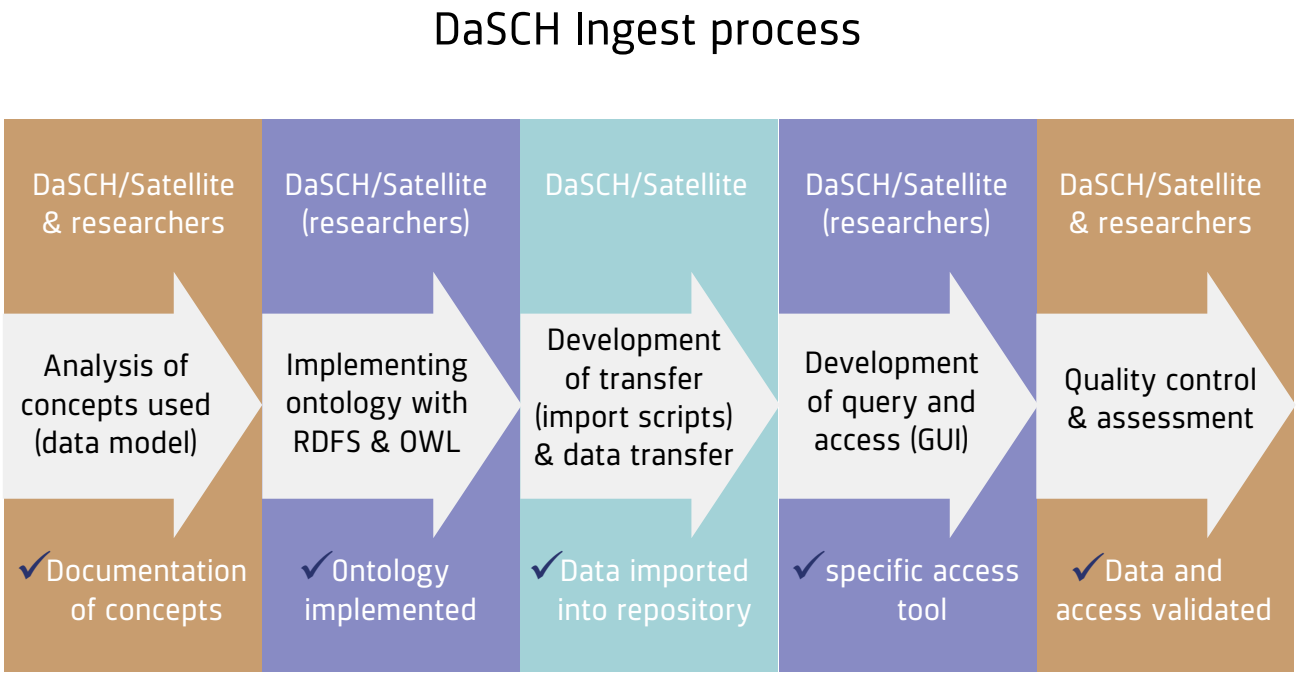


Figure 3: Steps of a typical ingest process into the Knora platform

3.4 Services and support

The purpose of importing data into the DaSCH repository is to make it available for future research, which may extend it, enrich it, and combine it with other digital information sources. Thus an important characteristic of the DaSCH is that it must offer advice and assistance to researchers interested in using the data in the repository. These services have to be coordinated with the researchers' local institutions (e.g. with their IT services departments.).

3.4.1 List of services provided

During the pilot phase, the DaSCH provided the following support and services:

Access to research data
Providing access to research data that the DaSCH curates is the most important service. Access to the data is provided on different levels:
a) using a web based generic virtual research interface (Salsah)
b) through a RESTful interface that allows research tools to remotely access the data in a machine-readable format for analysis and presentation
c) specialized, project-specific user interfaces
d) a full SPARQL-endpoint will be established in the near future.

Long-term preservation of research data
The long-term preservation is provided through two different paths:
a) Keep-alive archiving of the data using the Knora platform and highly redundant backup procedures. In the near future, these methods will be supplemented by the DISTARNET[49] system.
b) Interface to the long-term archive of the Swiss Federal Archives for selected datasets.

Advice & consulting on data analysis and modelling
This is the most important – and often the most difficult and time-consuming – service that DaSCH has had to provide. It is essential for research data at the end of its active life ("post-mortem" acquisition) and for projects that are still active or in their initial planning stages. It is crucial for an efficient and productive use of the research data during and after the lifetime of a research project.

Training and technical support for Knora
There is a constant need for training and technical support for the Knora platform. In the long run, the satellites should provide most of this support. However, during the pilot phase – especially in the beginning – the coordination unit has provided most of this support.

Advice & consulting on digital research methods
Many researchers are unaware of up-to-date digital research methods. They sometimes have some vague ideas but need help in putting them into practice. Existing tools and methods can often be adapted. Sometimes new tools (e.g. a specialized visualization) have to be developed.

3.4.2 Digital editions

The SNSF's call for applications in 2014 for long-term critical edition projects was like a small tsunami in the community. As a result, it became clear that critical editions would have to use digital tools and methods. While Knora in its current form is not a tool for the publication of critical editions, it is a very useful tool for collecting and annotating text, supplementary information, etc. Test cases such as the Anton Webern Edition demonstrate its usefulness as a research environment. In the course of the call for applications, many edition projects contacted the DHLab for support, and some promising collaborations have begun in recent weeks (e.g. Humboldt Edition [Bern], Bernoulli-Euler [Basel], and Sentences of Petrus Lombardus [Basel]).

3.5 Experiences with the test cases

The test cases processed so far are described in more detail in the annex. Here we present a summary of the results of the test cases. To date, we have worked on 13 test cases in different disciplines, with different degrees of complexity. These test cases cover all levels of project status: finished projects ("post mortem"), active projects ("in vivo"), and projects in the planning or starting phases ("ab ovo"). There are some common findings for all three types.

3.5.1 Data modelling

It is important to understand, at least to some degree, the subject or research of each project. This is necessary in order to translate (in the case of post mortem or in vivo) or create (ab ovo) an adequate data model to represent the data in the platform. The task of creating a data model requires considerable direct interaction with the researchers.

In the case of planned or starting projects ("ab ovo"), one difficulty is that researchers are not always familiar with the important concepts for their data models. Several examples (e.g. the Schweizerische Gesellschaft für Volkskunde, and the Anton Webern Edition) have shown that creating an adequate and efficient data model is essential for the success of a research project.

Already active projects ("in vivo") often use tools and data models that are not optimal for the given task. Migration into the platform is often an opportunity to clean up the project's data models.

Post-mortem integration poses the biggest challenge. In one of the major test cases, the Lexicon Iconographicum Mythologiæ Classicæ, there is no documentation available, and the people that created the data models and software are no longer available. However, there are still many active users who can help us to understand the concepts. Still, such projects require a great deal of time-consuming reverse engineering.

For different reasons, several of the test cases provide a drastic illustration of the challenge of understanding the existing concepts and transforming them (please refer to the annex to read about these test cases in more detail):
A) Lumières.Lausanne
B) DYLAN (Language dynamics and management of diversity) datasets
D) "Dessins de dieux" database
G) Anton Webern Gesamtausgabe
J) Lexicon Iconographicum Mythologiae Classicae (LIMC)

49 Distributed Archival Network.

3.5.2 Access

Knora's standard tools for accessing the repository (e.g. Salsah, the RESTful API, or direct SPARQL queries) are important for using the data within other research projects, as several test cases have successfully proven. However, in some cases, the application that was used to access the data within the original research project is important for understanding the project's concepts and for browsing and evaluating the data. In these cases, these access methods (including the fundamental characteristics of the original user interface) must be preserved. We are therefore developing a software library to facilitate this process. It has already been used successfully in a few test cases (e.g. St. Moritz, and to some extent Parzival, where more work is required).

The following test cases are good examples in which preserving the access methods is important:
E) Artists and Books (1880–2015): Switzerland as a cultural platform
F) Dokumentationsbibliothek St. Moritz
J) Lexicon Iconographicum Mythologiae Classicae (LIMC)
L) Parzival
O) HyperHamlet

3.5.3 Reusing and adding value to existing data

It is interesting and promising that, even though these test cases have only been available in the repository for a short time, there has already been substantial concrete interest in reusing the data from these projects within other research projects (e.g. by Harvard University for the LIMC and ETHZ for the Documentation Library of St. Moritz):
F) Dokumentationsbibliothek St. Moritz
J) Lexicon Iconographicum Mythologiae Classicae (LIMC)
K) VitroCentre Romont
O) HyperHamlet

3.6 Financial aspects

The pilot phase showed that experienced staff are needed to handle these demanding tasks. Skilled software developers are essential for developing the main infrastructure and implementing customized project-specific solutions, and qualified staff are needed to facilitate communication between the many project teams. We can draw the following conclusions on the basis of the budget for the pilot phase:

The configuration of resources as it currently stands (around CHF 300,000 p.a.) was to some extent sufficient for the pilot project (besides the direct funding of the pilot, the matching funds of about CHF 450,000 p.a. provided by the universities must also be taken into account). They are not sufficient for building up and running a national data center. Cutting back on staff would mean major drawbacks for the integration of existing projects and for the development and integration of new features. The integration of existing databases can be scaled in proportion to the size of the staff. Software development is similar regarding scalability, with one major difference: it is essential to have at least small working groups of three people, to keep efficiency at a reasonable level.

3.7 Further activities during the pilot phase

3.7.1 The DaSCH as a national point of contact for full Swiss membership of DARIAH

On the initiative of the DaSCH pilot (Professor Claire Clivaz), the DaSCH consortium members University Basel, Bern, Lausanne, and outside the consortium the University Geneva (with which the DaSCH have a close collaboration) has joined DARIAH have as cooperating partners for two years (2015–2016). The DaSCH is thus well prepared to become the national point of contact, which would have positive effects for the DH community in Switzerland and for the DaSCH in opening full access to DARIAH's activities and resources. As shown above, the DaSCH team is actively and successfully pursuing the acquisition of additional external funding in order to expand its knowledge base and to remain an institution with cutting-edge technological and methodological expertise. Moreover, the Academy requests a full membership of Switzerland of DARIAH-EU in the multi-year planning for 2017–2020 for the attention of the State Secretariat for Education, Research and Innovation (chapter 3.4 "Vollmitgliedschaft der Schweiz bei DARIAH-EU", multi-year planning 2017–2020 SASSH).

3.7.2 Swiss National Research Infrastructures

On 23 October 2013 the State Secretariat for Education, Research and Innovation (SERI) and the Swiss National Science Foundation (SNSF) launched a call for *"Update and renewal of the Swiss roadmap for research infrastructures of national relevance in view of the Federal Council Dispatch on the promotion of Education, Research and Innovation (ERI) for 2017–2020."* The DaSCH consortium, led by the DHLab and in close consultation with the SAHSS, submitted an application. The aim of this application was to ensure that the DaSCH would be included in the roadmap of the Swiss National Research Infrastructures. The consortium and the SAHSS decided that, as head office, the DHLab Basel would submit the application, as the SAHSS was legally not entitled to do

so. We are grateful to the rectorates of the universities of Basel, Bern and Lausanne who supported this application. The most important parts of the application are as follows:

- The center will offer a sustainable, reliable and trustworthy data platform for digital research data in the humanities
- Long-term accessibility and long-term archiving
- State-of-the-art data and access management providing a high degree of control over "who can do what"
- An extensible, open, and flexible toolbox of methods for data management, analysis, and visualization
- A toolbox/library of basic functions for the creation of new tools and research methods
- Connectivity to external data sources and repositories in accordance with the "linked data" standard
- Information and training about digital research methods and data management, focusing on qualitative data and digital sources.

For more details, see the full application text in the annex. The proposal gained the support of all three rectorates (see letters of support in the annex). The evaluation by the SNSF gave the proposal the highest rating of "A", with some minor points for improvement (also in the annex). One of the reviewers stated:

"The basic logic of this application of high scientific potential and of outstanding importance for Switzerland's research landscape is irrefutable: the humanities need infrastructures that make data available, connect them to appropriate tools and make data safe for a long period."

Another reviewer states:

"The application frequently mentions the intention to develop tools. It goes beyond the annoyingly monolithic approaches of infrastructure applications in other countries and acknowledges explicitly that it also has to and will provide a framework into which more specialized tools, to be developed later by the supported projects, can in turn be integrated."

On 12 November, the Rectors' Conference of the Swiss Universities wrote a letter stating that it will possibly support the infrastructure projects that were evaluated with an "A" with a maximum of the total budgeted amount with project-bound funding. The Universities had 2 weeks to confirm that they will apply for this project-bound funding, which the rectorate of the University of Basel (representing the national coordination unit of the DaSCH) did on 25 November 2014. Both documents can be found in the annex. Currently, it is not totally clear what the consequences of this statement are.

3.7.3 Collaborations beyond the humanities

The recent past has shown that the infrastructure, tools, and methods of the DaSCH are of interest well beyond the humanities. The advanced methods and tools developed for the humanities may also find interesting applications in the natural sciences, medicine, law, economics, etc. On the other hand, the DaSCH may also profit from developments in these areas. As described above, the collaboration with the DLCM-project, VITAL-IT and sciCore is a very promising start in this direction.

4. International comparison

In this chapter, we try to position the DaSCH in relation to other international institutions that provide a repository service to provide long-term access to research data. The comparison is based on the following topics:
– Mission
– Resources and organization (staff, finances, and institutional embedding)
– Services
– Technical base
– Embedding into national funding strategies

4.1 Data Center for the Humanities at the University of Cologne (DCH)[50] and Cologne Center for eHumanities (CCeH)

The DCH is a central facility of the Faculty of Humanities at the University of Cologne. Its mission is the permanent safeguarding, availability and presentation of digital research data. It is operated by the Cologne Center for eHumanities (CCeH)[51] using the IT infrastructure of the "Regionales Rechenzentrum der Universität". Thus, the missions of the DCH and CCeH have to be considered together because, in effect, they form a combined entity (even if formally they are separate entities). From the mission statement, this combination can well be compared to the DaSCH in terms of functions and priorities, even if the scope of its services and duties are local to the university while the DaSCH provides similar services on a national level.

Mission statement
Permanent safeguarding, availability and presentation of digital research data.
Improving the visibility of active research projects. Improving communication and coordination between similar research projects. Strengthening the existing interdisciplinary structures of the Faculty of Humanities. Clarifying and strengthening the special profile of the Faculty of Humanities inside and outside the University of Cologne.
Increasing digital expertise for postgraduates and staff of the university (training, workshops, symposia, etc.).
Support of ongoing research projects in relation to methodology and technology.
Support of the institutes and chairs in project development, acquiring external funds, and project implementation.
Networking on a national and international level with institutions in the field of eHumanities.
Coordination of activities from the different disciplines, institutes and chairs with other infrastructures of the university (university library, computer center, IT support for the faculty).
Support in teaching in Digital Humanities (new modules, courses of study, etc.).

These duties are very similar to the goals of the DaSCH. However, the DaSCH tries to reach this goal on a national level with a networked organization.

Resources and organization (staff, finances, and institutional embedding)
The DCH is under the auspices of the "Cologne Center for eHumanities", which has 22 staff. Most of these are dedicated specialists for different tasks such as data modelling, imaging, digital editions, access, digitizing, and publication systems. Some of the staff are directly assigned to the support of a few dedicated research projects.

Services
The DCH/CCeH provides a wide portfolio of services such as data modelling, analysis and digitization. It also explicitly includes the adaptation of existing digital research tools and the development of new ones.

Technical base
The DCH/CCeH seems to use a wide variety of 3rd party and custom software systems. A strong focus is on x-technologies (XML, XSLT, XQuery, XML-DB).

Embedding into the national funding strategies
The DCH/CCeH are funded primarily through the University of Cologne and the state.

The DCH/CCeH has an impressive portfolio of services and supports a large number of research projects with advice, consulting, and active participation (software development). However, to our knowledge, only a few projects have reached a stage where long-term preservation of access is an issue. Long-term preservation is still in the pilot phase.

50 http://dch.phil-fak.uni-koeln.de/startseite.html?&L=1
51 http://www.cceh.uni-koeln.de

Hence, the aims of the DCH/CCeH for the University of Cologne are very similar to those of the DaSCH on a national level, but are even more focused on support. However, it is evident that the DCH/CCeH has access to much more substantial resources than the DaSCH.

4.2 Data Archiving and Networked Services (DANS)

DANS is an institution financed by the Koninklijke Nederlandse Akademie Van Wetenschappen and the Netherlands Organisation for Scientific Research, which is the national research council in the Netherlands and has a budget of €650 million per year. NWO promotes quality and innovation in science. That is, like the DaSCH, DANS is a *national* service and repository for the Netherlands.

Mission statement
"DANS promotes sustained access to digital research data. For this, DANS encourages scientific researchers in archiving and reusing data in a sustainable form, for instance via the EASY online archiving system. With NARCIS, DANS also provides access to thousands of scientific datasets, e-publications and other research information in the Netherlands. With the Dutch Dataverse Network, DANS also supports researchers in data management during their research. The institute furthermore provides training and consultancy and carries out research on sustainable access to digital information.[52]"

Resources and organization (staff, finances, and institutional embedding)
DANS is a national institution that provides its services for all researchers in the arts and humanities in the Netherlands. Its finances are secured through the budget of the national research council. It has 47 staff consisting of a director, two deputy directors, and 44 specialists for different tasks (software development, data modelling, imaging, digital editions, access, digitizing, publication systems, communications, etc.). There is a steering committee, which is responsible for daily business and management, and consists of one representative each from the academy and from research council. The advisory board offers solicited and unsolicited advice to the steering committee and management of DANS. The board also responds to the DANS evaluation commission report concerning the six-yearly external evaluation. It consists of six national experts from the field.

Services
DANS's activities are centered on three core services: data archiving, data reuse, and training and consultancy.
Data archiving and data reuse
DANS offers 3 different systems for data archiving and reuse:
EASY: Easy is a file-based archive following the "traditional" OAIS reference model approach. A limited but quite wide palette of file formats is accepted. During data import, some important meta data about content, topic, etc. has to be entered into the system for later retrieval and reuse.
Dutch Dataverse Network (DDN): the DDN relies on the Dataverse Software of the Institute of Quantitative Social Science of Harvard University. It is a Java Netbeans application based on PostgreSQL and has R and Zelig data analysis components (both statistical packages). The primary target of DDN is quantitative social science data.
National Academic Research and Collaborations Information System (NARCIS) is an inventory of scholarly and scientific information about institutions, persons, publications, and datasets. NARCIS uses persistent identifiers and is based on an XML framework.
Training and consultancy
DANS offers a wide variety of training and consultancy. However, it does not offer direct research IT support to researchers and project teams.

Technical base
DANS relies on different systems and architectures for its services. With respect to research data, the Dataverse Software is an open-source application developed by Harvard University (using Java's NetBeans technology and a relational database).

Embedding into national funding strategies
DANS is a national institution funded by the research council of the Netherlands. It has very stable funding, which is crucial for the development of long-term strategies.

DANS is a very successful organization that offers a wide range of services. However, the model for long-term access is basically document-based. DANS is conducting its own research in the field of preservation and access to digital data. Interestingly, in a manifesto called "Exploring the long-term Availability of Research Data – DANS eResearch Programme 2012–2015"[53], some interesting statements/questions can be found:

How can data be made interoperable across infrastructural organizations, communities, and over time?

52 http://www.dans.knaw.nl/en/about/organisation-and-policy

53 http://dans.knaw.nl/en/about/research-and-innovation

How can principles of Linked Open Data be implemented in living archives?
How can a digital archive develop a standards-based, rich interface for fine-grained, interrelated data?
Can a seamless user interface for data archiving, data analytics, and data annotation be developed that is suitable for different disciplines?
What role can visualization of "raw" data and metadata play in the reuse of data?
How can we create and support new ways of navigating the growing landscape of data?

These are exactly the questions that were posed at the beginning of the development of the Knora/Salsah platform at the DHLab some six years ago. As the pilot of the DaSCH successfully demonstrated, the Knora/Salsah platform is able to carry out exactly the tasks raised by these questions. Compared to the DaSCH, DANS can be considered a combination of FORS and DaSCH with some focus on social science data and/or file-based information. DANS does not yet offer a solution for complex networked data such as databases, etc. In terms of resources (staff, finances), DANS is in a good position to respond to the needs of researchers nationwide.

4.3 UK Data Archive

The UK Data Archive is the curator of the largest collection of digital data in the social sciences and humanities in the United Kingdom. It hosts several thousand datasets relating to society, both historical and contemporary.

Mission statement
The UK Data Archive exists to support high-quality research, teaching and learning in the social sciences and humanities by acquiring, developing and managing data and related digital resources, and by promoting and providing access to these resources as widely and effectively as possible.

Resources and organization (staff, finances, and institutional embedding)
The UK Data Archive was founded in 1967 by the Social Science Research Council, which has committed to its long-term funding. This has been vital to the success of the archive[54]. It has around 70 employees. Most of its funding comes from the Economic and Social Research Council, the Joint Information Systems Committee, and the University of Essex, with some other sponsors.

Services
The UK Data Archive is definitively intended for the social sciences. In its statement about its services, it writes: "We provide a range of services supporting creators and users of social and economic data for research and teaching." From a preservation point of view, the UK Data Archive is generally conformant to the OAIS Reference Model, with additions and alterations that are specific to the materials held within the archive.

Technical base
The technical base is not known, but it seems that the UK Data Archive follows a strict OAIS model. Therefore it recommends that *qualitative data* should use XML-based formats, and that relational databases should be flattened to XML by the user before ingest.

Embedding into national funding strategies
The UK Data archive is backed by the Economic and Social Research Council, and thus has had stable funding for 40 years, which has allowed for long-term planning. Its longevity has given it a secure place in the UK research landscape.

While the UK Data Archive is – in the context of digital data – a very old institution, it is primarily aimed at quantitative social science data. Its approach to qualitative data, such as relational data bases, seems limited.

4.4 TextGrid (D)

TextGrid is a joint project between ten partners, funded by the German Federal Ministry of Education and Research (BMBF) for the period from June 2012 to May 2015. TextGrid consists of a data repository (TextGrid Repository) and tools (TextGrid Laboratory) based on the extensible editor Eclipse (which has to be installed locally on each user's PC, along with a series of plugins). TextGrid is explicitly intended for working with texts (TEI) and digital editions. Data is processed locally using the Eclipse plugin, then stored in the common repository. Collaborative work is then possible.

Mission statement
TextGrid lives in and through its community of users and supporters. It describes its mission as follows: "TextGrid is a Virtual Research Environment for scholars in the text-based humanities and cultural studies.

54 http://www.data-archive.ac.uk/about/archive

Among other features, it supports the creation of digital editions using free and project-specific expandable tools and services.[55]

Resources and organization (staff, finances, and institutional embedding)
TextGrid is an open-source, community-driven project, and the extent of its resources is not clear. Ten institutions are currently involved, with funding from the German Federal Ministry of Education and Research until 2015. It is not certain whether the current level of operation can be sustained with an open-source model, or whether a different revenue model will have to be implemented. In 2012, an association "TextGrid – Verein zum nachhaltigen Betrieb einer Virtuellen Forschungsumgebung in den Geisteswissenschaften e.V." was formed, with the task of finding a long-term institutional base and stable funding.

Services
In addition to providing plugins for the open-source Eclipse editor, the basic service provided is the central data repository, which is maintained by the Göttingen State and University Library (SUB) along with the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG). The community provides the support.

Technical base
TextGrid is based on a central repository (back end) built around three Java-based middleware applications: TG-auth for authorization of users, TG-search for queries, and TG-crud for the RESTful API. For TextGrid, "The world is flat. Ever was. The same goes for TextGrid. That is, TextGrid flattens all existing hierarchies. Hierarchies are only built by the application logic of the upper levels of the TextGrid infrastructure, primarily by the TextGrid Lab."[56] Metadata is stored in XML format in the form of tuples (metadata, data). An XML-based database (eXist) is used.
The front end is implemented using Eclipse plugins. Eclipse is a Java-based Integrated Development Environment (IDE) that is used mainly for software development, and can be extended with plugins. The user has to install Eclipse and the plugins on his or her local computer.

Embedding into national funding strategies
TextGrid is based on the principles of an open-source ecosystem, where many partners contribute to the development. The German Federal Ministry of Education and Research (BMBF) has provided the seed money, but this funding will stop in 2015, and the project's long-term future is therefore uncertain.

TextGrid provides an interesting technical solution that is clearly aimed at text-based research. Other data, such as images and video, can be handled as files with associated metadata. However, there are again no tools for handling complex qualitative data stored in databases of any kind (relational databases, NoQL databases, FileMaker, etc.) in an adequate manner. Judging by its architecture, it is also suboptimal for handling large numbers of media files.

4.5 Geisteswissenschaftliches Asset Management System GAMS (Graz)

Since 2003, the Centre for Information Modeling – Austrian Centre for Digital Humanities has provided an infrastructure for a variety of Digital Humanities projects.

Mission statement
In addition to applied research in the field of information processing in the humanities, the Centre for Information Modeling in the Humanities at the University of Graz offers support and cooperation for research projects in the humanities. It also operates GAMS, a repository for long-term access.

Resources and organization (staff, finances, and institutional embedding)
The center has 12 employees and is part of the Karl Franzens University of Graz (Austria). The University of Graz basically funds it.

Services
The center supports and advises research projects in all aspects of research IT. In addition, GAMS provides a central repository for long-term access to research.

Technical base
GAMS is based on the FEDORA Commons architecture (Flexible Extensible Digital Object Repository Architecture), in addition to some custom applications for content preservation and data preservation (Cirilo). As it is based on FEDORA Commons, it closely follows the OAIS model.

Embedding into national funding strategies
The center is locally funded by the University of Graz. It is well established as an important part of the Faculty of Humanities in Graz.

In addition to a whole set of services (support for research projects, specialized software development, etc.) for the humanities, the Center for Information Modeling provides, along with GANS, an excellent, document-based repository service for the University of Graz.

55 https://www.textgrid.de/fileadmin/materialien/TextGrid__Digital_editing_-_research_-_archiving.pdf

56 https://dev2.dariah.eu/wiki/display/TextGrid/The+Physical+View

4.6 TGIR Huma-Num (France)

TGIR Huma-Num is, in France, "the large-scale infrastructure (TGIR) that aims to facilitate the digital turn in research in the humanities and social sciences. To carry out this mission, TGIR Huma-Num is built on an original organization relying on a social structure (collective cooperation) and a technological structure (digital long-term services) at the European and national levels, based on a wide network of partners and operators" [57].

Mission Statement
TGIR Huma-Num offers a service matrix focused on
– long-term preservation
– tools and methods
– presentation and visualization
It offers these services on a national level.

Resources and organization (staff, finances, and institutional embedding)
TGIR Huma-Num is funded by the CNRS (Centre national de la recherche scientifique), l'Université d'Aix-Marseille and the Campus Condorcet. It has stable funding, with 15 permanent employees at three locations in France. These employees operate its central infrastructure (Isidore, Nakala, Services Grid, long-term archival, etc.).

Services
There are 4 basic services:
Intégration de services, interconnexion de données de la recherche et de l'enseignement (Isidore) [58]: Isidore is the central repository and working environment for Huma-Num. It collects, enriches, and offers unified access to digitized documents and data in the SHS, and promotes interoperability between databases.
Nakala: Nakala offers permanent links to, and citability of, research data; it is enriched by the individual researchers, in a bottom-up approach.
Long-term archiving [59]: long-term archiving works together with the Archives Nationales and offers a service for importing data into the AN.
Services Grid [60]: the project's concept of supporting researchers is particularly interesting: this support is organized through consortia that are arranged around disciplines. The following consortia have been established:
Consortium Musica
Consortium 3D
Consortium Mémoires des archéologues et des sites archéologiques
Consortium Sources Médiévales, Consortium Cartes et photographies pour les géographes
Consortium Archives des sciences sociales du politique
Consortium Archives des mondes contemporains
Consortium CAHIER – Corpus d'Auteurs pour les Humanités: Informatisation, Édition, Recherche
Consortium Corpus Oraux et Multimodaux
Consortium Corpus Écrits
Consortium Archives des ethnologues.
Each of these consortia is organized as a network within the discipline and offers support, special tools, etc. The consortia are also financed by the TGIR's funds.

Technical base
The whole infrastructure of TGIR Huma-Num is based on Semantic Web technology using triple stores [61] and RESTful APIs. It holds around 3 million resources.

Embedding into national funding strategies
TGIR Huma-Num is part of France's national strategy for advancing research in the humanities and social sciences. The CNRS is investing large sums into this crucial infrastructure.
The technical concept of TGIR Huma-Num can be well compared with the DaSCH. It uses the same technologies and also unifies the service infrastructure around Semantic Web technologies. Long-term archiving is – as planned for the DaSCH – based on cooperation with the Archives Nationales through the provision of an easy gateway. Thus, TGIR Huma-Num has a very similar approach as the Knora/Salsah platform. However, the Knora/Salsah platform is much simpler and therefore easier to manage. It allows for a decentralization of services and storage and thus is more suited to a federalist approach. In addition, it is much easier to adapt and extend Knora/Salsah to include special requirements.

Where TGIR Huma-Num differs significantly from the DaSCH is in the organization of support, advice and cooperation: it is based on a number of disciplinary networks. In our view, the approach of embedded satellites is more promising in Switzerland, a small country with only a handful of universities. However, our contacts with Huma-Num have led to a stimulating reflection on how including a bottom-up approach could help the DaSCH to reach the largest possible number of humanities researchers in the medium term.

It should be underlined that, as DANS does it, the humanities and social sciences are both covered by Huma-Num. This is not currently the case in DARIAH, which remains focused on the humanities, but this point is evolving at a general level. On 16 January 2015, a European Association

57 http://www.huma-num.fr/la-tgir-en-bref
58 http://www.huma-num.fr/service/isidore
59 http://www.huma-num.fr/service/archivage-a-long-terme
60 http://www.huma-num.fr/service/grille-de-services
61 It uses the commercial triple store from Antidot (see http://www.antidot.net).

of Social Sciences and Humanities (EASSH) was founded in Paris to enable the humanities and social sciences, particularly in the European Horizon 20/20 projects, to speak with one voice when budgets for SSH have to be strongly defended.

4.7 OpenEdition

OpenEdition is a digital open publication platform that also offers many related tools. It collaborates with TGIR Huma-Num and with humanities researchers in several countries.

Mission statement
Its mission is to promote the development of electronic publishing in the humanities and social sciences, and to participate in the dissemination of skills related to electronic publishing.

Resources and organization (staff, finances, and institutional embedding)
OpenEdition has around 30 employees and is located at the Université d'Aix-Marseille, l'EHESS and l'Université d'Avignon et des Pays de Vaucluse. It is funded by the CNRS and also receives funds through its Freemium model [62]: two thirds of these funds are going to the journals and to the partners, one third to OpenEdition.

Services
OpenEdition offers four complementary platforms for open publications in the humanities and social sciences: Revues.org, OpenEdition Books, Calenda and Hypothèses. It includes some high-quality automated bibliographical research tools such as Bilbo [63]. These services are well-known among researchers. They support strongly open-access publication, with interesting economic models such as Freemium [64].

Technical base
OpenEdition uses "Logiciél d'édition électronique" (Lodel), which is based on a classical LAMP (Linux, Apache, MySQL and PHP) architecture.

Embedding into national funding strategies
OpenEdition is developed and operated by the Centre for Open Electronic Publishing, with stable funding from the CNRS.
OpenEdition is an interesting platform that is more targeted at publications and digital editions. Again its success is due to stable funding with a long-term perspective in mind, combined with some regular external funding from the Freemium model.

4.8 Results of the international comparison

The international comparison shows that no single repository can meet everyone's needs. There are various approaches to ensuring long-term access to research data. Only DANS represents a truly national repository; most of the others are specific to one particular institution, a group of collaborating institutions, or a local state. Different courses of action, organizational structures and funding strategies have been adopted in accordance with different scopes, aims, and missions. Furthermore, many of these projects are still in a development phase or even a conceptual phase.

However, there is considerable common ground for data centers in the humanities. This is also stated in the position paper that is in preparation by the AG Datenzentren der Digital Humanities im deutschsprachigen Raum (working group of data centers for the Digital Humanities in the German-speaking area). There is a general agreement on the following issues:

Consulting and participation in research projects as a partner, able to adapt or create new digital tools, cooperation, training, etc., must be an integral part of a data center.
There are four areas where a data center should offer solutions and support:
Long-term preservation of the digital data
Securing direct and easy long-term access to the data
The possibility to link the data to other digital sources or repositories, e.g. via Linked Open Data (LOD)
Presentation and research environments.
The staff at the data centres must have broad knowledge and experience of information technologies, software development, digital long-term preservation and – very importantly – research in the humanities. This type of data center therefore requires staff with a highly interdisciplinary orientation. In addition to having very good IT skills, the staff must understand to a certain extent the content of the data they have to deal with.
Stable, long-term funding is absolutely crucial.
There seems to be a consensus that purely centralized solutions are no longer favored, particularly because support, advice, consulting, and even cooperation are very important services. Networked and cooperative solutions that guarantee the close proximity of support to the researcher who is using the facilities seem to be optimal.
Participation in international communities of data centers, but also more broadly in the Digital Humanities community, is very important. Technology and digital meth-

62 See http://cleo.openedition.org/openedition/freemium; http://cleo.openedition.org/pilotage/abonnes

63 http://www.revues.org/; http://lab.hypotheses.org/; http://books.openedition.org/; http://calenda.org/

64 http://www.openedition.org/8873

ods for the humanities are evolving at an extremely rapid pace. The data center has to keep up with these developments in order to avoid becoming obsolete in a short time.

One important aspect is the organization of support and service. While most of the institutions we examined for comparison take a more centralist approach, where the researchers have to contact the center for advice and help, Huma-Num again provides a very interesting model of organizing support and service along the lines of interest groups in the different disciplines. We consider this approach to be a very promising one. It corresponds with our experience that the support and service must be very close to the researcher. However, this approach would probably not work well in Switzerland, as a small country with only few universities, because the communities would be too small. In addition, we consider cross-disciplinary synergies to be very important. Therefore, our approach with local satellites is optimal in relation to service level, efficiency and cost.

This brief comparison of data centers for the humanities also shows that there is no single solution to the problem of long-term access to research data. It is interesting to note that DANS has a research program that addresses exactly the same issues (fine-grained connect data, open linked data, etc.) for which the DaSCH has a working solution that has been tested on reasonably large real-world cases.

The solution proposed by the DaSCH is an advanced approach that is very promising for the future, and has shown great potential. During the pilot phase, it has been extensively tested with success on many different real-world cases.

The choice of using Semantic Web technologies is not an exotic one; the TGIR Huma-Num initiative is using this technology on a large scale at national level with great success.
The international comparison has also shown that the available resources are definitively not adequate for the task. The current level of funding would be at the lower limit of what is required even for a local repository for a single university. All comparable initiatives have 5–10 times more resources available.

We conclude that the proposed DaSCH is very well positioned in an international comparison. It uses an adequate, very advanced technology that is very promising for the future. All signs from the Digital Humanities community (such as presentations at academic conferences) show that Linked Open Data[65] will be the technology of choice for sharing research data in the future. It follows best practices and is interoperable with other repositories.

65 Which, incidentally, is a mature technology that has been used for more than a decade in many fields and branches of science, industry and business.

5. Implications for the implementation of the DaSCH

The international comparison clearly shows that with a comparatively small budget the DaSCH has achieved the goal of establishing a working data and service platform for long-term access to research data that is in high demand. During the pilot phase, the general approach of migrating data from the different research projects onto a common technical platform without imposing strict rules on data formats etc. has proven to be practicable, efficient and attractive. It is interesting that one of the larger, successful national initiatives, Huma-Num in France, adopts a similar approach and uses similar technologies.

5.1 Organization and governance

5.1.1 Organization

The concept of a national coordination unit which:
- Coordinates and drives the technological development
- Provides second level support
- Provides documentation and good practice
- Develops demanding tools and methods

and locally anchored satellites which are close to the researchers using the services of the DaSCH and offer:
- First level support
- Data analysis and import of existing data collections
- Project collaboration for new research projects
- Small development

seems to be optimal for Switzerland as a federalist, multilingual country. To bring this organization into use, a close interaction between the national coordination unit and the satellites is very important. It has been an advantage that the head office of the pilot is located within an institution (the DHLab) that also provides first level support for the researchers in the faculty of Humanities at the University of Basel.
The diagram below (Fig. 4) illustrates the proposed organization for a permanent installation of the DaSCH.
Such a distributed organization clearly requires a good balance between local and global decision-making. According to the lessons learned during the pilot phase, we propose to leave as much authority as possible within the local satellites. However, some national policy and decision-making has to be implemented.

5.1.2 Governance

The governance structure depends to some degree on the legal structure of the future DaSCH, which is unknown at present. However, independent of the legal form, we propose to have two bodies of governance: the Executive Board is the governing body of the DaSCH. The second, the Scientific Board, guides the DaSCH with regards to content.

5.1.2.1 Executive Board

The executive board is the primary governing body that is responsible for strategic decisions and overseeing the operations of the DaSCH. It can adopt whatever measures are necessary to ensure that the DaSCH works well, and it ensures that the funds received by the DaSCH are spent for the intended purposes. It also supports the DaSCH in achieving its goals on a political and conceptual level. The board is responsible for optimal embedding of the DaSCH within the scientific landscape of Switzerland. The members of the executive board shall represent the major stakeholders of the DaSCH:

One member appointed by the SAHSS
The SAHSS represents the different disciplines of the humanities and is therefore an important stakeholder.

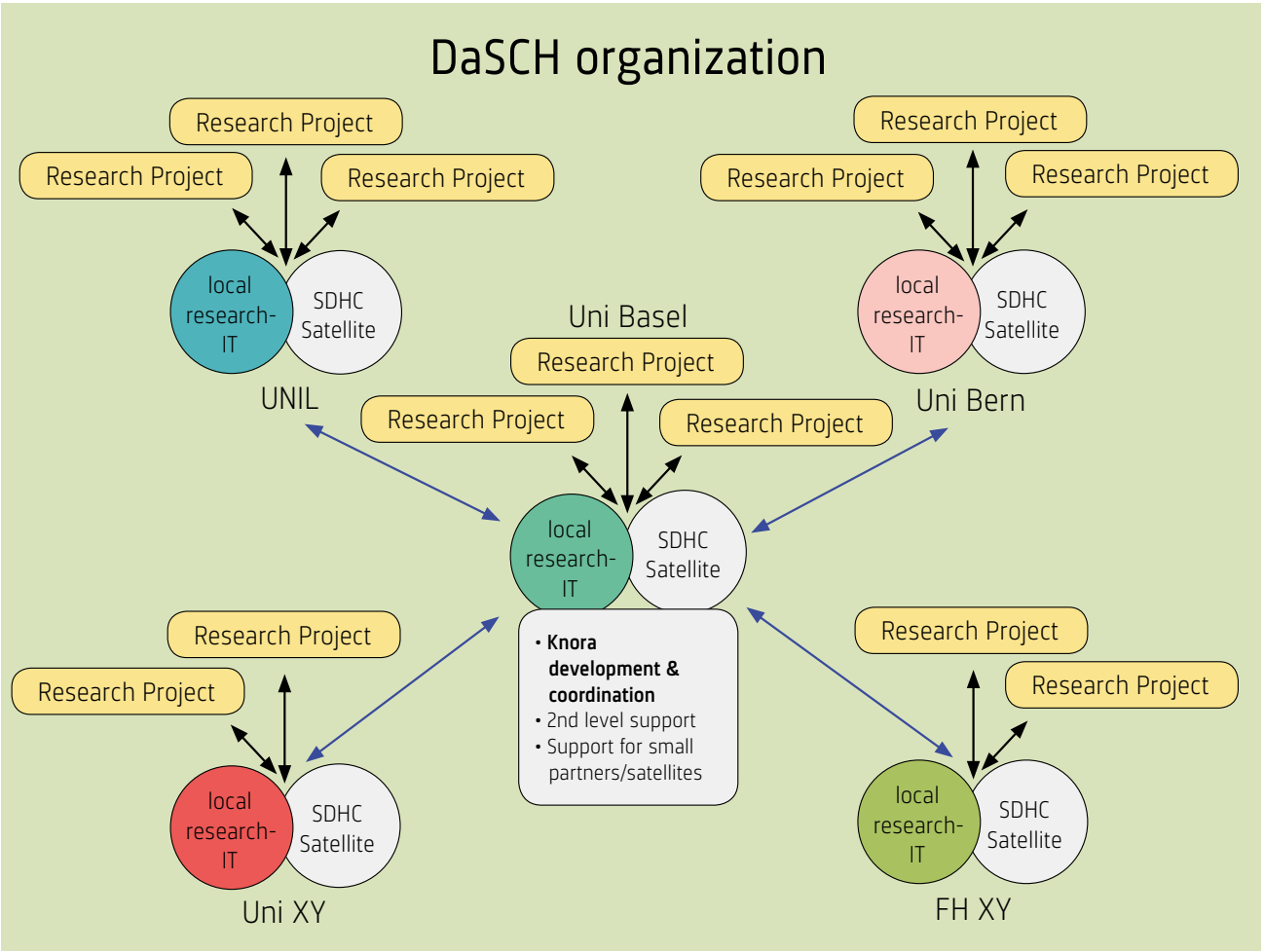


Figure 4: Proposed organization of the DaSCH as a network of satellites and a central development unit also providing second level support

One member appointed by the SERI
The state secretariat represents a major funder of the DaSCH. If so desired, it may nominate a delegate to the executive board.

One member appointed by the SNSF
The SNSF represents one of the most important funding agencies for humanities research that has a direct interest in research data created with its funds remaining accessible in order that it may be valorized by future research. The SNSF may impose rules regarding long-term access when granting funds.

One member appointed by the SwissUniversities
SwissUniversities represents the universities and universities of applied sciences.

One member appointed by the Swiss Federal Archives
The SFA represents an important partner of the DaSCH.

One delegate from each university which operates a satellite and provides some matching funds
The universities which host satellites and provide matching funds for the operation of the satellites are also important stakeholders. They should delegate a person who is able to represent the strategic interests of the university.

The executive director of the DaSCH reports to the executive board twice a year in relation to activities and finances. The executive board, in collaboration with the scientific board, conducts an evaluation of the DaSCH every four years.

The executive board should convene once or twice a year.

5.1.2.2 Scientific Board

The scientific board supports both the executive board and the executive director of the DaSCH with scientific and operational matters that are at the core of the activities of the DaSCH. The scientific board consists of seven to ten leaders of national or international standing in the field of Digital Humanities and long-term access to research data. The members of the scientific board should cover a wide range of disciplines in the humanities.

The scientific board drafts and revises the rules for the selection of research data that will be curated by the DaSCH. These rules have to be approved by the executive board.

The scientific bard should convene at least once a year.

5.1.2.3 Ad hoc technical committees

The executive director may convene ad hoc technical committees for advice and consultation.

A technical committee should support the management of the DaSCH in making strategic decisions about the development of the technical infrastructure and provide assistance with the implementation of these strategies.

5.1.2.4 Organization of the national coordination unit

An executive director who is responsible for the realization of the strategic goals and operative work leads the national coordination unit. He reports to the executive board. The national coordination unit consists of four departments that work very closely together.

Software development
This department should be responsible for the continuous, long-term maintenance of the platform software in order to add features, respond to user feedback, and adapt the software to changes in technology and industry standards. Moreover, new tools, and interfaces for those tools, have to be developed. As the platform software will be an open-source project, this department should also coordinate contributions from external developers, enforce quality standards, ensure that contributions are documented, and integrate them into the release management process, etc.

Infrastructure and deployment
This department should manage the servers and storage, third-party software (such as the RDF triple store), and the network. Although we should aim to share as much infrastructure as possible with the host university's scientific computing department, for example, a great deal of special-purpose installation and maintenance will have to be carried out.

Archiving
Long-term archiving is a very demanding task. This department implements the center's long-term archiving strategies, and maintains the gateway to the Swiss Federal Archives, in cooperation with the software development department.

Support and documentation
This department, together with the software developers, is responsible for providing second level support to satellite institutions, and provides some first level support to smaller institutions or projects that have no access to a satellite. In addition, it is responsible for producing high-quality documentation of the platform's processes, components, tools, and interfaces.

Service and consulting
While service and consulting is primarily carried out at the satellites, the national coordination unit has to maintain an excellent and responsive second level support. In case of very demanding projects, the staff at the national coordination unit may take a leading role if necessary.

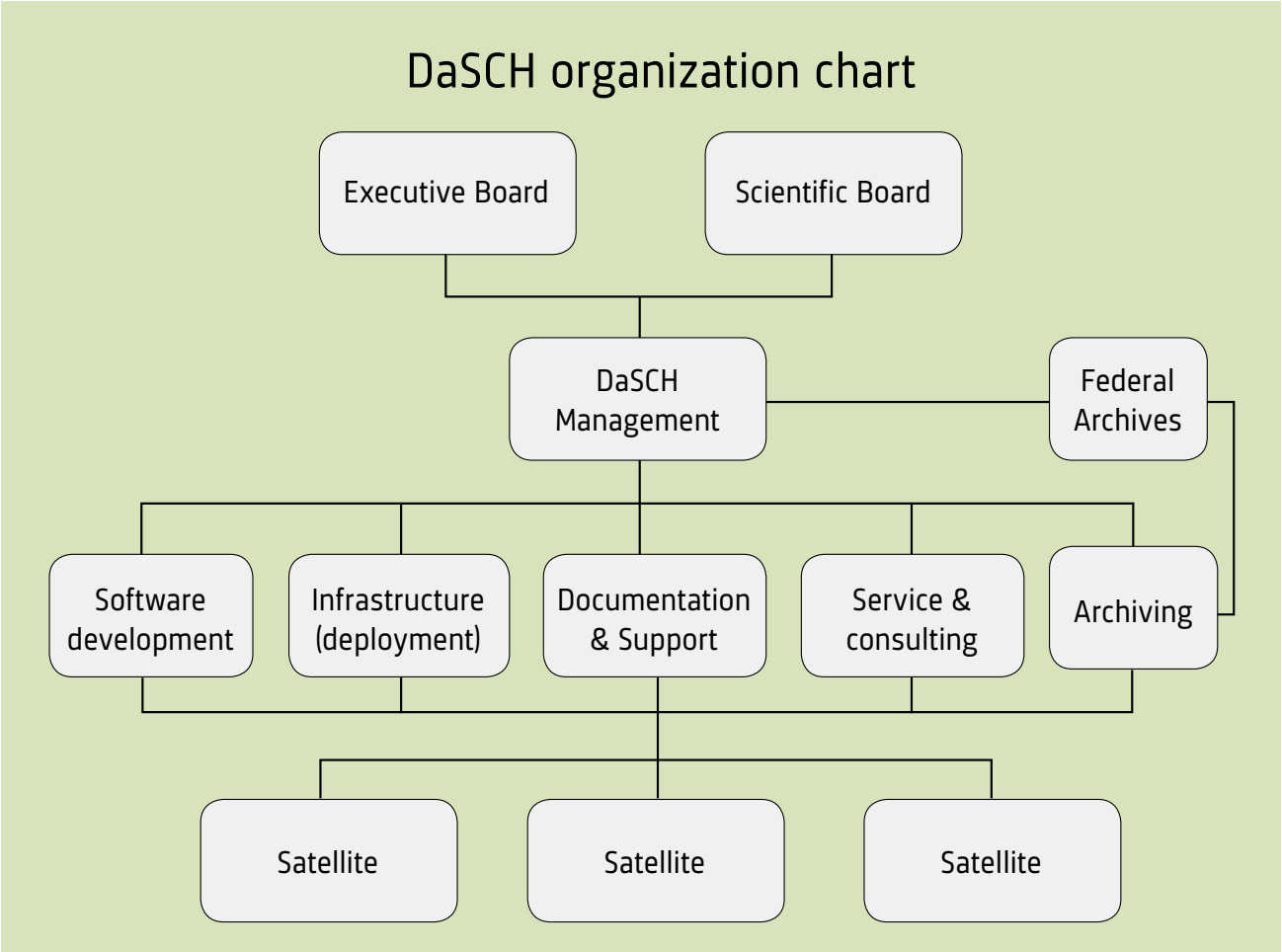


Figure 5: Governance of the DaSCH

5.1.2.5 Organization of the satellites

Proper embedding of the satellites into the host institution is crucial for the success of this model. During the course of the pilot project, none of the institutions created permanent structures, but there were some efforts made to create models showing how the embedding might work.

The internal organization of the satellites in terms of embedding it into the local structures has to be decided by each participating university.

5.2 Financial aspects

The pilot phase was funded with approx. CHF 300,000 p.a. It is obvious that a sustainable operation of the DaSCH to meet the (increasing) demand is not possible with this amount of money. Comparable institutions, as described above, have considerably more funding available. However, we believe that, with the proposed funding below, a highly efficient and cost effective institution can be established and maintained.

The estimate below is based on full cost accounting. It distinguishes between the national coordination unit responsible for driving and coordinating the software development, providing second level support and support for running the machinery. The full cost of the national coordination unit is estimated for about three of four satellites. If there are many more satellites, the budget of the national coordination unit will have to be adjusted.

5.2.1 Estimate of full cost of national coordination unit including long-term archiving

A reasonable allocation of resources would be as follows, assuming 5 local satellites (costs including social security, etc.):

1 FTE executive director & chief software architect (CHF 190,000 p.a.)
3 FTE software/IT specialist with DH background, adaption of the platform to technological change, tools and methods (CHF 130,000 p.a. per person x 3 = CHF 390,000 p.a.)

1 FTE infrastructure specialist, server and network (CHF 110,000 p.a.)
1 FTE secretary & communication, administration, finances, communication (CHF 80,000 p.a.)
Total personnel: CHF 770,000 p.a.

Computer infrastructure (server, PC, etc.), network, etc. (CHF 100,000 p.a.)
Office rent, etc. (CHF 80,000 p.a.)
Travel, meeting and support of small projects (CHF 50,000 p.a.)
Total other cost: CHF 230,000 p.a.

Total cost of the national coordination unit: CHF 1,000,000 p.a.

For the support of long-term archiving by the SFA we estimate costs of CHF 100,000 p.a.

Thus, the *full cost o*f operating the national coordination unit, including the computer infrastructure, offices, etc. and long-term archiving is approx. CHF 1,100,000 p.a.

5.2.2 Estimate of full cost of a satellite

The full cost of the satellites of course depends on several factors: the size of the university, own resources available, service level required, etc. However, we believe that our estimate represents a good average.
A satellite should consist of 1–3 FTE supporters with a strong background in the humanities in addition to good IT knowledge.
Including infrastructure cost, we estimate an average total cost of approx. CHF 300,000 p.a. and per satellite as appropriate. Based on this sum, funds totaling CHF 1,500,000 p.a. would be required for the satellites. In case it is not possible to obtain the indicated sum, we recommend cutting the average total cost of a satellite to CHF 200,000 p.a. and starting with smaller satellites instead of creating a lower number of satellites. We consider nationwide anchorage and coverage of the satellites to be more important than starting with larger satellites. Five smaller satellites would require funds of CHF 1,000,000 p.a.

5.2.3 Further costs

5.2.3.1 Swiss Federal Archives

For the support of long-term archiving, data ingest into the SFA (adaption of SIP, support and service, etc.), we estimate around CHF 100,000 p.a.

5.2.3.2 DARIAH national membership

DARIAH membership will cost about CHF 45,000 p.a. This amount (€35,000) is only necessary if Switzerland chooses to become a full member of DARIAH, and if the DaSCH is selected to be the national point of contact.

5.2.4 Summary of effective costs

The table below summarizes the total cost per annum, on the basis of six active satellites:

National coordination unit (staff, infrastructure, long-term archiving)	CHF 1,100,000
5 satellites (CHF 200,000 per satellite)	CHF 1,000,000
DARIAH	CHF 45,000
Total effective cost p.a. (on the basis of six sat.)	CHF 2,145,000

On an international scale, the required resources are very modest. Nonetheless, we believe that an excellent service for the research community can be maintained with this funding. However, the funding must be allocated on a stable basis, as preserving long-term access to digital data makes long-term funding absolutely indispensable.

5.3 Technical infrastructure

5.3.1 Knora/Salsah platform

The Knora/Salsah platform proved to be a flexible and stable environment for the given task of providing a keep-alive archive and, at the same time, a research platform. The modularization that has replaced the rather monolithic architecture of the early Salsah implementation has been essential for success.

The decision to use the JPEG2000 format for storing images and facsimile has been forward-looking. The DaSCH has been approached by the sciCore-facility[66] of the University of Basel to use the JPEG2000 image server developed by the DaSCH.
In the near future, some important developments will be completed, which are crucial for providing a stable and future-oriented platform for keep-alive archiving.

5.3.1.1 Going open source

As soon as the new scala-based platform is fully tested and documented, its source code will be given into open source. We plan to establish a GitHub repository for the source code. The DaSCH will be committed to further coordinate the development and contribute substantially to add new features as required. The goal is to build an international community of contributors

66 Scientific Computing Core Facility, a group dedicated to providing high-performance computing and large-scale storage for research in the faculties of Science and Medicine at the University of Basel.

which will keep the code base up to date and add interesting features. Going open source also reduces the dependency on one developer team and thus contributes substantially to the longevity of the software base.

5.3.1.2 DISTARNET

The Distributed Archival Network is a long-term archiving method that extends the ideas of LOCKSS[67] to arbitrary digital objects. The base is a peer-to-peer network which automatically creates redundant copies of digital information and distributes them optimally on the participating nodes. Through elaborated algorithms, the system recognizes the failure of nodes and automatically maintains the required redundancy. If a node fails completely, all data on this node can be reconstructed using the distributed, redundant copies. The system also allows for format migration. Hardware and storage migration is achieved by simple exchange of the hardware/storage. The system automatically recognizes that a node is missing and a new empty node has been attached to the network and thus starts to increase the redundancy and fill the new node with data. DISTARNET offers a state-of-the art archiving service and is ideally suited to keep-alive archives. The development of DISTARNET has been funded by the SNSF.

5.3.1.3 Knora-Cloud

We plan to extend the Knora platform in order to become a true distributed system. Thus the Knora platform will have some characteristics of a cloud system. The advantage is that each participating institution may have its own server and store its data locally. However, for a user with the proper permission, the local servers will act like a single large system.

5.3.1.4 User interface

As one of the primary tools for accessing the Knora platform, the user interface of Salsah has to be redesigned. We started using AngularJS as the primary tool to create a modern, extendable web-based front end that will be easy to use. AngularJS implements a MVC model on the browser side.

5.3.1.5 2D/3D and CAD using WebGL

Knora and Salsah will be extended in order to natively support also 2D and 3D objects such as CAD[68] drawings and GIS[69] information. There is a twofold motivation behind developing these new tools: on one hand, our own research at the DHLab (e.g. digital materiality) requires such tools. On the other hand, disciplines including archaeology would profit considerably from the inclusion of new digital objects such as 2D, 3D drawings and GIS information. The tools will support industry standards such as the DXF[70] file format. The research team at the DHLab worked closely with the Swiss industry leader in this field[71]. The tools will be web-based using the new WebGL standard that has been adopted by all major browser developers.

5.3.2 Computer infrastructure

The hardware infrastructure has been extended in order to provide a high level of reliability and security against data loss. The virtualization using vmware has added a great deal of flexibility. Storage is provided by a NAS system with 75TB using RAID6 with hot spare. The NAS is connected to the server via a dedicated 10GB Ethernet.

In the near future we plan to collaborate with the sciCORE facility at the University of Basel, which offers high performance computing and large-scale storage services.

67 The LOCKSS Program, based at Stanford University Libraries, provides libraries and publishers with award-winning, low-cost, open-source digital preservation tools to preserve and provide access to persistent and authoritative digital content. See http://www.lockss.org

68 Computer Aided Design.

69 Geographical Information Systems.

70 Drawing Interchange Format, an industry standard for 2D and 3D CAD drawings.

71 CADWORK AG, see http://www.cadwork.ch/indexL1.jsp?wsid=2

6. Overall conclusions

Urgent need
All clarifications and investigations carried out since 2008 clearly indicate an urgent need for a data and service center for research data in the humanities. In recent times, the demand for IT support has again been accentuated with about 20 requests issued to the management of the pilot project. The digital transformation of research methods associated with demands for Open Access, Open Data and Linked Open Data that are also IT related resulted in a massive increase in complexity on the technical side. As a result, "normal" researchers face very difficult challenges, as the impact of researchers trained in Digital Humanities methods only begins to show up gradually. The services offered by DaSCH allow an efficient distribution of a collaborative research process, which is advantageous for all parties. Moreover, the research funding organizations themselves also have a growing need for expertise that the DaSCH can provide.

Technical feasibility
The tests carried out using real data in different formats and different contents (editions, databases, images, texts, etc.) have shown that a platform built on the principles of the Semantic Web using the RDF data model meets the requirements for long-term access and long-term preservation of research data. The data transfer, the most complex and labor-intensive part of the ingest, can be carried out on the basis of standardized processes and tools with economically acceptable costs. The long-term preservation of data is guaranteed in cooperation with the Federal Archives. Due to its modular design, the Knora platform is suitable to respond to future needs that are as yet unknown.

Organization as a national service with localsatellites at the universities
The pilot has already clearly shown that the organizational form featuring a national coordination office and "satellites" as local contact points is best at meeting the needs of researchers. The coordination office has to remain in close contact with research, which requires a local connection to a university. At the same time, it must be able to remain independent of the universities in order to be nationally recognized and to be able to develop a coordinating effect. The division of labor between the national coordination office and the local satellites mainly provides different types of support levels. Furthermore, the coordination office may provide subsidiary first level support to universities without a local point of contact. The range of services includes the four areas of long-term preservation and accessibility, linkage with other sources and provision for secondary use, associated with specific consulting services.

Governance
The Commission's accompanying pilot project has clearly concluded that the critical success factors – a national independent coordination, acceptance of the research community and continuous funding – can best be ensured by the connection to the SAHSS. The Academy will assume responsibility for the central coordination unit of the DaSCH, while the universities have control over the satellite. The legal form of this facility is yet to be determined. An executive board of the partner institutions and a scientific advisory board (analogous to FORS) are planned as governing bodies. This body could form a core group of a national board of stakeholders of the ERI (BFI) sector that the Academy and the SNSF propose to establish in the medium-term in order to coordinate research infrastructures of at least national relevance[72].

Financing
The Commission and Board of SAHSS propose a shared funding model: The national coordination office is being funded with CHF 1 million per annum on the basis of § 11 paragraphs 6 and 7 of the Federal Research and Innovation Promotion Act (FIFG) as a special task of the SAHSS. The remaining CHF 100,000 of the total costs incurred by the national coordination office (incl. long-term archiving) of CHF 1.1 million will be financed through revenue for chargeable services and/or third-party funding. The satellites are funded by project-bound contributions on the basis of § 47 paragraph 1 lit. c) of the Federal Act on Funding and Coordination of the Swiss Higher Education Sector (HEdA), with CHF 600,000 annually. The remaining CHF 400,000 are contributed by the universities from their own resources. The university that hosts the national coordination office may coordinate the applications from the individual universities. The business model of DaSCH foresees that consultations and services for completed research projects of average complexity are offered free of charge. Data transfer and hosting services for ongoing projects will be charged according to an as yet undefined cost model which will provide some revenue. In addition, the DaSCH may apply for research grants to receive some third-party funding. The individual universities contribute financially to the satellites, primarily through the provision of infrastructure and – if possible – staff. The costs given in section 5.2.3.2

72 Compare chapter 3.3 of the multi-year planning of the SAHSS.

for full membership of EU-DARIAH for Switzerland are expected to apply from 2018 onwards. As we expect a slight annual growth in federal contributions to the DaSCH, these membership costs can be financed through this increase and through the DaSCH's own funds. The other financing schemes of the Confederation – DaSCH as a scientific auxiliary service (wissenschaftlicher Hilfsdienst) under § 15 FIFG or exclusively as a coordinated task of universities, e.g. according to § 3 lit. h) HEdA – are not considered appropriate, as mentioned above under point c), as critical success factors cannot be guaranteed in this way.

7. Annex

A. Model for embedding "satellites" (example Lausanne)

The following two diagrams (authors: Béla Kapossy and Nadia Spang-Bovey) are based on discussions with the UNIL Rectorate, and take into account the experience gathered by the UNIL's DDZ/CDPDDZ/CDP unit during the pilot phase. They present a possible model for the successful collaboration between a regional DDZ/CDPDDZ/CDP and UNIL's IT structures. Diagram 1 presents a conceptual model of the various tasks needed to support existing and develop new DB. Diagram 2 shows the required institutional infrastructure.

It proposes a flexible, non-hierarchical structure requiring continuous collaboration between UNIL technical units (front end) and DDZ/CDP (back end), while granting the DDZ/CDP a high degree of autonomy. Lausanne already has a large number of DB, some of which will want to transfer to Salsah/Knora. The model proposes that the migration should take the following steps (Diagram 1, section Green, Fonctions techniques internes à l'UNIL):

Analysis and remodelling of existing DB by UNIL IT consultants;
Development of specific apps and functionalities (front end) by UNIL IT developers;
Transfer to Salsah/Knora by UNIL and DDZ/CDP IT developers;
New DB will be developed in collaboration with faculty-based IT consultants and developers directly in Salsah/Knora;
Existing or new DB that cannot or do not wish to transfer to Salsah/Knora will be supported by faculty-based IT developers and the UNIL technical infrastructure.

For a stable but flexible IT support infrastructure, the model (Diagram 2) suggests that the DH research and teaching infrastructure coordinated by LADHUL should be separated from the technical infrastructure and that the two should be overseen by a Piloting Committee consisting of representatives from the rectorate, the faculties, the technical coordination group and the LADHUL.

The technical infrastructure consists of three different groups: UNIL IT personnel in charge of local DB support (ingénieurs pédagogiques, développeurs R&D, analystes), DDZ/CDP in charge of Salsah/Knora, and FORS. Representatives of each group constitute a "Groupe de coordination technique" under the guidance of a coordinator. A representative of LADHUL has observer status and ensures the flow of information between LADHUL and the IT infrastructure.

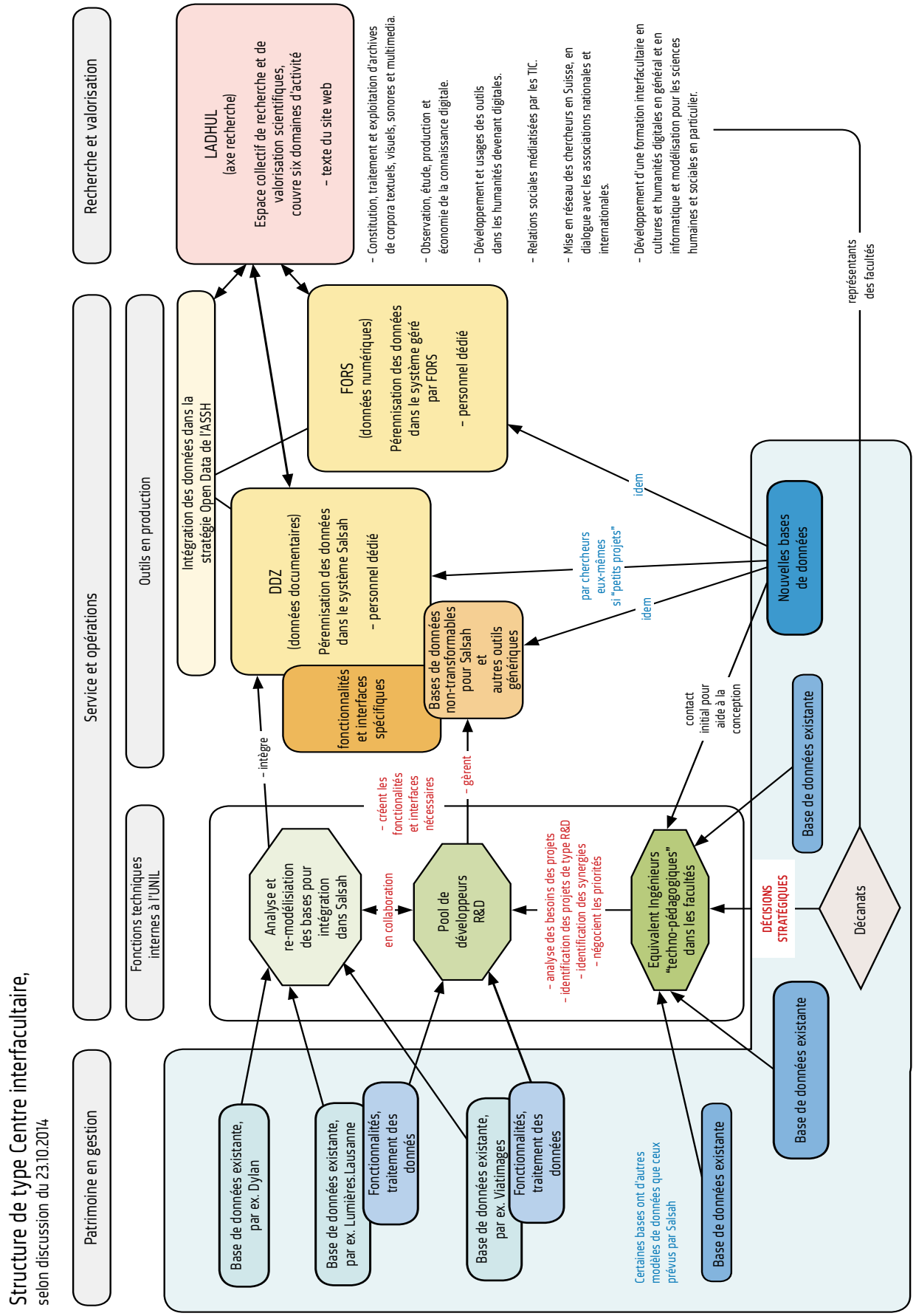


Diagram: Conceptual model of the various tasks needed to support existing and to develop new DB

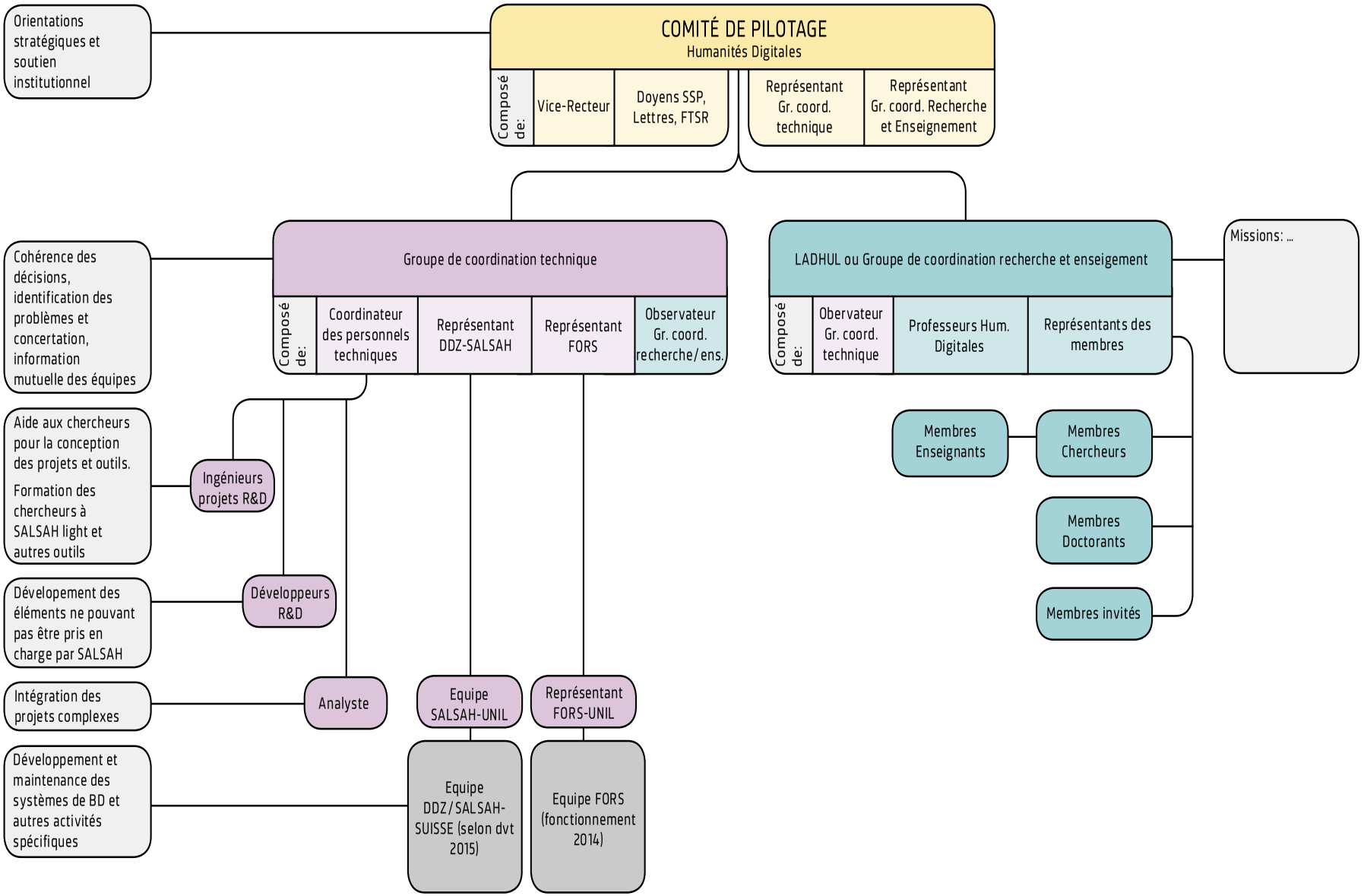


Diagram: Technical infrastructure and LADHUL

B. Detailed description of test cases

Data off the following projects were used as test cases:

A) Lumières.Lausanne (dir. B. Kapossy), UNIL
B) DYLAN (Language dynamics and management of diversity) datasets (dir. A.-C. Berthoud, J. Jacquin), UNIL
C) Viatimages (dir. D. Vaj & Ch. Kaiser), UNIL
D) "Dessins de dieux" database (dir. P.-Y. Brant), UNIL
E) Artists and Books (1880–2015): Switzerland as a cultural platform (dir. Ph. Kaenel & Nathalie Dietschy), UNIL
F) Dokumentationsbibliothek St. Moritz
G) Anton Webern Gesamtausgabe (AWG)
H) Schweizerische Gesellschaft für Volkskunde (SGV)
I) Postkarten Russland
J) Lexicon Iconographicum Mythologiae Classicae (LIMC)
K) VitroCentre Romont
L) Parzival
M) Hotel de Musique
N} Humboldt Edition
O} HyperHamlet

A) Lumières.Lausanne (dir. B. Kapossy), UNIL

Overview

Aim and research interest of the Lumières.Lausanne database[73]:
Lumières.Lausanne's model was conceived both as a research database and a methodological tool geared towards postgraduate research assistants. Its main objects are extensive bibliographic and biographic records on 18th century authors and their literary productions. Literary production is linked to authors, and authors themselves are linked to each other according to a range of defined relationships. As this literary production includes manuscripts, a specific editing tool was designed in order to edit the content. This task, as well as producing new biographic and bibliographical records, is part of the training of students. It should be noted that several colleagues from UNIL and other academic institutions participate in this database.

Technical framework

In its current state, Lumières.Lausanne is a 96-table MySQL database that contains approximately 112,000 records and was developed under the Django framework. Most of the database content is available on line, partially mixed with news-type content. Lumières.Lausanne is an "old database", which means that its use over time has involved several structural changes, numerous adjustments and improvements. However, the database did not undergo any thorough re-examination, hence several minor inconsistencies.

The educational aspect of Lumières.Lausanne comes with a high level of granularity in relation to access (200 different permissions) and includes a validation process which – although not very complicated – we chose not to duplicate in Salsah at the moment.

Main difficulties

In this case, the complete absence of documentation and difficult access to the original developers (from the IT staff of the Université de Lausanne and from a private company) resulted in a long-lasting data analysis in order to understand the data model.
The main contacts with whom we required collaboration were:
the primary developer of Lumières.Lausanne database, a current IT Staff member from the Université de Lausanne;
the secondary developer of the database, an IT technician from a private company, who has been in charge of the development of Lumières.Lausanne for more than a year;
the person responsible for data entry, a graduate assistant who was not familiar with the back end of the database.

Data ingest process

Data analysis and modelling:
The dump of the current state of Lumières.Lausanne MySQL database was available on 3 March 2014. Given the absence of documentation and the difficulty of gaining access to the original developers, data analysis of the MySQL database (96 tables, 112,000 records) took up to three weeks. By the end of one month, we produced a comprehensive model of the current state of the SQL relational database. We were then able to prioritize the various content and aspects of the Lumières.Lausanne database: the research data and the validated educational content gained priority over the news-type content and the validation process geared toward students. The overcomplicated data access and granularity also remains to be completed. However, these aspects have more to do with the usability of the database once data has been ingested in Salsah/Knora than with the sustainability of the research data itself.

73 https://lumieres.unil.ch

RDF modelling and preparing Salsah for data import:
The RDF modelling of the SQL database and the preparation of Salsah/Knora using the administration interface to create the resources, properties, selections and hierarchical lists, and to attach properties to relevant resources, took nearly two weeks.

Data import:
Data import was made in Basel's DHLab. The process required a full collaboration between the person in charge of data analysis and modelling and the person responsible for writing the import script. It took 11 days to complete the process and resulted in a Python script with embedded

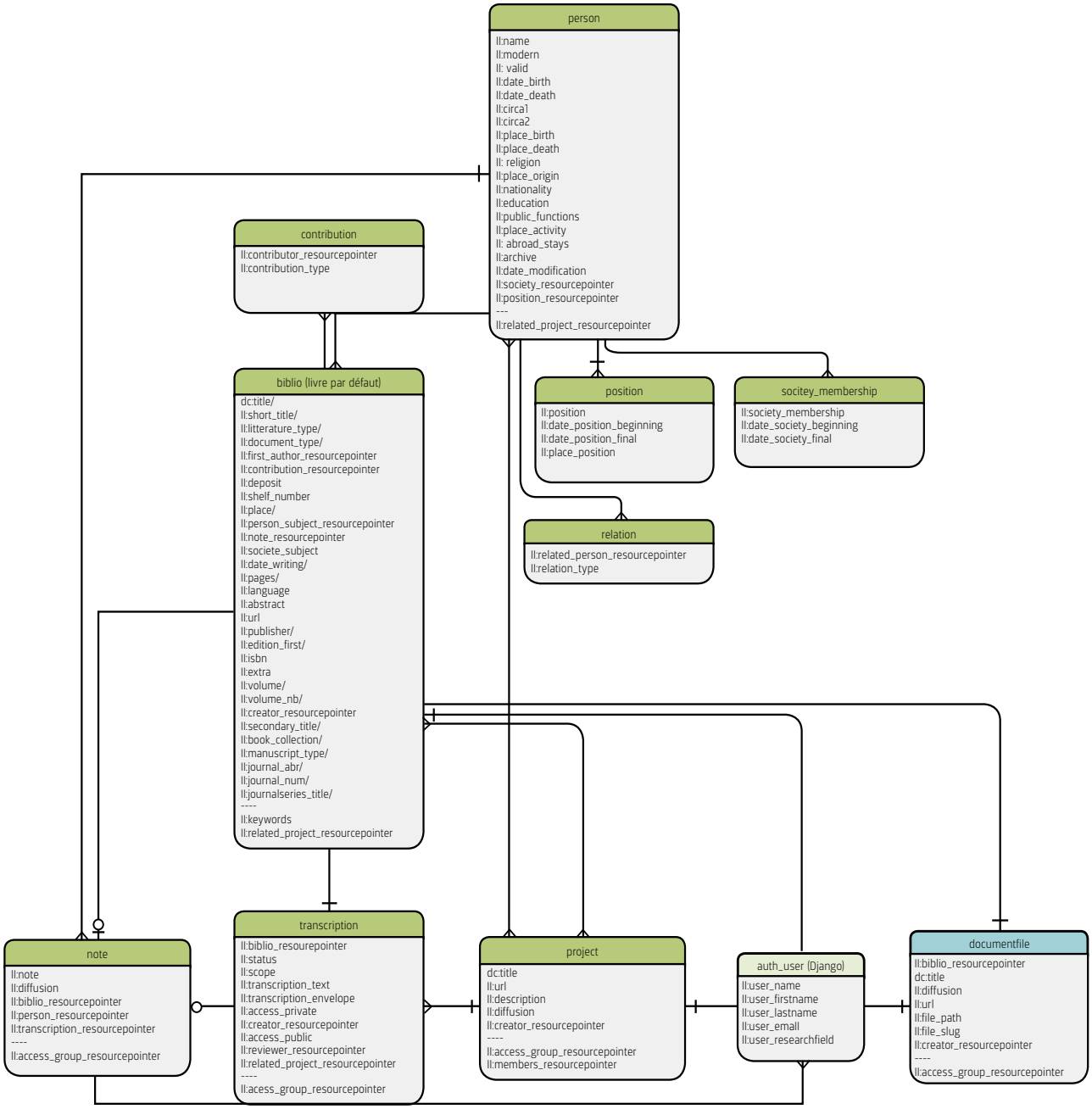


Diagram: Modelling of Lumières.Lausanne in Salsah/Knora RDF

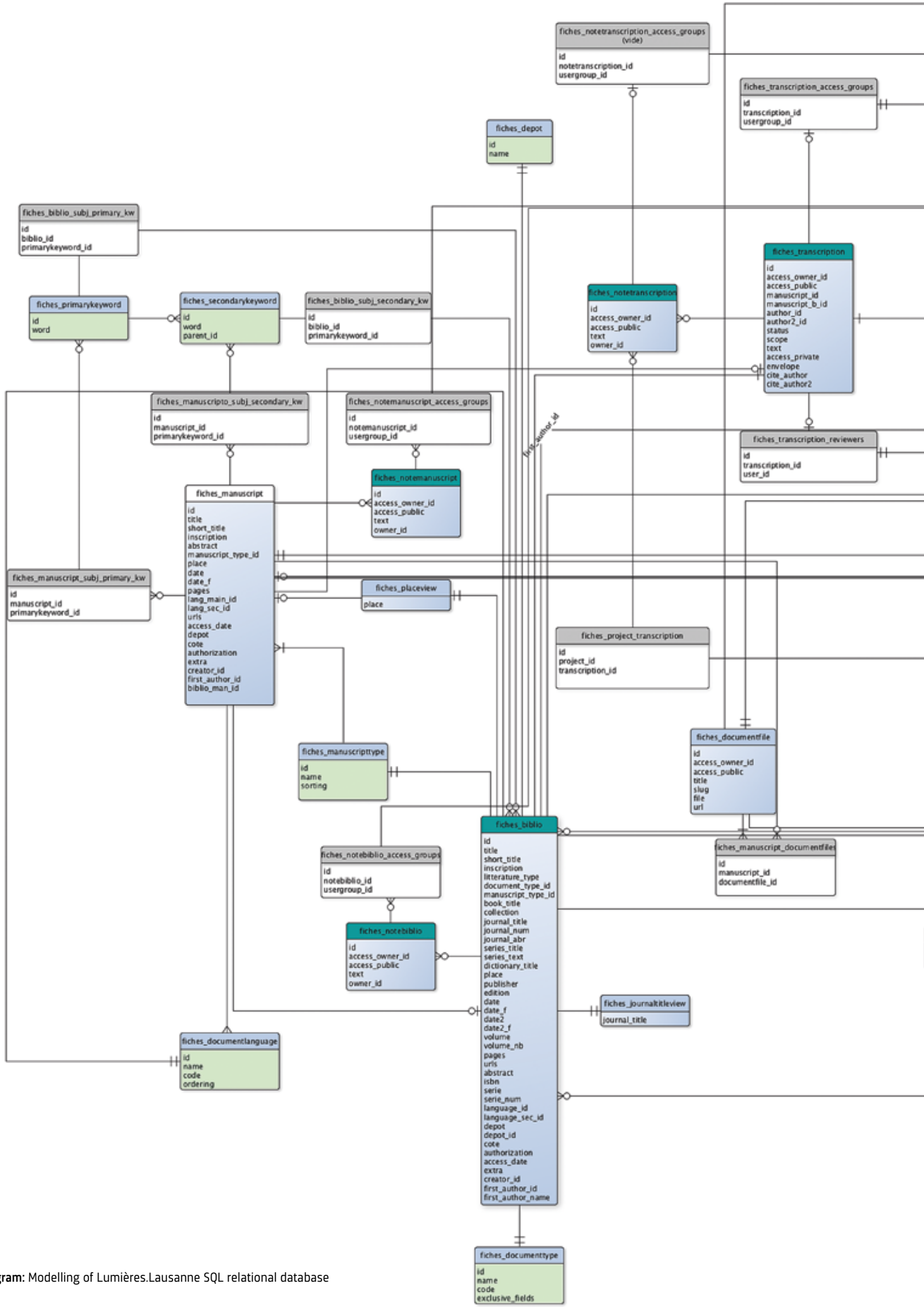
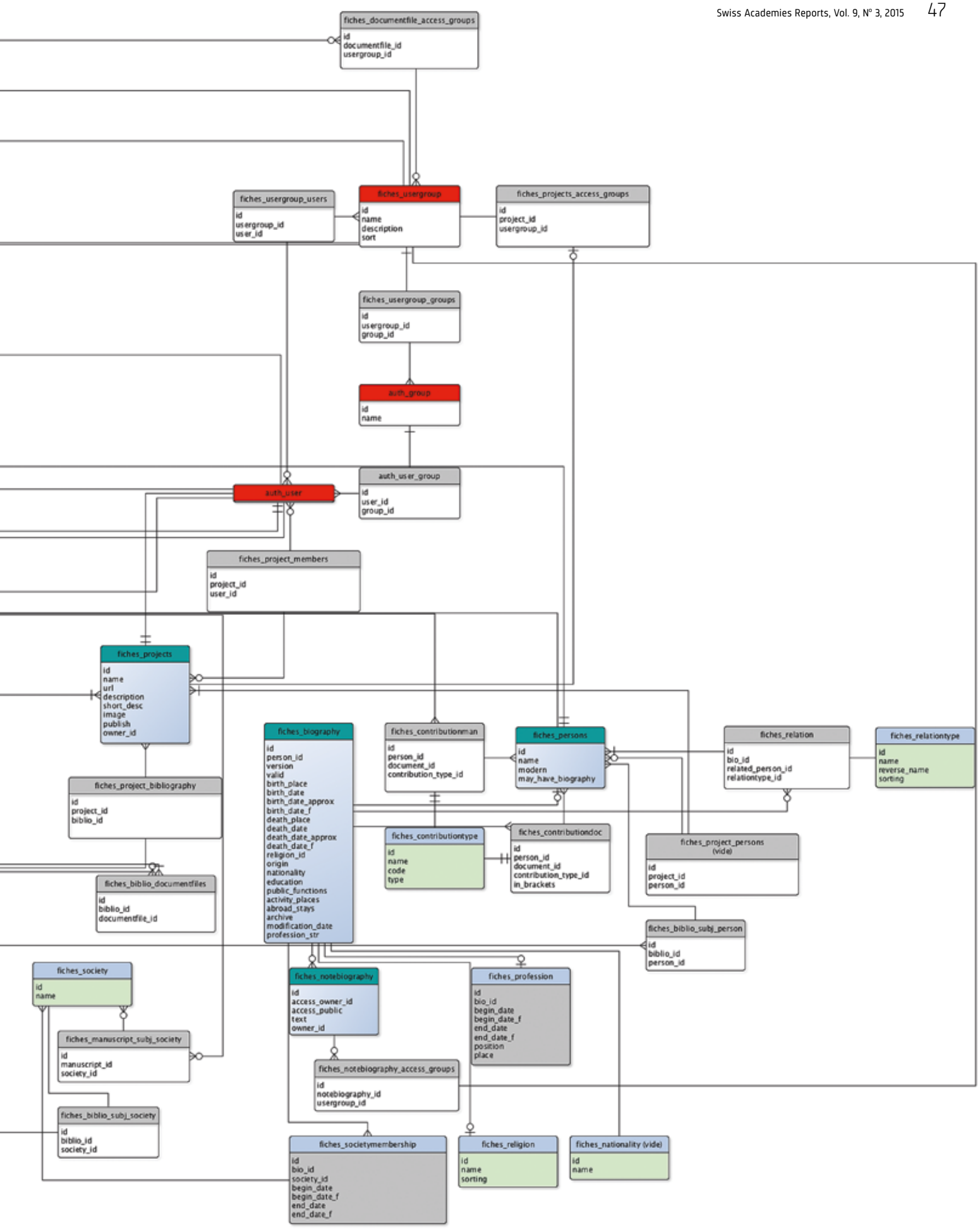


Diagram: Modelling of Lumières.Lausanne SQL relational database


fiches_documentfile_access_groups
id
documentfile_id
usergroup_id
fiches_usergroup_users
id
usergroup_id
user_id
fiches_usergroup
id
name
description
sort
fiches_projects_access_groups
id
project_id
usergroup_id
fiches_usergroup_groups
id
usergroup_id
group_id
auth_group
id
name
auth_user
auth_user_group
id
user_id
group_id
fiches_project_members
id
project_id
user_id
fiches_projects
id
name
url
description
short_desc
image
publish
owner_id
fiches_biography
id
person_id
version
valid
birth_place
birth_date
birth_date_approx
birth_date_f
death_place
death_date
death_date_approx
death_date_f
religion_id
origin
nationality
education
public_functions
activity_places
abroad_stays
archive
modification_date
profession_str
fiches_contributionman
id
person_id
document_id
contribution_type_id
fiches_persons
id
name
modern
may_have_biography
fiches_relation
id
bio_id
related_person_id
relationtype_id
fiches_relationtype
id
name
reverse_name
sorting
fiches_project_bibliography
id
project_id
biblio_id
fiches_contributiontype
id
name
code
type
fiches_contributiondoc
id
person_id
document_id
contribution_type_id
in_brackets
fiches_project_persons
(vide)
id
project_id
person_id
fiches_biblio_documentfiles
id
biblio_id
documentfile_id
fiches_biblio_subj_person
id
biblio_id
person_id
fiches_society
id
name
fiches_notebiography
id
access_owner_id
access_public
text
owner_id
fiches_profession
id
bio_id
begin_date
begin_date_f
end_date
end_date_f
position
place
fiches_manuscript_subj_society
id
manuscript_id
society_id
fiches_notebiography_access_groups
id
notebiography_id
usergroup_id
fiches_biblio_subj_society
id
biblio_id
society_id
fiches_societymembership
id
bio_id
society_id
begin_date
begin_date_f
end_date
end_date_f
fiches_religion
id
name
sorting
fiches_nationality (vide)
id
name

complex SQL queries in order to get the data from the SQL database and to import it into Salsah/Knora.

Value of Lumières.Lausanne as a test case for the data ingest process in Salsah/Knora

Lumières.Lausanne is definitely an evolved relational database, whose understanding was complicated and slowed down by both a lack of information and inconsistencies. Its main interest regarding a test case for the DDZ project was the complexity of its architecture/model, the high number of records, and the content itself: bibliographic and biographic data are a staple in humanities research data. The data ingest process of Lumières.Lausanne proved that Salsah/Knora is capable of receiving and hosting data of this type. Nevertheless, if the data imported in Salsah/Knora were to become the real Lumières.Lausanne, some more adjustments would be required. A versioning and a validation process to control students' data entry should be integrated, in addition to a custom data access model.

Moreover, the data ingest process somehow reassessed the development choices made for the RDF-implementation or raised new ones: users coming as a whole resource, "clickable" path to a document stored in the server, a hybrid property between resource pointer and selection, etc. The embedded editing tool in Lumières.Lausanne is also a further development track that might be pursued in the next step of the DDZ project.

Institutional and decentralized working model questions raised in the process of data ingest

The mid-term availability and accessibility of Lumières.Lausanne research data could have been considered secure because the database is hosted by an institutional server in Lausanne (UNIL). However, the further development and maintenance of the database seemed to have reached an impasse. Indeed, as the local IT staff at the University of Lausanne did not have the skills or the time required to undertake the development of Lumières.Lausanne, this task was transferred to a private company a few years ago. As a result, the costs charged by the company began to exceed Lumières.Lausanne's resources. Moreover, the fact that Lumières.Lausanne's principal investigator was forced to rely on a developer external to the Université de Lausanne created additional obstacles. The private company was not allowed to access the institutional server hosting the database. Thence, an institutional IT technician at the Université de Lausanne had to implement the database modifications on a regular basis.

Being the first test case processed in Lausanne, Lumières Lausanne gave the DDZ team the opportunity to test a decentralized working model. A new server dedicated to Salsah/Knora was opened on 15 June 2014 at the Centre Informatique in Lausanne as the first satellite to Basel's DHLab.

Data analysis of the current state of the Lumières.Lausanne database and data modelling in RDF were achieved in Lausanne. However, data import was processed in Basel with the DHLab's help. This point proved that an embedded IT technician in Lausanne would be crucial for the DDZ project and would very much improve the efficiency of the work undertaken here.

B) DYLAN (Language dynamics and management of diversity) datasets (dir. A.-C. Berthoud, J. Jacquin), UNIL[74]

Overview

Aim and research interest of the DYLAN datasets:

The DYLAN corpus test case happens to consist of two separate datasets, both belonging to the same European project, funded from 2006 to 2011, and involving 19 research partners in 12 countries.

The DYLAN main corpus (thereafter DYLAN-main) gathers all the administrative and dissemination documents related to the European project whereas the DYLAN-Lausanne corpus consists of working data produced in order to address one "task" in one "work package" of the DYLAN project.

The DYLAN project focused on multilingualism in European institutions, and DYLAN-Lausanne's approach mainly concerned multilingualism in higher education institutions.

Technical framework

DYLAN-main dataset consists of 935 files, distributed in a file tree of 221 repositories on 5 hierarchy levels. DYLAN-Lausanne consists of 12,282 files distributed in a file tree of 222 repositories on 8 hierarchy levels.

Main difficulties

None of these datasets are databases. In fact, all the current work undertaken with the two researchers responsible for these datasets (one researcher per dataset) was counselling and education regarding the best way to arrange and structure the data so that data could be further ingested in a database – namely Salsah/Knora.

A short interview with the ex-project manager (also responsible for the data storage in a private company [SCI-PROM]) made it clear that she has not a lot of time to help understand the main threads behind DYLAN-main file tree.

Given their total lack of experience in this matter, the whole process of counselling and supporting the researchers takes some time, both being well aware of the fact that their reflections regarding data architecture and modelling might be pioneering work that is eventually re-used by some of their colleagues.

74 http://www.dylan-project.org/

Moreover, given the fact that the DYLAN project had come to an end a long time ago, both researchers are occupied by their academic careers, and the on-going reflection on their archives is not a top priority – even though the researcher responsible for DYLAN-Lausanne was able to find some funding for the appointment of an assistant student who will be in charge of data entry. It is very likely that, once the data modelling is set, data entry will take some time, especially in relation to DYLAN-main datasets, because some properties linked to secondary resources (not the documents themselves) will require data to be dug and extracted from the current website, and copy-pasted. This division of labor is not completely clear yet: does it fall under the responsibility of the DDZ team during the pilot project or within the researcher's remit? Most likely, tasks will be equally distributed.

If the DYLAN project had been involved in the beginning, our collaboration would have arrived at tangible results more quickly because researchers would have been more deeply involved in the discussions.

Data ingest process

Data analysis:

To date, the main work undertaken in relation to DYLAN data has been data architecture and modelling counselling. However, this task could not have been achieved without data analysis as a first step.

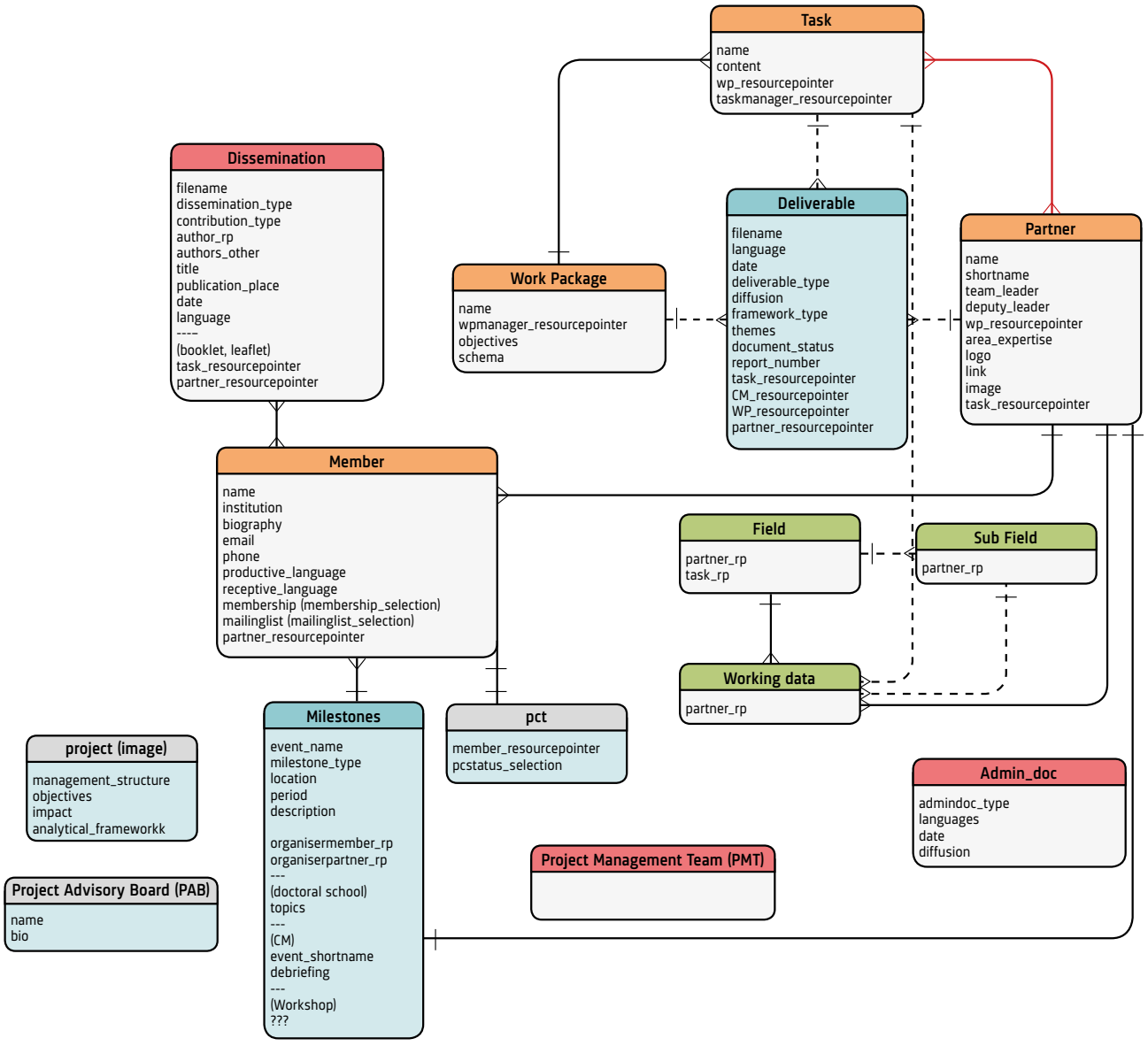


Diagram: Modelling of DYLAN datasets in Salsah/Knora RDF

We visited the DYLAN-Lausanne inventory on 23 March 2014 and received an inventory of the files and data from DYLAN-main hosted by SCIPROM on 25 March 2014. The former required data to be cleaned as a top priority (remove duplicates, collect missing data, archiving raw data [Audacity repositories]). This dataset was deemed satisfactory on 30 September 2014 and a new inventory was undertaken (367 files replaced the former 12,282 files, file tree remains nearly unchanged).

RDF modelling:
Meanwhile, the researchers initiated a discussion with some of their peers in order to try to define relevant metadata regarding their corpus – especially regarding DYLAN-Lausanne working data. Hard work and lengthy discussions have since been necessary to come to an agreement regarding the main resources and their properties. We estimate that it took us (researchers and DDZ team) up to one full working week to settle these questions – and some uncertainties remained that will be clarified during the data entry process.
As for the RDF modelling, we have been working with a view to merging both datasets into a single one. In the diagram above, the resources related to DYLAN-Lausanne appear in green.

It is very difficult to give an estimate of the time (around two weeks) needed to come up with this result because this work was primarily carried out intermittently. Once a modelling proposition was made, it was systematically submitted for the researchers' approbation, because it is supposed to address their needs specifically. Given that discussions take time, it can sometimes be difficult to find a slot quickly. This coming and going was a staple during this phase of the work.

Data import:
Whereas the final modelling of DYLAN-main still requires approval for some details, data modelling of DYLAN-Lausanne has been achieved. The assistant student appointed will undertake data entry on a spread sheet designed specifically for this purpose (with controlled vocabulary, etc.). An initial training session on data entry is planned for 16 December 2014. The assistant student will then be left on her own and a revision will be made on a regular basis by both her supervisor and our team. Once data entry has been completed, data ingest will begin as usual.

Value of DYLAN datasets as test cases for data ingest process in Salsah/Knora

Some of the DYLAN-main corpus had been displayed on a website (http://www.dylan-project.org/) by the project manager of DYLAN project, who currently owns a private company (SCIPROM). All of the European project material has been made available to the project participants via a login and password (http://www.dylan-project.org/Dylan_en/members/members.php). However, SCIPROM would not be in a position to maintain the DYLAN material storage and accessibility beyond 2015, unless any further funding can be found – and this will not happen.

As for DYLAN-Lausanne, by the time we contacted the researcher in charge, data were hosted in a repository on his personal computer. He was recovering from a crash, gathering data scattered across various media in order to rebuild the original archive and store it on the personal storage space allocated by the Université de Lausanne (Centre Informatique).
Given the obvious defects of the data management plan highlighted by both DYLAN datasets, the DYLAN project looks like a customized test case for Salsah/Knora and the DDZ pilot phase.
In terms of the data itself, DYLAN-main material is comprised almost exclusively of text in various formats, spread sheets and small images. DYLAN-Lausanne consists of much more heterogeneous and complicated material: mainly video, audio and text files. Most of these 3 media had been linked and synchronized using the ELAN software (generating a *.eaf file), a tool for annotation on video and audio resources (https://tla.mpi.nl/tools/tla-tools/elan/). The ingestion of DYLAN-Lausanne into Salsah/Knora will require the development of a bespoke front end if the data are to be re-used and displayed in Salsah/Knora. The ongoing research in Basel's DHLab has already proceeded in this direction. Moreover, the audio resources collected during the project also raise the question of the sustainability of their format. By the time we got in touch with the DYLAN team, Audacity had become the only software that can be used to read and convert *.wma audio files.
In terms of the DYLAN-main dataset, the interest in this material as a test case – apart from the fact that Salsah/Knora would address its sustainability issues – does not lie in the technical field. But all European project coming roughly with the same architecture, part of the modelling and vocabulary ("work package", "task", "partner", "milestone", "deliverable", "dissemination", etc.) created in order to fit the DYLAN-main datasets will be re-usable for any European project willing to archive its data using Salsah/Knora.

Institutional and decentralized working model questions raised in the process of data ingest

It has been quite difficult to deal with the restrictions set by the Centre Informatique of the Université de Lausanne. In this phase of the work on DYLAN-main dataset, we needed to collect and host the data previously stored by SCIPROM. Finding a satisfactory way to access the Salsah/Knora server at the UNIL has been quite difficult, and communication with the Centre Informatique in this matter took two weeks to come to a more or less successful conclusion.

Both DYLAN datasets have raised the question as to what would be a reasonable division of the tasks. DYLAN is a complete project and, as such, it proved difficult to find funding for secondary but crucial tasks such as data entry. Had DYLAN been a project in the beginning, the question would not have come up. But the DDZ team – and a Swiss DH Center even more so as a service provider – will most likely have to deal with the finished project with no funding to complete basic tasks such as data entry. How will we address such a question?

C) Viatimages (dir. D. Vaj & Ch. Kaiser), UNIL

Overview

Aim and research interest of the Viatimages database:
Viatimages was developed from 2003 onwards in relation to the Viaticalp project, initiated in 2002. Viatimages (http://www2.unil.ch/viatimages/) is currently a multilingual database (4 languages) based on an extensive geographical thesaurus. Its current content is a collection of any kind of landscape reproductions, cross-matching with the bibliographic resources from which the reproduction has been taken. The peculiarity of this data collection is that, whenever possible, the object represented on photographs or drawings is georeferenced and can be located on Google Maps (and in the future on Swiss Topo).

Technical framework

By the time the DDZ team and the Viatimages team met (21 May 2014), the developer of the database (who is a full time lecturer in cartography and geovisualization at the Université de Lausanne) was in the process of re-implementing the Php/MySQL database into PostgreSQL, under the Django framework.
Viatimages, under PostgreSQL, is a very clean database in accordance with best practices. The dump we were given consists of 54 tables with a data sample sufficient for the DDZ team to test the ability of Salsah/Knora to ingest the categories of data addressed by Viatimages.

Main difficulties

The researchers responsible for the database were reluctant to give full access to their data. They have been asked many times by various colleagues and institutions to share the data model of Viatimages and they have consistently refused. With the guarantee that Salsah/Knora would only access data using proxy objects, and that Viatimages will remain the only real data repository, we came to an agreement and were given a dump of the database re-implemented in PostgreSQL on 1 July 2014.

As usual, we were forced to deal with the researchers' busy schedule: it was not until 2 months after the initial contact that we were given the dump, and yet we had to wait until the search module for the Viatimages API was developed. Finalized on 15 October 2014, this module was crucial for accessing data using proxy objects.

Data ingest process

Data analysis and modelling:
Once the PostgreSQL dump was in our possession, data analysis was relatively quick: it took up to two days to come up with an exhaustive model of the database.

Data import:
The work on Viatimages database is still in progress. The various tasks we are involved in require us to prioritize, and Viatimages was temporarily excluded.

Value of Viatimages as a test case for data ingest process in Salsah/Knora

Multilingualism, extensive geographical thesaurus, georeferencing, and the fact that Viatimages is a renowned database are several aspects justifying its selection as a test case in Lausanne. Moreover, our solution for data access in Viatimages via proxy objects provided another opportunity to demonstrate the efficiency of a technical procedure already successfully used in different cases in Basel (as with e-codices, for example).

Questions about institutional and decentralized working models raised in the process of data ingest

It might be worth noting that developing the Viatimages database in the first place, and implementing it in PostgreSQL after it has been in operation for several years, falls within the remit of the two researchers who were involved in the Viaticalp project from the very beginning. It was made clear from several discussions with these researchers that they could not find the long-term reliability and support they would have needed from the institutional IT staff at the Université de Lausanne to sustain their project. Fortunately, one of the researchers had the required IT skills to set up a SQL relational database.

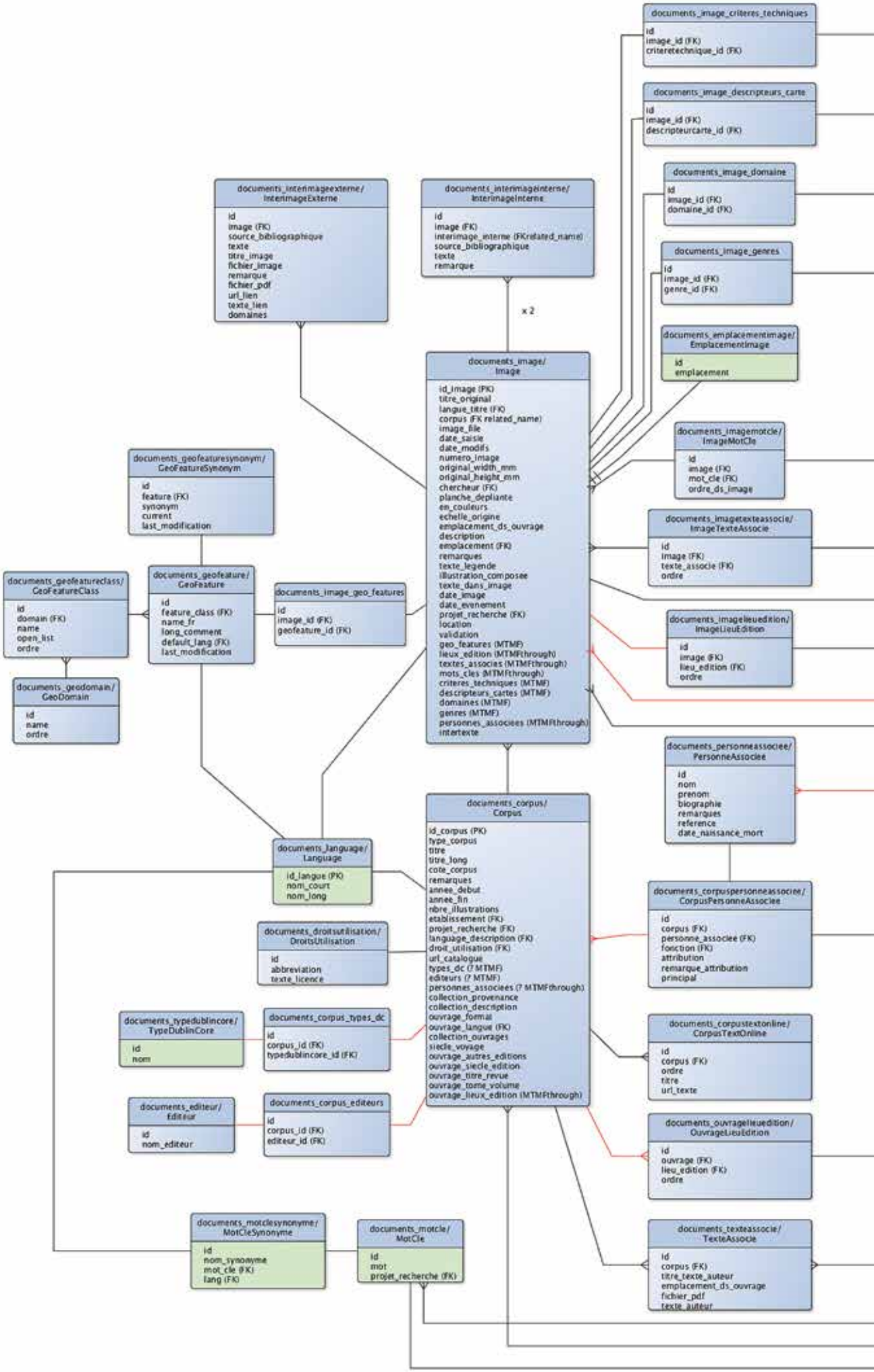
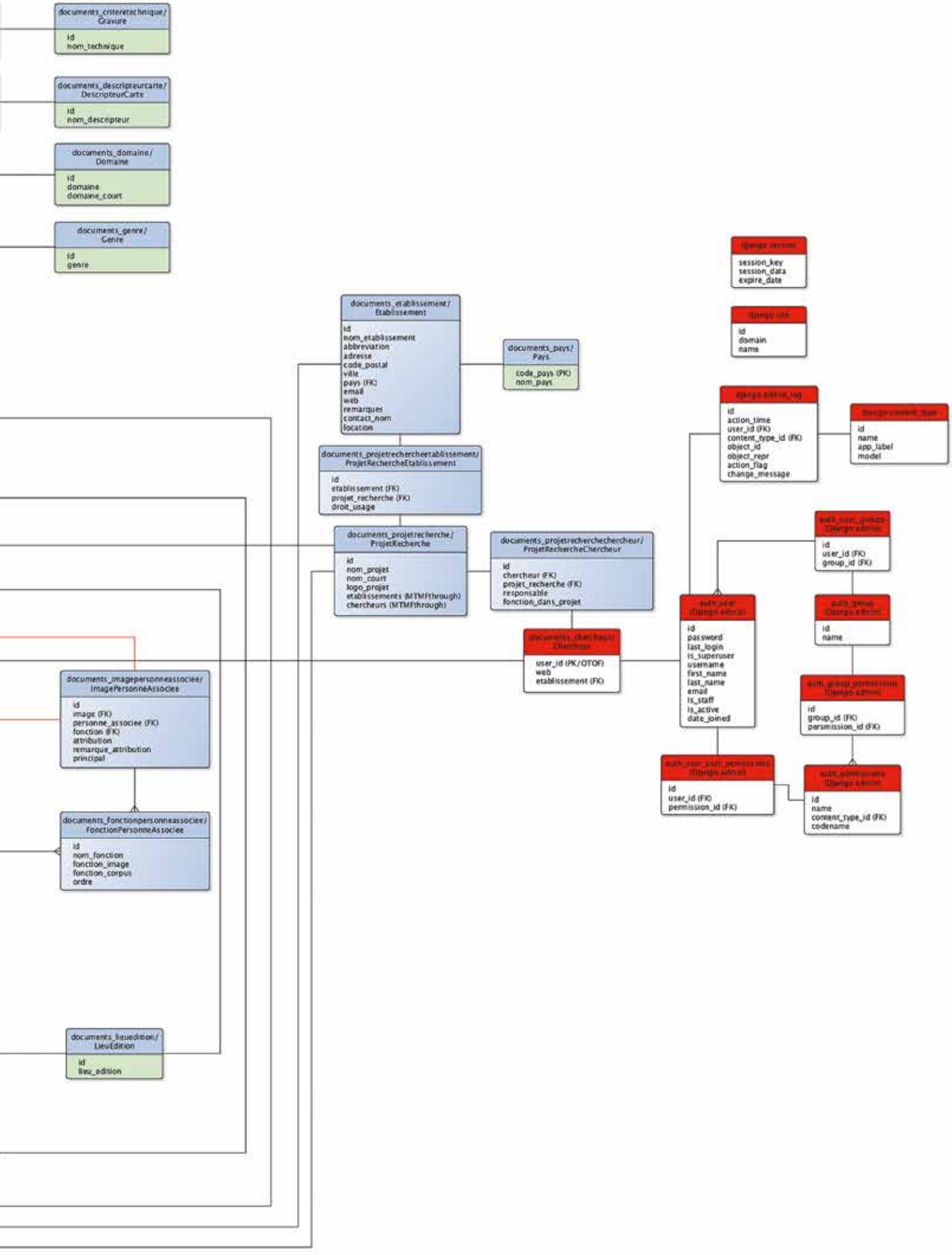


Diagram: Modelling of Viatimages PostgreSQL database


documents_critieretechnique/
Gravure
id
nom_technique
documents_descripteurcarte/
DescripteurCarte
id
nom_descripteur
documents_domaine/
Domaine
id
domaine
domaine_court
documents_genre/
Genre
id
genre
documents_etablissement/
Etablissement
id
nom_etablissement
abbreviation
adresse
code_postal
ville
pays (FK)
email
web
remarques
contact_nom
location
documents_pays/
Pays
code_pays (PK)
nom_pays
session_key
session_data
expire_date
id
domain
name
id
action_time
user_id (FK)
content_type_id (FK)
object_id
object_repr
action_flag
change_message
id
name
app_label
model
document_projetrechercheetablissement/
ProjetRechercheEtablissement
id
etablissement (FK)
projet_recherche (FK)
droit_usage
documents_projetrecherche/
ProjetRecherche
id
nom_projet
nom_court
logo_projet
etablissements (MTMFthrough)
chercheurs (MTMFthrough)
documents_projetrecherchechercheur/
ProjetRechercheChercheur
id
chercheur (FK)
projet_recherche (FK)
responsable
fonction_dans_projet
user_id (PK/OTOF)
web
etablissement (FK)
id
user_id (FK)
group_id (FK)
id
password
last_login
is_superuser
username
first_name
last_name
email
is_staff
is_active
date_joined
id
name
id
group_id (FK)
permission_id (FK)
id
user_id (FK)
permission_id (FK)
id
name
content_type_id (FK)
codename
documents_imagepersonneassociee/
ImagePersonneAssociee
id
image (FK)
personne_associee (FK)
fonction (FK)
attribution
remarque_attribution
principal
documents_fonctionpersonneassociee/
FonctionPersonneAssociee
id
nom_fonction
fonction_image
fonction_corpus
ordre
documents_lieuedition/
LieuEdition
id
lieu_edition

D) "Dessins de dieux" database (dir. P.-V. Brant), UNIL

Overview

Aim and research interest of the "Dessins de dieux" database:
The very first collection of these drawings of gods made by children was gathered in 2003–2004. To date, this bilingual database currently displays several collections by children from Switzerland, the United States, Russia, Japan and Romania (http://dddtiresias.unil.ch/dessinsdedieux/).
The potential users of the database come from all these countries and beyond, when the network research is extended.

Technical framework

The current MySQL database is new implantation of a former Filemaker Pro file. Our dump consists of 11 tables of 6,332 records, with most of the records belonging to one single table and 4 tables being empty. Recent additions have been made that are not accounted for in our dump.

Main difficulties

The principal difficulty has been to find a reliable person who could account for the choices made for the data structure. Given that 4 of 11 tables are empty, we have been asking for an up-to-date dump in order to clarify the uncertainties of data modelling. However, the maintenance of the database is no longer guaranteed because the student in charge has terminated his contract. Therefore, no one has been able to provide us with a new dump.

Data ingest process

Data analysis and modelling:
The first contact with this project's team was made on 20 May 2014, whereas the project's team was in the process of re-implementing their database.
The dump we were expecting arrived on 3 September 2014, with no documentation.

An attempt at modelling was made that requires the confirmation that a new dump would provide.

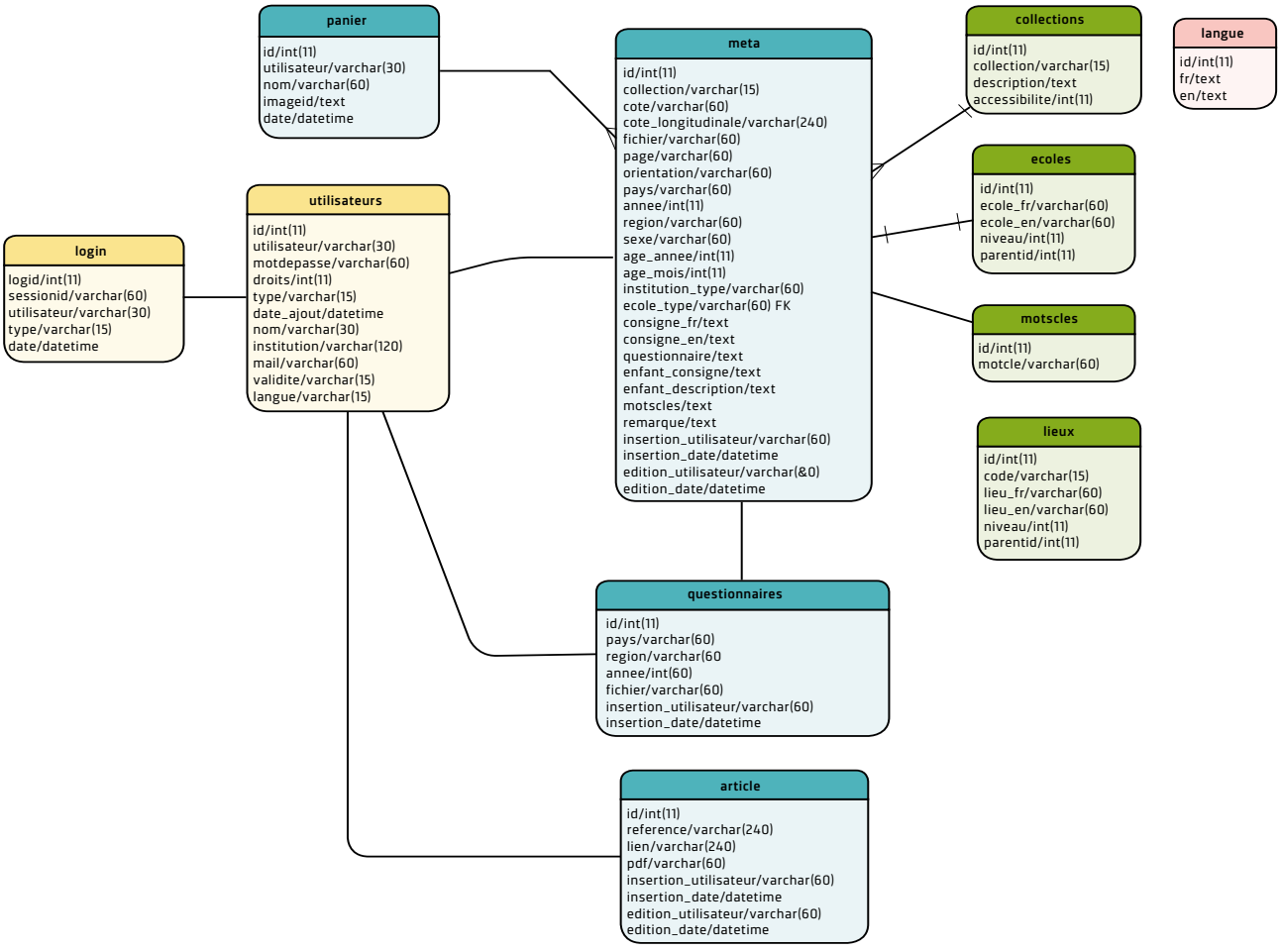


Diagram: Modelling of "Dessins de dieux" PostgreSQL database

RDF modelling and getting Salsah reading for data import:

Given that some aspects of the relational database modelling have to be clarified, the RDF modelling is only a proposition.

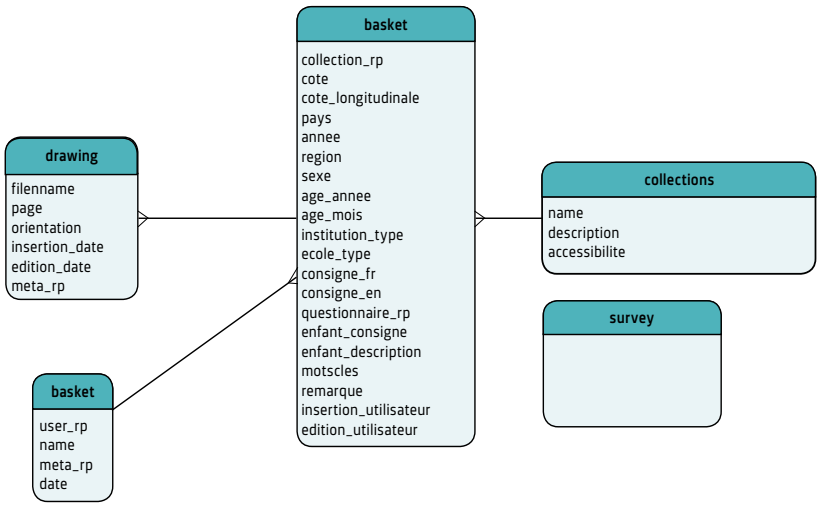


Diagram: Proposition of modelling of "Dessins de dieux" database in Salsah/Knora RDF

Data import:
Data import will be the next step once the PostgreSQL and the RDF modelling have been set up.

Value of "Dessins de dieux" database as a test case for data ingest process in Salsah/Knora

The database in its current state is rather basic. However, it has potential for interesting future development, including the integration into the front end of image computational algorithms. Being accessed from various foreign countries, the database could also be interesting for Salsah/Knora.
Furthermore, if we take into account the difficulties experienced by the principal investigator in finding reliable IT staff to sustain the database, the backup of these research data collections (which remains the primary aim of Salsah/Knora) can be considered a staple.

Questions about institutional and decentralized working models raised in the process of data ingest

The MySQL implementation of the Filemaker Pro database had been the subject of a mandate assigned to an assistant student at the Université de Lausanne who is also an independent designer and developer. It is most likely that this choice came up because, at the moment of implementation, none of the IT staff of the Université de Lausanne were available to get involved in this project.
However, once the new implementation had been achieved, the original developer (who found another appointment outside of the university) was replaced by a new student in charge of maintenance. A few weeks later, the new student terminated his contract, leaving the project's principal investigator with no IT support.
Since then, it has been impossible to provide a satisfactory answer to the few questions raised by the architecture of the database.

E) Artists and Books (1880–2015): Switzerland as a cultural platform (dir. Ph. Kaenel & Nathalie Dietschy), UNIL

Overview

Aim and research interest of the "Artists and Books project":
The "Artists and Books project" is a new project funded by the Swiss National Science Foundation. Initiated in 2013, it will end in autumn 2016.
The aim of this project is, in collaboration with the Swiss National Library, to help to define the object "book of artist" within the time span 1880–2015. Data will be collected from sources including the SNL Helveticat catalogue and input into a research database customized for the project. The choice of the more relevant metadata to describe the resource will facilitate a more accurate definition of the "book of artist" in general.

Technical framework

This project involved close collaboration with the SNL and other libraries. One peculiarity of the project was to

deal with the MARC1 format used by libraries in general (and by the SNL in particular) to input data in RDF/Salsah/Knora, and to provide an opportunity to output data from RDF/Salsah/Knora into MARC21 in order to feed libraries' catalogues with new metadata.

Main difficulties

Apart from the MARC21/RDF conversion, which represents a challenge for Salsah/Knora, no real difficulty was raised at this stage in relation to our forthcoming collaboration with this project.
However, the process of convincing the project team to collaborate with us took some time. Two contacts (7 October 2014; 27 October 2014) and a work session (15 December 2014) with staff from the SNL and three members of the DDZ team were required.

Data ingest process

So far, the work carried out regarding this database has consisted of advice and counselling, and advocating the Salsah/Knora and RDF option for a research-oriented database.

Value of the "Artists and Books project" as a test case for data ingest process in Salsah/Knora

Several aspects of this project are very interesting for the DDZ project. For the first time at the Université de Lausanne, we will be in charge of data for a project from the very beginning. This means that data modelling, data input, etc. will take place as part of ongoing research work. Consequently, researchers and assistants for data entry will be willing to discuss at any stage the conception of a customized data modelling, and the dedicated front end for the project. Salsah will therefore be the main and only data tool for the researchers working on this project. This timetable coincided with the appointment of an embedded IT technician in Lausanne so that the DDZ team in Lausanne could begin to work as a more independent satellite under the supervision of Basel's DHLab.
Moreover, the aspect of data exchange from Salsah/Knora has been left aside until recently. The "Artists and Books project" will address this question. Indeed, the team wants to share the additional data enhancement with catalogues' libraries which imply a process of converting MARC21 to RDF and from RDF to MARC21. Furthermore, researchers are also very eager to download the results of their future database queries: the export process still has to be designed and developed, so that they can access the result of their queries for further and in-depth treatment.

F) Dokumentationsbibliothek St. Moritz

Overview

Aim and research interest of the "Dokumentationsbibliothek St. Moritz":
This is a unique collection of photographs and other material about the history of St. Moritz. The photographs document the development of the site in terms of architecture (historic buildings) and society (sports events, local celebrations, etc.). The photographs are supplied with metadata (of varying depths). Not yet digitized, but there are plans to include journal articles, movies, etc. The digital library is open to the public and for research. An ETHZ research project is currently using these assets in its research.

Technical framework

The first database was developed in the early 2000s when digitization began. It was based on an older version of PHP and MySQL with an integrated web front end which was used both for user access and management of the digital assets. It used a project-specific metadata scheme which simulated the old filing cabinet catalogue. The database contained only lower-quality JPEG images. In order to provide high-quality TIFF images, the library staff had to copy the files from on-shelf hard disks.

Main difficulties

The ageing software has led to many inconsistencies in the data, such as dead or missing links to images, double entries, etc. One of the goals was to clean up the data while transferring it to the repository. This requires numerous test runs and a great deal of interaction with the Documentation Library staff.
The original database structure was of medium complexity and some documentation was available embedded in the source code. Thus the effort to understand the data model and create an RDF model was average.
One special problem was to replace JPEG images with high-quality TIFF images during the data import (which then – during the import process – were converted to JPEG2000 images with loss-less compression). As the naming convention for the original TIFF images was different to that used in the database, a complex translation process had to be programmed.

Data ingest process

Finally, the data ingest process worked flawlessly, and it was possible to rectify a large number of inconsistencies at the same time. During the data ingest process, a few problems arose with some of the high-quality TIFF images which were stored in an incompatible format. We decided to adapt the JPEG2000 conversion in order to cope with this special TIFF variant.

Lessons learned

This project showed that, even in a database which initially appears to be well maintained and have a solid base, many unexpected problems may arise which require close interaction between the "customer" of the repository and the repository staff responsible for the ingest.

The problem with the TIFF files was difficult to track down and required an in-depth knowledge of digital imaging and image file formats.

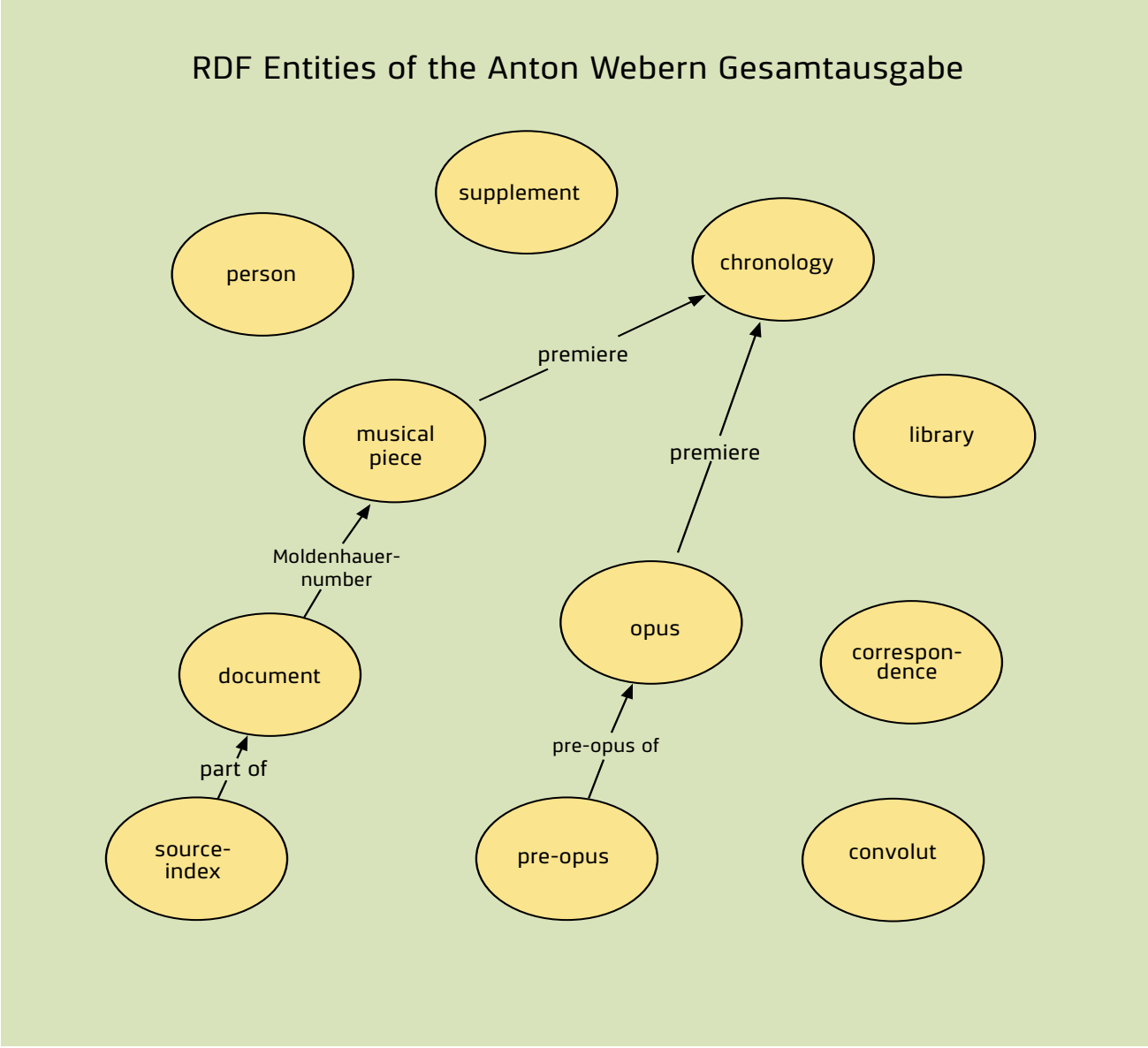
It was obvious from the beginning that the project requires a tailor-made GUI. It has been shown that, with the advanced tools that the platform provides, such a specialised GUI can be developed efficiently in a relatively short time frame.

G) Anton Webern Gesamtausgabe (AWG)

Overview

Aim and research interest of the Anton Webern Gesamtausgabe:

The AWG is a large, long-term edition research project (expected to last around 10 years). The goal is to create a complete edition of the works of Anton Webern as musical writing (to be used by musicologists and performing artists) complemented by an online edition containing all the supplementary information such as letters, personal notes, etc.



Above is an entity diagram of the AWG ontology. Only the directly defined links are shown. Most linkage between these entities is carried out using generic comments.

Technical framework

The project started using a few FileMaker databases (single table) and MS Word for the transcription of supplementary information. This setup soon proved to be inadequate. The AWG team then contacted the DHLab in Basel for a better solution. We suggested transferring all data into the SDHC repository and continuing to work directly in the repository using Salsah.

Main difficulties

The most interesting and at the same time challenging part was to create an optimal data model using RDF. On one hand, the SDHC team had to understand the specifics of the research project (which are the objects of interest, e.g. what is a "Moldenhauernummer", etc.). On the other hand, the AWG team had to learn that, to some extent, the digital data requires strict categorization. Building the data model took around 3–6 months of intensive discussion.

Data ingest process

The existing FileMaker databases were imported using custom-built PHP scripts. Some rectification and cleanup of inconsistencies was necessary, but did not pose any major problems.

Lessons learned

It is very advantageous if a research project decides to work directly in the repository at an early stage. Working together with the SDHC from the beginning results in data structures which are efficient and well adapted to the project. The AWG team is very happy. Up to now, the ontology has become quite extensive and complex, but is very well-suited to the needs of the research.

H) Schweizerische Gesellschaft für Volkskunde (SGV)

Overview

Aim and research interest of data collection of the Schweizerische Gesellschaft für Volkskunde (Swiss Society of Folklore):
The SVG has a large collection of documents (photos, films and written descriptions) of local folk tradition, old craftsmanship, etc. These assets are currently being digitized and annotated. Making these resources available for research has a great deal of potential.

Technical framework

The metadata has been collected using MS Excel. This has proven to be a very poor solution because there were many inconsistencies and errors in the data, mostly due to the fact that Excel is not suited to this task.

Main difficulties

For Import, the Excel files had to be converted to tab-separated lists, which raised several problems. Firstly, it was difficult to find the right settings in order to transfer umlauts and special characters correctly. Secondly, there were several Excel tables with connecting relations. Due to typing errors it was hard to resolve these relations during the import process.

A second problem also arose. In a first attempt to create the proper data model, the SVG insisted on a complex metadata scheme with several dozen properties per entity. Most of these fields remained empty because the metadata was not available, or too tedious to enter. In the end a complete redesign was required in order to simplify the ontology to a manageable complexity.

A third point was a hierarchical concept of geographical locations. At first, the hierarchy was (against the recommendation of the SDHC team) designed down to the individual street address. Finally we decided to connect the SDHC repository to the location service of geonames.org, which provides a RESTful interface for rich geolocations including coordinates, links to Wikipedia, etc.

Data ingest

Because of the inconsistencies and variations of the many Excel tables, the ingest has been quite a painful process. For each batch, the import scripts had to be adapted or rewritten. Finally we were able to agree on a standard format. Together with the redesigned ontology the import is now efficient and without major problems.

Lessons learned

MS Excel is a problematic tool for gathering metadata. Even a primitive database such as FileMaker offers more possibilities to ensure a certain consistency. The ontology should be "as complex as necessary, but as simple as possible"[75]. The redesign has been a success, despite consuming a lot of time and resources. Within the next 12–18 months, around 100,000 documents (mainly images, some movies) are expected to be integrated and annotated and therefore made available for further research.

I) Postkarten Russland

Overview

Aim and research interest of data collection of Postkarten Russland:
This is a very small collection of around 700 postcards from Russia around the turn of the century. It was collected and annotated in a PhD thesis using FileMaker.

75 An aphorism attributed to A. Einstein.

Technical framework
The metadata was collected using FileMaker, the images were referenced by file name and had to be retrieved by hand.

Main difficulties
None – it was a simple, straightforward 2-day project to import this data into the repository. The data modelling was reduced to just one object class.

Data ingest
Small PHP script.

Lessons learned
Such small projects with a limited complexity can be processed very efficiently. Using a template script that can be adapted even with little IT knowledge is sufficient. Due to the very flexible import scheme for images, integrating and combining the images with the metadata is very easy (using the FileMaker solution, the images had to be opened in PhotoShop or a similar program by copying/pasting the file name from the database to the image viewer.

J) Lexicon Iconographicum Mythologiae Classicae (LIMC)

Overview
Aim and research interest in LIMC:
The LIMC is one of the best examples which shows the difficulties but also the chances of preserving long-term access to an important research database. The LIMC is a multi-volume encyclopedia cataloguing representations of mythology in the plastic arts of classical antiquity. Published in a series from 1981 to 2009, it is the most extensive resource of its kind, providing "full and detailed information"[76]. Entries are arranged alphabetically, with black and white illustrations indexed to their respective entries. The work was prepared by international scholars from nearly 40 countries who contributed in their language of choice, resulting in entries written in English, German, French and Italian.
LlMC has been called an "indispensable research instrument"[77], "monumental"[78] and "magnificent"[79].

LlMC also offers a multilingual online database that is updated independently of the print publication. However, the database itself (not its content!) has not been maintained for nearly a decade. The company that was responsible for the development went out of business leaving no documentation behind. It still runs on an old server where even the password is unknown. We therefore decided to migrate this database into the SDHC repository.

There have been several requests, notably by the Center for Hellenic Studies (CHS) at Harvard University, to gain access through the RESTful interface to combine the LlMC data with their digital Homer commentary (Prof. G. Nagy, head of CHS).

Technical framework
The LlMC database uses Microsoft SQL-Server on an outdated Windows server at the central lT services of the University of Basel. Our investigation showed that part of the software was developed using an outdated version of dot.net.nuke, a content management system for the ASP.NET habitat. As a first step, we copied the whole server onto a virtual test system, as the database is still heavily used.

Main difficulties
There are many challenges. The most central is the total lack of any documentation, which therefore requires reverse engineering of the original data model. The only information available was from users of the database. Fortunately, the Dept. of Classical Studies and ancient History has some very knowledgeable users. lt took many meetings, sessions with some key users and experiments in order to understand the very complex data structures.

The data structures are very complex and often inconsistent (e.g. redundancies, unused tables filled with data never referenced, etc.).

As a second step, again with constant interaction with the selected users, we were able to design a data model/ontology for the repository. We are currently implementing it using the RDFS/OWL vocabulary. At the same time, the development of the transfer programs has started. For several technical reasons, the transfer software must be programmed in C# and run on the LIMC server using the RESTful interface of the repository to push the data. We expect the data be transferred at the end of February 2015. Building the special user interface is expected to take around 80–100 hours of development time in order to allow comfortable access to the data.

Data ingest
C# programs using the RESTful interface of the SDHC repository.

Lessons learned
Besides having a very complex and sometimes inconsistent data structure, the lack of documentation is a major

76 Robin Hard, The Routledge Handbook of Greek Mythology (Routledge, 2004), p. 691.

77 Pura Nieto Hernández, Mythology: Oxford Bibliographies Online Research Guide (Oxford University Press, 2010), p. 45.

78 William Hansen, Classical Mythology: A Guide to the Mythical World of the Greeks and Romans (Oxford University Press, 2004), p. 14.

79 Hard, Routledge Handbook, p. 691.

issue. However, in such cases it is extremely important to have access to "power users". Looking at the database from the user's perspective helps 1) to understand the existing data structures and 2) to create a meaningful and efficient RDFS/OWL ontology.

K) VitroCentre Romont

Overview

Aim and research interest of data collection of the VitroCentre:
The VitroCentre in Romont is the national point of contact for the inventory of stained glass. On one hand, it published the inventory of stained glass in Switzerland. Until now, the publication was in the form of a traditional book on paper. However, in future, digital publication will become standard. For some years, FileMaker-based database solutions have been used to prepare the data. This has been the source of some problems: different researchers used different variants of the FileMaker template (with different definitions of the data fields). As the version of FileMaker that was used could not be shared, different copies of each database were used, which, with time, ended up out of sync. The images are available – externally from the FileMaker databases – as high-quality TIFF images. There is a trifold aim for the transfer of the data into the SDHC repository: a) the consolidation of all different databases into one common, sharable and collaborative environment allows for a much more efficient internal workflow. b) The RESTful interface of the repository allows to open selected parts of the data (including lower resolution JPEGs of the images) to the general public through a bespoke portal. c) The exchange of information with other national inventories becomes much easier. In fact, there are plans for a European inventory which might possibly use the platform developed for the SDHC.

Technical framework

The VitroCentre uses a common shared drive to store the different FileMaker databases. If a researcher wants to use or modify it, he/she copies it to his/her local drive to work on it. The result is that the FileMaker databases become inconsistent and to a certain degree incompatible because the data schemes have been extended in a non-systematic way. A first test (which is online) has been carried out with the inventory of the Canton of Geneva. There have been some minor inconsistencies with the data. The main problem is that different object classes have been fattened into one FileMaker file by just copying the data. As an example, the building information was just copy-pasted for each window of a given church. This changed the information about the building as it was copied to all stained glass window entries. This is very cumbersome and error-prone.

Main difficulties

On one hand, the different FileMaker databases are not consistent in terms of the data model. Often, some fields have been added or modified. Since we were not able to have a look at versions at once, often the ontologies had to be adjusted.
On the other hand, the FileMaker files contained a lot of redundancies, because several concepts have been merged into one file. For example, each stained glass window belongs to a building. In the databases, the building information (around 5–10 fields) was just copy-pasted for each window. Taking apart these different concepts during import was quite tricky.

Data ingest

PHP scripts using the ODBC FileMaker gateway.

Lessons learned

Even a "simple" FileMaker-based database with only one FileMaker file may be quite tricky.

L) Parzival

Overview

Aim and research interest in Parzival:
Parzival is an ongoing research project of Prof. Stolz, University of Bern, with quite a long history. On one hand, it consists of a relational database and a webinterface for the editing process (which partially uses TUSTEP software). On the other hand, there have been 3 CDs of 3 different editions. As the first step, we decided to integrate the CDs, starting with "Parzival-Handschrift, Burgerbibliothek Bern, Cos. AA91". The goal is to present the edition in the same way as it is presented on the CDs whose software is slowly becoming obsolete.

Technical framework

The CD is accessed using a web browser. It does not use a database. All information is encoded using either HTML, XML or JavaScript. There are long text files in either format with embedded links and structural information. The images are stored as JPEG.

Main difficulties

There were two challenges:
As all the data was encoded in HTML, XML and JavaScript, it was quite difficult to extract a suitable data model.
Parsers for HTML, XML and JavaScript then had to be written to extract the information and connection between the concepts.

Creating the user interface which simulated the original one from the CD was straightforward using the tools from the SDHC repository.

Data ingest
The parser was written in PHP using a standard XML-parser. The ingest uses the RESTful interface.

Lessons learned
The user interface – i.e. the way a user accesses the data for browsing – may be essential for the reuse of data. The RESTful interface becomes crucial if the data is incorporated into other research at a later stage.

M) Hotel de Musique

Overview
Aim and research interest in Hotel de Musique:
This as yet non-public database contains theatre and opera performances from the 19th century at the Hotel de Musique in Bern. It contains more than 10,000 entries about persons, figures, pieces and performances. It will be very interesting to make this database available to members of the musicology who are interested in the musical and theatrical performances in Bern.

Technical framework
The database was created using FileMaker.

Main difficulties
The database was imported into the repository by an IT specialist with no prior experience of Knora/Salsah. It was interesting to note that, even with the limited documentation available, he was able to create the data modelling, implement the ontology and transfer the data with very little support from the head office in Basel.

Data ingest
The transfer scripts were developed using the python language. The data was directly transferred from Bern into the repository in Basel during the import using the RESTful interface.

Lessons learned
It is possible for an IT specialist with good knowledge in the humanities (or vice-versa) to transfer an existing database into the Knora/Salsah framework with relatively little help from the second level support.

N) Humboldt Edition

The Humboldt Edition is a very recent project. The goal is to use TEI-encoded transcriptions of Humboldt's scientific articles and papers for distant reading. One of the advantages of using the SDHC repository is that the structural information encoded in TEI can be easily separated (and joined again without losing any information) using standoff markup. Thus, using the RESTful interface, statistical and tools for distant reading can be applied with little effort.

We are currently at the stage of creating the adequate RDFS/OWL ontology to represent the data.

O) HyperHamlet

Overview
Aim and research interest in HyperHamlet:
HyperHamlet is a database of references to Shakespeare's most famous play. Structured as a hypertext of Hamlet, it gives access to thousands of extracts from later texts that quote particular lines. Extracts which mention certain characters or scenes can be searched for these names and motifs.

HyperHamlet could be described as a dictionary-in-progress which does not tell us where phrases come from (as other dictionaries do), but rather where Shakespeare's phrases have gone.

References to Hamlet in literature, in the visual arts, in political discourse, and, more recently, in advertising and merchandising can tell us a great deal about the status and understanding of the play.

Technical framework
HyperHamlet uses a quite complex relational database and Sphinx as a full-text index. The application is written in PHP.

Main difficulties
HyperHamlet has a very elaborate user interface which allows anyone to add a reference to any line in the text of Hamlet (crowd sourcing). However, this entry becomes only visible after an editorial team has reviewed and approved the entry. In order to make HyperHamlet available, this interface has to be recreated.

Data ingest
The data ingest was carried out using PHP scripts.

Lessons learned
Since the HyperHamlet database was originally created by Professor L. Rosenthaler in 2003, extensive knowledge about the goals and data structures has become available.

Due to the limited capacity, this project is momentarily on hold. We expect to finish the transfer into the repository, including all editorial functions, in spring 2015.

C. Selected further reading on the DaSCH topic from members of the pilot project (published 2013–2015)

Bornet, P., Clivaz, C., Durisch Gauthier, N., Honoré, E. (eds), *L'Homme augmenté* (eTalks), Université de Lausanne: VITAL-DH (publication in July 2015).

Clivaz, C., "Covers and corpus wanted! Some Digital Humanities fragments", in J. Nyhan (ed.), *Digital Humanities Quarterly, special number* (forthcoming).

Clivaz, C., "En quête de couvertures et corpus. Quelques éclats d'humanités digitales", in Carayol, V., Morandi, F., in *Les Humanités digitales, le tournant des sciences humaines,* Presses Universitaires de Bordeaux (à paraître).

Clivaz, C., Dilley, P., Hamidovic, P. (eds), in collaboration with Thromas, A., *Ancient Worlds in a Digital Culture (Hellenistic Studies),* Washington: Center for Hellenic Studies, Trustees for Harvard University Press (manuscript to be forwarded in April 2015).

Clivaz, C. and Vinck, D. (dir.), *"Les humanités délivrées"*, in *Les Cahiers du Numérique 10 (2014/3)*, 2014.

Clivaz, C., "New Testament in a Digital Culture: A Biblaridion (Little Book) Lost in the Web?", in *Journal of Religion, Media and Digital Culture 3 (2014/3)*, 2014, p. 20–38.

Clivaz, C. and Hamidovic, D., "Critical Editions in the Digital Age", in Ryan M.-L., Emerson, L., Robertson, B. (eds), *The Johns Hopkins Guide to Digital Media and Textuality,* Johns Hopkins University Press, John Hopkins, 2014, pp. 94–98.

Clivaz, C., Gregory, A., and Hamidovic, D. (eds), in collaboration with Schulthess, S., *Digital Humanities in Biblical, Early Jewish and Early Christian Studies (Scholarly Communication 2),* Leiden: Brill, 2013.

Keller, S., Keller, A., Neuroth, H. and Rosenthaler, L., "Project management and sustainable revenue models in the Digital Humanities", in *Proceedings Digital Humanities 2014* (accepted, to be published).

Kilchenmann, A., Rosenthaler, L., "Annotation and Linkage of Motion Picture in an Interactive and Collaborative Environment", in *Archiving 2014,* Society for Imaging Science and Technology (accepted, to be published).

Rosenthaler, L., Fornaro, P. and Clivaz, C., "National Data Curation and Service Center for Digital Research Data in the Humanities", in *Proceedings Digital Humanities 2014* (accepted, to be published).

Rosenthaler, L., Fornaro, P., Clivaz, C., "Data and Service Center for the Humanities", in *DHS (LLC) (1/2015),* Oxford: OUP, (publication in June 2015).

Rosenthaler, L. and Fornaro, P., "Big Data – Bedrohung oder Chance für das Kulturerbe?", in *NIKE Bulletin 6,* 2014.

Rosenthaler, L., "Technische Herausforderungen in den Digital Humanities", in *Bulletin (4/2013),* SAGW, Bern, 2013.

Rosenthaler, L. and Schweizer, T., "Salsah – eine webbasierte Forschungsplattform für die Geisteswissenschaften", in *Bulletin (1/2012),* SAGW, Bern, 2012.

Rosenthaler, L:, "Virtual Research Environments. A New Approach for Dealing with Digitized Sources in Research in Arts and Humanities", in Claire Clivaz u.a. (ed.): *Reading Tomorrow. From Ancient Manuscripts to the Digital Era,* Lausanne 2012, p. 661–670.

Schweizer, T., Rosenthaler, L. and Subotic, I., "Visualisierung von Annotationen und Verknüpfungen in Salsah", in *Geschichte und Informatik,* Chronos Verlag (accepted, to be published).

Schweizer, T., Wassmer, A., and Rosenthaler, L., "Long-term Access to Primary Research Data as a Challenge to Migration", in *Archiving 2014,* Society for Imaging Science and Technology (accepted, to be published).

Schweizer, T. and Rosenthaler, L., "Building Digital Editions on the Basis of a Virtual Research Environment", in *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: HRI Online Publications, 2014. http://www.hrionline.ac.uk/openbook/book/dhc2012 (27.3.2015)

Subotic, I., Kilchenmann, A., Schweizer, T. and Lukas Rosenthaler, "Future Development of a System for Annotation and Linkage of Sources in Arts and Humanities", in *Proceedings Digital Humanities 2014* (accepted, to be published).

Subotic, I., Rosenthaler, L. and Schuldt, H., "A Distributed Archival Network for Process-Oriented Autonomic

Long-Term Digital Preservation", in *ACM Proceedings of the Joint Conference On Digital Libraries 2013,* 2013.

Subotic, I., Rosenthaler, L., Schuldt, H., "A Benchmark for RDF-based Metadata Management in Distributed Long-Term Digital Preservation", in *Proceedings of the 3rd International Workshop on Data Engineering Meets the Semantic Web* (DESWEB, ICDE 2012 Proceedings), 2012.

Terras, M., Verhoyen, D., Clivaz, C., Kaplan, F. (eds), *Digital Humanities in Scholarship* (DH2014 special number) 1 (2015), Oxford: OUP (publication in June 2015).

Vinck, D. and Clivaz, C. (dir.), "Les humanités délivrées", in La *Revue d'Anthropologie des Connaissances 8 (2014/4),* 2014.

"Digital religion out of the book: the loss of the illusion of the 'original text' and the notion of a 'religion of the book'", in *Scripta Instituti Donneriani Aboensis 25:* Digital Religion, B. Dahla, R. Illman (eds.), Turku: The Donner Institute for Research in Religious and Cultural History, 2013, p. 26–41.

"Digital Humanities – un nouveau défi", *Bulletin (4/2013),* SAGW, 2013, p. 32–33. http://www.sagw.ch/dms/sagw/bulletins_sagw/bulletins_2013/SAGW_Web (27.3.2015)

"Internet Networks and Academic Research: the Example of the New Testament Textual Criticism", Clivaz, C., Gregory, A., and Hamidovic, D. (eds), in collaboration with Schulthess, S., *Digital Humanities in Biblical, Early Jewish and Early Christian Studies* (Scholarly Communication 2), Leiden: Brill, 2013, p. 151–173.

D. Legal documents

Rechtliche Bewertung des DDZ

Allgemein

1. Zielsetzungen des DDZ

Das Daten- und Dienstleistungszentrum (DDZ) beabsichtigt, den Zugang zu den Werkzeugen der *Digital Humanities* zu vereinfachen. Dazu gehört einerseits ein Fachportal, welches Forschende bei Fragen unterstützt und ihnen bereits beim Aufbau von Projekten zur Seite steht. Andererseits haben interessante Forschungsprojekte die Möglichkeit, ihre Forschungsdaten durch das DDZ digitalisieren und archivieren zu lassen. In einem zweiten Schritt sollen diese Primärdaten dann aufbereitet und Drittparteien für Sekundäranalysen zur Verfügung gestellt werden.
Sofern die übernommenen Forschungsdaten keine Personendaten enthalten und keine geschützten[80] oder verbotenen[81] Inhalte aufweisen, steht einer Verbreitung zur freien Nutzung (Open Data) grundsätzlich nichts entgegen. Sind diese beiden Voraussetzungen aber nicht gegeben, stellt sich eine Vielzahl von rechtlichen Fragen, von denen hier einige kurz erörtert werden sollen.

2. Open Data und Open Access

Vorerst ist zu unterscheiden zwischen einer Publikation von Forschungsdaten im Sinne von **Open Data** und einer Publikation von Forschungsergebnissen im Sinne von **Open Access**. Ein wissenschaftlicher Artikel oder eine Studie können zwar ebenfalls Forschungsdaten enthalten, doch die rechtlichen Fragen, die sich stellen, sind nur teilweise vergleichbar.
Einer **wissenschaftlichen Publikation** unter einer offenen Lizenz (Open Access) liegt ein Gegenstand zugrunde, dessen Werkcharakter im Normalfall unbestritten ist und dessen Rechte meist beim Urheber oder Verleger liegen. Der Einbezug fremder Werke ist üblicherweise nicht notwendig oder durch die Zitierfreiheit erlaubt.
Bei **Forschungsdaten** ist es umgekehrt. Der Werkcharakter ist umstritten und wird abgesehen von Ausnahmefällen nicht gegeben sein. Es besteht jedoch zumindest bei geisteswissenschaftlichen Forschungsdaten eine grosse Wahrscheinlichkeit, dass Werke von Drittparteien darin enthalten sind, deren Rechte nicht durch den Datenlieferanten übertragen werden können.

80 Immaterialgüterrechte.

81 Z. B. persönlichkeitsverletzende oder rassistische Inhalte sowie verbotene Darstellungen von Gewalt oder Sexualität.

Datenschutz

3. Geltungsbereich des DSG

Mit dem Datenschutzgesetz (DSG) wird versucht, die Persönlichkeit und die Grundrechte einer Person zu schützen. Insofern stellt die Datenschutzgesetzgebung eine Konkretisierung von Art. 28 ZGB[82] (Persönlichkeitsschutz) und 13 BV[83] (informationelle Selbstbestimmung) dar. Der Geltungsbereich des DSG erstreckt sich auf **natürliche** und **juristische Privatpersonen** sowie auf **Bundesorgane**, wobei der Umstand, dass auch juristische Personen, also z. B. eine Aktiengesellschaft, Schutzobjekt sein können, eine Spezialität der Schweiz darstellt. Nicht in den Schutzbereich des DSG fällt folglich das Handeln von **kantonalen** oder **kommunalen Organen**[84]. Dies gilt selbst dann, wenn sie Bundesaufgaben wahrnehmen[85].
So müssen sich z. B. eine Universität oder eine Vormundschaftsbehörde, aber auch mit kantonalen oder kommunalen Aufgaben beliehene Private, bei ihren Datenbearbeitungen nach dem jeweiligen **kantonalen Datenschutzgesetz** richten[86].
Mittlerweile hat sich die Schweiz durch **internationale Verträge** (z. B. das Schengen-Abkommen) zur Einhaltung eines gewissen Mindeststandards verpflichtet, der auch auf Kantonsebene gelten muss. Folglich besteht in den meisten Kantonen ein dem Bundesrecht **vergleichbarer Schutzumfang.**

4. Bundesorgane, Kantone und Private

Das DSG sieht bereichsweise für Privatpersonen andere (weniger strenge) Regeln vor als für **Bundesorgane**. Deswegen muss eine nicht immer einfache doppelte Abgrenzung vorgenommen werden, zwischen Bundesorganen und Privatpersonen, aber auch zwischen Bundesorganen und kantonalen (kommunalen) Organen.
Unter den Begriff der Bundesorgane fallen einerseits die **Behörden und Dienststellen** des Bundes sowie deren **dezentrale Verwaltungseinheiten** (für eine Auflistung siehe Anhang

82 Zivilgesetzbuch.

83 Bundesverfassung.

84 In Art. 2 Abs. 2 DSG sind Ausnahmen vom Anwendungsbereich auch auf Bundesebene vorgesehen. So z. B. bei hängigen Verfahren oder für Beratungen in den eidgenössischen Räten.

85 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz.

86 Kantonale Organe richten sich nur im Fall einer Lückenfüllung der kantonalen Datenschutzgesetzgebung nach dem DSG.

RVOV[87]), aber auch die mit öffentlich-rechtlichen Aufgaben **beliehenen Privaten**. Entscheidend ist, ob auf die der Datenverarbeitung zugrunde liegende Tätigkeit **mehrheitlich öffentliches Recht** oder **Privatrecht** zur Anwendung kommt.

Datenschutzrechtlich ist eine Privatperson auch dann als Organ des Gemeinwesens einzustufen, wenn sie mit **hoheitlichen Befugnissen** ausgestattet wird und somit anderen Privaten Weisungen erteilen kann. Ein hoheitliches Auftreten ist jedoch keine Voraussetzung für die Einstufung als Bundesorgan. Auch das Handeln gestützt auf **Leistungsvereinbarungen** oder eine **mehrheitliche Finanzierung** durch das Gemeinwesen kann dazu führen, dass der öffentlich-rechtliche Charakter überwiegt und eine Qualifizierung als Bundesorgan vorgenommen wird. Diese Abgrenzungsfrage ist wegen der angesprochenen regelungstechnischen Zweiteilung des Datenschutzgesetzes entscheidend. Diese Zweiteilung wird da durchbrochen, wo **Bundesorgane privatrechtlich handeln**. In diesen Fällen erklärt Art. 23 DSG die **Regeln für Privatpersonen** für anwendbar. Damit soll eine Benachteiligung der Bundesorgane verhindert werden. Die Aufsicht bestimmt sich jedoch weiterhin nach den Regeln für Bundesorgane und bleibt somit dem Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten (EDÖB) unterstellt[88].

5. Gelten für das DDZ die DSG-Regeln für Bundesorgane?

Auftraggeber beziehungsweise Initiantin des DDZ ist die SAGW, welche somit den Anknüpfungspunkt darstellt. Von ihrer Rechtsnatur her ist die SAGW ein wissenschaftlicher Verein mit einer öffentlich-rechtlichen Zielsetzung. Sie stützt sich bei ihrem Handeln auf eine Leistungsvereinbarung mit dem SBFI[89], und auch die Finanzierung übernimmt zum grossen Teil der Bund. Als Forschungsinstitution im Sinne des FIFG[90] sind ihre groben Zielsetzungen und Zwecke in Art. 11 FIFG umschrieben. Die SAGW ist unseres Erachtens als Organ des Gemeinwesens (Bundesorgan) einzustufen[91].

Somit muss sich ihr Handeln nach den Datenschutzregeln für Bundesorgane richten. Dazu gehört auch die Initiierung des Projekts für ein DDZ, welche auf eine Leistungsvereinbarung mit dem SBFI zurückgeht. Somit handelt durch das DDZ ein Organ des Gemeinwesens (ein mit öffentlich-rechtlichen Aufgaben beliehener Privater).

87 Regierungs- und Verwaltungsorganisationsverordnung. SR 172.010.1.

88 Maurer-Lambrou/Vogt, Datenschutzgesetz.

89 Das Staatssekretariat für Bildung, Forschung und Innovation. Ein Teil der zentralen Bundesverwaltung im Departement für Wirtschaft, Bildung und Forschung (WBF).

90 Bundesgesetz über die Förderung der Forschung und der Innovation.

91 HMU 22, 247 ff. Zuordnung von Verträgen 1057 ff.

6. Handelt das DDZ als Privater?

Das DSG erlaubt Bundesorganen, Daten nach den weniger strengen Regeln für Private zu bearbeiten, wenn sie in einem bestimmten Bereich privatrechtlich und ohne hoheitliche Gewalt handeln. Die daran anschliessende Frage ist, ob das Handeln des DDZ im vorgesehenen Aufgabenbereich als öffentlich-rechtlich klassifiziert werden muss oder ob eine Privilegierung im Sinne von Art. 23 DSG zulässig ist. D.h.: Ist das Verhältnis des DDZ zu seinen Nutzern und Datenlieferanten als privatrechtlich oder als öffentlich-rechtlich einzustufen? Ein öffentlich-rechtliches Handeln wird dann angenommen, wenn sich das Handeln auf eine **öffentlich-rechtliche** Grundlage stützt und ein **Subordinationsverhältnis** vorliegt. Ein privatrechtliches Handeln durch Bundesorgane kann vorliegen: bei der Leistungsverwaltung, der Finanzverwaltung, bei der administrativen Hilfstätigkeit sowie bei einer privatwirtschaftlichen Staatstätigkeit (z. B. ein Restaurant). Die Beziehung des DDZ zu den Datensponsoren ist voraussichtlich nicht hoheitlich und beruht auf einer freiwilligen Vereinbarung. Auch die Benutzer stehen nicht in einem Subordinationsverhältnis zum DDZ, jedoch bezweckt das Handeln des DDZ die Umsetzung einer **öffentlich-rechtlichen Aufgabe** (der Leistungsauftrag des SBFI), welche sich wiederum auf eine Grundlage in einem öffentlich-rechtlichen Gesetz (FIFG) stützt. Es sprechen also sowohl Argumente für ein privatrechtliches als auch für ein öffentlich-rechtliches Handeln. Vorerst wird hier davon ausgegangen, dass der Schwerpunkt auf dem öffentlichen Recht liegt.

Eine präzisere Beantwortung der Frage wird vorgenommen, wenn die genaue Mechanik der Vorgänge feststeht.

7. Gesetzliche Grundlage für die Datenbearbeitung des DDZ

In Art. 17 DSG konkretisiert das Datenschutzgesetz den allgemeinen Grundsatz von Art. 5 BV, wonach sich jedes staatliche Handeln auf eine gesetzliche Grundlage stützen muss. Es verlangt für die Bearbeitung von ordentlichen, d.h. normalen Personendaten eine gesetzliche Grundlage und für das Bearbeiten von besonders schützenswerten Personendaten oder Persönlichkeitsprofilen sogar eine ausdrückliche formell gesetzliche Grundlage. Eine **formell gesetzliche Grundlage** stellt z. B. ein Bundesgesetz wie das FIFG dar.

Das Handeln der Akademien und somit auch der SAGW ist in Art. 11 FIFG geregelt. Darin wird festgehalten, dass die Akademien berechtigt sind, **Datensammlungen und ähnliche Forschungsinfrastrukturen** zu unterstützen (Art. 11 Abs. 6 FIFG), und, sofern eine Leistungsvereinbarung mit dem SBFI sie dazu ermächtigt, auch zu betreiben (Art. 11 Abs. 7 FIFG).

Personendaten

8. Datenschutzrechtlich relevante Daten

Nicht alle Informationen oder Angaben müssen nach den Vorgaben des Datenschutzgesetzes behandelt werden. Dem DSG sind nur Angaben unterstellt, die als Personendaten zu klassifizieren sind. Aus diesem Grund ist der Begriff zentral für das Datenschutzrecht. Nach Rosenthal setzt sich ein Personendatum aus drei Elementen zusammen: Es muss eine **Information** beinhalten (1), die sich auf eine **real existierende** Person bezieht (2), welche **bestimmt oder bestimmbar** ist (3)[92].13 Die Bestimmbarkeit hat ein relatives Element. So kann ein Foto von einer Party ein Personendatum darstellen für Person X., welche mind. eine der abgebildeten Personen kennt, für Person Y., die keinen Bezug zu den Personen hat, stellt es kein Personendatum dar, sofern nicht damit gerechnet werden muss, dass Person Y. ein Interesse daran hat und die Möglichkeiten besitzt, eine der Personen zu bestimmen.

Ein Personendatum kann auch eine Sachinformation sein, wenn damit **indirekt eine Aussage** über eine Person gemacht werden kann. So ist z. B. der Wert eines Grundstücks erst einmal eine Sachinformation, wird aber zum Personendatum, wenn damit gerechnet werden muss, dass der Inhaber des Grundstücks für den Empfänger dieser Information bestimmbar ist. Welche Fähigkeiten eines Adressaten für die Abschätzung der Bestimmbarkeit zugrunde gelegt werden müssen, ist relativ und hängt von der Art der Daten ab. Je sensibler Daten sind, desto besser müssen sie geschützt werden.

9. Besonders schützenswerte Personendaten

Abhängig vom Lebensbereich, auf den sich Informationen beziehen, liegen **ordentliche Personendaten** oder **besonders schützenswerte Personendaten** vor. Besonders schützenswert sind Daten, die sich auf Vorstrafen, die Intimsphäre, die Weltanschauung, Religion, Gesundheit, gewerkschaftliche Tätigkeiten, Rassenzugehörigkeit oder Sozialhilfe beziehen. Ebenfalls als besonders schützenswerte Personendaten gelten **Persönlichkeitsprofile.** Ein Persönlichkeitsprofil liegt überall dort vor, wo genügend Informationen über eine Person zusammenkommen, um eine Beurteilung **wesentlicher Aspekte der Persönlichkeit** vorzunehmen (z. B. im Rahmen der Cumulus- oder Supercard-Programme). Diese Daten sind potenziell persönlichkeitsverletzend und deshalb besonders schutzwürdig, weshalb der Gesetzgeber an diversen Stellen strengere Regeln für diese Datenkategorie vorgesehen hat.
Die Kategorie der besonders schützenswerten Personendaten ist nicht immer auf Anhieb erkennbar und kann leicht übersehen werden. Z. B. erfüllt bereits ein Foto das Kriterium, wenn darauf der Gesundheitszustand oder die Rassenzugehörigkeit einer Person erkennbar sind.

10. Informationen über eine verstorbene Person

Der **Schutz der Persönlichkeit,** und somit auch der Schutz der mit ihr verbundenen Daten, **endet mit dem Tod.** Sofern aber Verwandte oder eng verbundene Personen des Verstorbenen davon betroffen sind, z. B. bei Informationen über eine vererbbare Krankheit oder sonstigen Daten, die einen direkten Bezug zu den Überlebenden aufweisen, können diese sich auf ihren eigenen Persönlichkeitsschutz berufen. Dazu gehört in gewissem Masse auch das Ansehen eines Angehörigen.

11. Die Anonymisierung von Daten

Eine **Anonymisierung** ist dann erfolgreich abgeschlossen, wenn eine Person durch niemanden mehr bestimmbar ist[93]. Werden also irreversibel anonymisierte Daten durch das DDZ bearbeitet, so stellt dies (nach strittiger Auffassung) keine Bearbeitung von Personendaten mehr dar und ist unbedenklich. Es kann jedoch aus diversen Gründen notwendig sein, den Personenbezug wieder herzustellen. In diesen Fällen kann eine **Pseudonymisierung** (auch Teilanonymisierung genannt) vorgenommen werden. Nach einer Pseudonymisierung ist der Bezug zur Person nur noch durch einen «Schlüssel» möglich. Die Daten stellen also in der Folge nur noch für den **Inhaber des Schlüssels** Personendaten dar, nicht aber für alle anderen (strittig).
Allgemein anerkannt ist, dass Personendaten, die für die Forschung verwendet werden sollen, so schnell wie möglich zu anonymisieren sind. Das Gesetz nennt keinen konkreten Zeitraum, sondern verlangt lediglich eine Anonymisierung, sobald es der **Bearbeitungszweck zulässt** (Art. 22 DSG). In der Lehre spricht man von max. 6 Monaten, die verstreichen dürfen, bis zumindest eine Teilanonymisierung vorliegen muss. Nach spätestens einem Jahr sollten die Daten vollständig anonymisiert sein[94].
Auch bei anonymisierten Daten ist zu beachten, dass durch die **Vernetzung** mit anderen Informationen ein «Wiederaufleben» als Personendaten möglich ist, wenn durch die Vernetzung Personen wieder bestimmbar werden. Es ist also zu vermerken, dass der Datenschutz und die Wissenschaft teilweise unterschiedliche Zielsetzungen verfolgen. Ein Wissensgewinn aus Daten ist allgemein erwünscht, doch sobald mit statistischen Methoden eine immer engere Eingrenzung des Personenkreises möglich wird, werden ab einem gewissen Punkt die Interessenbereiche des Datenschutzes berührt.

92 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz.

93 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz, S. 35 ff.

94 Maurer-Lambrou/Vogt, Datenschutzgesetz.

12. Der Begriff der Datensammlung

Von einer Datensammlung wird dann gesprochen, wenn Personendaten von mehreren Personen in einer Art und Weise vorliegen, die sie **individuell erschliessbar** machen. Diese individuelle Erschliessbarkeit muss sich aus der Struktur der Datenbank ergeben und darf nicht auf das Wissen einer Person über den Datensatz zurückgehen. Es genügt jedoch bereits eine Freitextsuche, wenn damit kein grosser Aufwand für das Auffinden der Person verbunden ist. Die einzelnen Datensätze müssen zudem einen gewissen **thematischen Zusammenhang** aufweisen. Eine Datenbank beinhaltet also in den meisten Fällen eine Datensammlung. Die Begriffe sind jedoch nicht deckungsgleich, da eine Datensammlung immer auch Elemente enthalten wird, die sich nicht auf konkrete Personen beziehen, und somit nicht in das **theoretische Konstrukt** «Datensammlung» passen[95]. Da der Begriff der Datensammlung relativer Natur ist, kann dasselbe Personendatum gleichzeitig Teil von mehreren Datensammlungen sein.

Bearbeitungsregeln

13. Die Persönlichkeitsverletzung

Wenn eine Abklärung ergeben hat, dass **Personendaten** vorliegen, müssen gewisse Regeln bei der Bearbeitung beachtet werden, wobei im rechtlichen Sinne unter dem Begriff **Bearbeiten** jeder Umgang mit Personendaten verstanden wird. Art. 3 lit. e DSG nennt dafür nicht abschliessend: das Beschaffen, Aufbewahren, Verwenden, Umarbeiten, Bekanntgeben, Archivieren oder Vernichten.
Bei all diesen Bearbeitungsschritten müssen die allgemeinen Bearbeitungsgrundsätze des DSG beachtet werden. Wenn eine der Bearbeitungsregeln verletzt wird, liegt eine Persönlichkeitsverletzung vor. In der Folge müsste dann geprüft werden, ob die Verletzung aus einem der in Art. 13 DSG genannten Gründe gerechtfertigt war.
Keine Persönlichkeitsverletzung liegt vor **wenn**:

1) Jeder Bearbeitungsschritt **rechtmässig** ist, womit gemeint ist, dass durch die Bearbeitung keine Gesetze verletzt wurden.
2) Die Bearbeitung nach dem Grundsatz von **Treu und Glauben** erfolgte.
3) Die Bearbeitung der Personendaten für den beabsichtigten Zweck geeignet und erforderlich war und für den Betroffenen keine unzumutbaren Folgen hatte (**Verhältnismässigkeitsprinzip**).
4) Die Daten **transparent** erhoben wurden, d.h. der Betroffene sowohl das Erheben an sich als auch den Zweck des Erhebens erkennen konnte. Sofern der Zweck nicht aus den Umständen ersichtlich war, muss die Person aufgeklärt und darauf hingewiesen worden sein. Eine erteilte Einwilligung kann jederzeit widerrufen werden. Die Beweislast für eine gültige Einwilligung liegt beim Inhaber der Datensammlung.
5) Die Personendaten nur zu den **Zwecken** bearbeitet wurden, die aus den Umständen ersichtlich waren, die angegeben wurden oder die gesetzlich vorgesehen sind.
6) Die Daten für den beabsichtigten Zweck **richtig** waren und durch angemessene technische und organisatorische Massnahmen gesichert wurden (Art. 7 DSG **Datensicherheit**).
7) Vor dem systematischen Erheben von Personendaten durch Bundesorgane die Probanden ausreichend informiert wurden, so dass eine gültige Einwilligung möglich war. Insbesondere muss über den Inhaber der Datensammlung, den Zweck der Bearbeitung sowie die Kategorien der Datenempfänger und das Auskunftsrecht nach Art. 8 DSG[96] aufgeklärt werden. Diese **Informationspflicht** besteht selbst dann, wenn die Daten über Dritte beschafft werden, wie dies beim DDZ vermutlich der Fall sein würde. Folglich ist wichtig, dass im Zeitpunkt der ursprünglichen Erhebung eine weitreichende, gültige Einwilligung eingeholt wurde. Falls dies nicht geschehen ist, kann bei einer Drittbeschaffung ausnahmsweise von einer Information abgesehen werden, nämlich wenn das Gesetz dies vorsieht oder der Aufwand dafür unverhältnismässig wäre.

14. Rechtfertigung einer Persönlichkeitsverletzung

Sofern einer der oben genannten Bearbeitungsgrundsätze verletzt wurde, liegt eine **Persönlichkeitsverletzung** vor, und es muss geprüft werden, ob sie gerechtfertigt war[97].
Die wichtigste «Heilungsmethode» für eine Persönlichkeitsverletzung ist die gültige Einwilligung der betroffenen Personen, d.h. die **freiwillige Einwilligung** nach einer **angemessenen Aufklärung** über den Zweck der Bearbeitung. Bei besonders schützenswerten Personendaten muss die Einwilligung ausdrücklich erfolgen. Für Bundesorgane ist die Berufung auf eine Einwilligung nur ausnahmsweise möglich.
Wenn keine Einwilligung der betroffenen Person vorliegt, muss die Verletzung durch ein **Gesetz** vorgesehen sein oder ein **überwiegendes öffentliches und privates Interesse** nachgewiesen werden. Wobei überwiegende private Interessen zurückhaltender angenommen werden als überwiegende öffentliche Interessen. Genannt sei in diesem Zusammenhang das Interesse an Personen des öffentlichen Lebens. Dieser Rechtfertigungsgrund soll insbesondere die Arbeit von Historikern und Journalisten verein-

95 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz.

96 Siehe Art. 18a DSG.

97 Auf die Rechtfertigungsgründe von Art. 13 DSG können sich grundsätzlich nur Private berufen, da sich das Verhalten von Bundesorganen immer auf eine gesetzliche Grundlage stützen können muss.

fachen. Dabei wird unterschieden zwischen **absoluten Personen** des öffentlichen Lebens (bekannte Schauspieler, Sportler, Politiker usw.) und **relativen Personen** des öffentlichen Lebens, bei denen ein Eingriff in die Persönlichkeit nur im Zusammenhang mit einem bestimmten Ereignis gerechtfertigt ist: z. B. ein Straftäter in Zusammenhang mit seiner Tat oder ein Amtsträger in Zusammenhang mit einem Missbrauch seiner Stellung. Für Bundesorgane ist jedoch höchstens eine analoge Anwendung dieser Rechtfertigungsgründe denkbar[98].

15. Die Folgen einer nicht gerechtfertigten Persönlichkeitsverletzung

Falls keine gültige Einwilligung in eine Persönlichkeitsverletzung vorliegt und auch kein anderer Rechtfertigungsgrund gefunden werden kann, liegt eine **widerrechtliche Persönlichkeitsverletzung** und damit eine unerlaubte Handlung vor. In der Folge kann vor Gericht die Feststellung, Beseitigung und Richtigstellung der Verletzung gefordert werden. Auch ein Anspruch auf Schadensersatz und Genugtuung kann sich daraus ergeben. Strafrechtliche Sanktionen sind nur in seltenen Fällen vorgesehen, so z. B. bei einer fehlenden Aufklärung beim Beschaffen von besonders schützenswerten Personendaten oder Persönlichkeitsprofilen gemäss Art. 14 DSG oder bei einem Geheimnisverrat (Bankgeheimnis, ärztliche Schweigepflicht).

Auftragsbearbeitung

16. Bearbeiten von Daten durch Dritte

In der Praxis werden Personendaten oftmals durch Dritte bearbeitet: z. B. dort, wo ein Unternehmen gewisse Aufgabenbereiche ausser Haus gibt (*«Outsourcing»*) oder wenn Dienstleistungen in Anspruch genommen werden, die zu ihrer Erfüllung eines Transfers bestimmter Informationen bedürfen (z. B. Anwalt, Internetprovider, *Cloud Computing*). Die Voraussetzungen für eine dahingehende Erlaubnis werden in Art. 10a DSG festgelegt. Wer alles genau als Dritter im Sinne von Art. 10a DSG zu gelten hat, ist umstritten. Teilweise wird darunter jede vom Auftraggeber unterschiedene natürliche oder juristische Person verstanden[99]. Andere sind der Ansicht, dass der Struktur des DSG besser entsprochen wird, wenn Art. 10a DSG auch auf gewisse interne Vorgänge anwendbar bleibt. Unter einer Auftragsbearbeitung soll nach dieser Ansicht jede Datenbearbeitung verstanden werden, die von *einer Person* an eine *andere* abgegeben wird (dies kann auch innerhalb des gleichen Unternehmens geschehen und schliesst sogar die Beauftragung des eigenen Arbeitnehmers mit ein)[100].

17. Die Voraussetzungen für eine Datenbearbeitung durch Dritte

Die Beauftragung eines Dritten ist dann unbedenklich, wenn die Voraussetzungen von Art. 10a DSG beachtet werden. Dazu gehört die **Verpflichtung des Beauftragten,** nur Handlungen vorzunehmen, zu denen auch der Auftraggeber berechtigt wäre. Weiter dürfen der Datenübertragung keine **Geheimhaltungspflichten** entgegenstehen, und die betroffenen Personen dürfen **nicht schlechter gestellt** sein, als wenn der Auftraggeber die Datenbearbeitung selber vornehmen würde. Die Beauftragung darf auch keinen Verstoss gegen die allgemeinen **Bearbeitungsgrundsätze** (Art. 4 DSG) darstellen.

Es liegt jedoch auch nach einer korrekten Beauftragung nach Art. 10a DSG in der **Verantwortung des Auftraggebers** (in diesem Fall also der SAGW), für eine rechtmässige Bearbeitung durch den Beauftragten zu sorgen[101]. In vielen Fällen ist es zu empfehlen, mittels Vertrag oder Weisung sicherzustellen, dass die Bearbeitung sachgemäss erfolgt. Im vorliegenden Fall ist eine solche Vereinbarung nicht notwendig, da das DHLab als Teil der Universität Basel dem kantonalen Datenschutzgesetz untersteht, was einen vergleichbaren Schutz garantiert.

Es bleibt jedoch die Pflicht des Auftraggebers, die Einhaltung des Datenschutzes zu überwachen, wobei aber eine allzu grosse Einflussnahme nicht verlangt werden kann und auch unwirtschaftlich wäre. Der Auftraggeber muss jedoch die Kontrolle über den Prozess behalten (Instruktionsrecht), so dass im Notfall eingegriffen werden könnte.

18. Die Folgen einer Einordnung als Auftragsbearbeiter

Sind die Voraussetzungen für eine gültige Beauftragung eines Dritten eingehalten worden, kommt es zu einer sogenannten **Bekanntgabeprivilegierung** zwischen dem Auftraggeber und dem Beauftragten. D.h., es dürfen untereinander Personendaten (auch besonders schützenswerte Daten oder Persönlichkeitsprofile) bekanntgegeben und ausgetauscht werden. Zudem kann sich der Beauftragte auf die Rechtfertigungsgründe des Auftraggebers berufen (Art. 10a Abs. 3 DSG).

98 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz.

99 Maurer-Lambrou/Vogt, Datenschutzgesetz, S. 206.

100 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz.

101 Analog zu Art. 55 OR, *cura in eligendo, instruendo und custodiendo*. Dazu gehört es auch, sicherzustellen, dass der Dritte die notwendigen Voraussetzungen für eine Übertragung erfüllt und die Datensicherheit gewährleisten kann.

19. Der Inhaber der Datensammlung?

Der Auftraggeber darf aber nicht mit dem **Inhaber der Datensammlung** verwechselt werden. Sie können zwar in derselben Person zusammenfallen, was aber nicht zwingend der Fall sein muss. Es ist auch möglich, dass der Auftragsbearbeiter selber Inhaber der Datensammlung ist (z. B. ein Privatdetektiv oder eine Immobilienverwaltung), oder dass Auftraggeber und Beauftragter gemeinsam Inhaber der Datensammlung sind [102]. Der Inhaber einer Datensammlung ist die für die Datensammlung verantwortliche Person, wobei entscheidend ist, wer die **tatsächliche Kontrolle über die Daten** hat und nicht wer auf dem Papier (z. B. in einem Vertrag der Parteien) als Inhaber benannt wird.

Das Bearbeiten von Daten durch das DHLab stellt also eine Auftragsbearbeitung dar. Die SAGW ist in diesem Verhältnis der Auftraggeber. Das DHLab könnte jedoch datenschutzrechtlich trotzdem als Inhaber der Datensammlung eingestuft werden, wenn es selbstständig über den Inhalt der Datensammlung entscheiden kann. In der Regel wird jedoch bei Bundesorganen das **verantwortliche Organ mit dem Inhaber der Datensammlung identisch** sein [103]. So oder so bleibt aber ein Bundesorgan für eine in Auftrag gegebene Bearbeitung verantwortlich (Art. 16 DSG), auch wenn es keine tatsächliche Kontrolle mehr über die Datensammlung haben sollte.

Mit der Einstufung als Inhaber der Datensammlung werden im Datenschutzrecht gewisse Aufgaben verbunden. So liegt es z. B. an ihm, den Informations- und Auskunftspflichten nachzukommen.

Registrierung

20. Die Registrierungspflicht für Datensammlungen

Nach Art. 11a Abs. 2 DSG müssen Bundesorgane all ihre Datensammlungen beim **EDÖB** registrieren. Das DSG erlaubt jedoch dem Bundesrat mit der Verordnung, gewisse Datenbearbeitungen durch Privatpersonen oder Bundesorgane von der Anmeldepflicht auszunehmen. Von dieser Kompetenz hat der Bundesrat Gebrauch gemacht und in Art. 4 VDSG alle Datensammlungen, die zu **nicht personenbezogenen Zwecken** betrieben werden, von der Anmeldepflicht ausgenommen. Damit sollte auch das DDZ von einer Registrierungspflicht befreit sein.

Bekanntgabe und Forschung

21. Privilegierung der nicht personenbezogenen Bearbeitung

Die Datenbekanntgabe ist von ihrer Natur her ein besonders heikler Schritt, da der bisherige Dateninhaber ab diesem Zeitpunkt die faktische Kontrolle über den Datensatz verliert. Deswegen ist die Datenbekanntgabe auch teilweise zusätzlichen Regeln unterworfen. So braucht z. B. ein Bundesorgan, welches gesetzlich ermächtigt ist, Personendaten zu bearbeiten, nochmals eine zusätzliche Rechtsgrundlage, um Daten bekanntzugeben (Art. 19 Abs. 1 DSG). Eine Ausnahme von diesem zusätzlichen Erfordernis besteht dann, wenn die Personendaten lediglich zu **nicht personenbezogenen Zwecken** verwendet werden (Art. 22 Abs. 2 lit. c DSG). D.h, die gesetzliche **Grundlage** der SAGW **für die Bearbeitung** (gestützt auf Art. 11 FIFG) **ermöglicht auch die Bekanntgabe** der Daten, da für nicht personenbezogene Zwecke keine separate Ermächtigung notwendig ist.

Die Bearbeitung zu nicht personenbezogenen Zwecken wie Forschung, Planung oder Statistik hat auch noch andere Erleichterungen zur Folge. So darf entgegen der Regel von Art. 4 Abs. 3 DSG eine Zweckänderung vorgenommen werden (nur in Richtung nicht personenbezogene Zwecke), und es ist auch für das **Bearbeiten von besonders schützenswerten Personendaten und Persönlichkeitsprofilen** keine ausdrückliche Ermächtigung in einer formell gesetzlichen Grundlage notwendig, wie dies sonst der Fall wäre [104].

Diese Privilegierungen der Forschung, Planung und Statistik nach Art. 22 DSG sind den folgenden Voraussetzungen unterstellt:

- Die Daten müssen anonymisiert werden, sobald es der Bearbeitungszweck erlaubt.
- Der Empfänger der Daten darf sie ebenfalls nur zu nicht personenbezogenen Zwecken bearbeiten.
- Eine Weitergabe darf nur mit Zustimmung des Bundesorgans stattfinden.
- Durch die Veröffentlichung der Ergebnisse dürfen die Personen nicht bestimmbar werden.
- Fraglich bleibt, ob die geisteswissenschaftliche Forschung im konkreten Fall die Daten «nicht personenbezogen» im Sinne von Art. 22 DSG verwendet. Ein Historiker z. B., der über eine bestimmte Person nachforscht, benutzt die Daten sehr wohl personenbezogen und wäre damit nicht berechtigt [105], womit insbesondere im Bereich der Geisteswissenschaften Forschungen, die auf Personendaten angewiesen sind, ein Problem bekommen könnten. Zwar kann sich der einzelne Historiker auf die Rechtfertigungsgründe in Art. 13 DSG berufen, doch ein Bundesorgan darf dies üblicherweise

102 In der EU ist der Auftraggeber immer auch der Inhaber der Datensammlung (in beiden Fällen wird vom *«Controller»* gesprochen). Da in der Schweiz die Sorgfaltspflichten des Inhabers der Datensammlung auch den Auftragsbearbeiter treffen können, ist, zumindest theoretisch, die Auftragsbearbeitung in der Schweiz riskanter als in der EU.

103 Astrid Epiney/Patrizia Zbinden/Tamara Civitella, Epiney Zbinden – Datenschutzrecht in der Schweiz, S. 38.

104 Art. 17 Abs. 2 DSG.

105 Rosenthal/Jöhri, Handkommentar zum Datenschutzgesetz.

nicht und hätte somit dem Historiker die Daten gar nicht erst zugänglich machen dürfen.

22. Die Informationspflicht bei der Beschaffung

Die Informationspflicht für Bundesorgane wurde bei der letzten Revision des DSG verschärft. So müssen nach Art. 18a DSG sowohl bei der eigenen Beschaffung, als auch bei der **Beschaffung bei Dritten** (was beim DDZ die Regel sein wird), die Betroffenen informiert werden. Diese Informationspflicht gilt sowohl für ordentliche als auch für besonders schützenswerte Personendaten. Eine Privilegierung für die Forschung ist nicht vorgesehen. Einschränkend wirkt, dass mit dem Begriff «Beschaffung» nicht jedes Beschaffen von Personendaten, sondern nur das **systematische Beschaffen** gemeint sein kann. Trotzdem wird das DDZ davon betroffen sein. Die Informationspflicht entfällt, wenn **bereits informiert** wurde, ein Gesetz die Speicherung der Bekanntgabe ausdrücklich vorsieht oder eine Information der Betroffenen nur mit unverhältnismässigem Aufwand möglich ist.
Sofern eine Informationspflicht bestehen bleibt, muss der betroffenen Person Folgendes mitgeteilt werden:

- der Inhaber der Datensammlung
- der Zweck der Bearbeitung
- die Kategorien der Datenempfänger
- ihr Auskunftsrecht nach Art. 8 DSG

Folglich ist es wichtig, im Vertrag mit dem **Datenlieferanten** die **Zusicherung** einzuholen, dass die im Datensatz vorhandenen Personen über all diese Punkte bereits informiert wurden und eine Bearbeitung, wie sie das DDZ beabsichtigt, berücksichtigt wurde. Falls dies nicht der Fall ist, müsste spätestens bei der Speicherung der Daten über die fehlenden Punkte **nachinformiert** werden.

23. Die Auskunftspflicht

Der **Inhaber der Datensammlung** muss jeder anfragenden Person bekanntgeben, ob Daten über sie bearbeitet werden. Damit eine solche Anfrage überhaupt möglich ist, muss eine **Kontaktperson** dafür bezeichnet worden sein. Falls dann Daten über die anfragende Person vorliegen, ist anzugeben, welche dies sind, zu welchem Zweck sie bearbeitet werden, auf welche gesetzliche Grundlage sich die Bearbeitung stützt und woher die Daten kommen. Eine solche Auskunft ist **üblicherweise kostenlos**, es kann jedoch eine Beteiligung im Umfang von max. Fr. 300 erhoben werden, wenn die ersuchende Person innerhalb der letzten zwölf Monate bereits einmal ein Gesuch gestellt hat und keine guten Gründe für das erneute Gesuch vorliegen. Ausserdem ist eine Kostenbeteiligung denkbar, wenn die Recherche besonders aufwändig ist, wovon aber nur ausgegangen werden kann, wenn der Datensatz nicht nach Einzelpersonen erschlossen ist. Auf ein Auskunftsersuchen muss innert nützlicher Frist (üblicherweise 30 Tage) geantwortet werden. Wenn der Auskunftspflicht nicht nachgekommen wird, kann dies mit einer Busse bestraft werden.
Die Auskunftspflicht gilt sowohl für Private als auch für Bundesorgane und stellt für Betroffene die wichtigste Handhabe dar, um die Kontrolle über ihre Daten zu behalten.

Fazit

Das Ziel des DDZ, Datensätze als «*Open Data*» für jedermann zugänglich zu machen, ist nur möglich, wenn sie keine Personendaten mehr enthalten. Die Personendaten müssten also vorher anonymisiert werden. Wenn der Bearbeitungszweck eine Anonymisierung nicht erlaubt, ist ein Zugang nur nach einer vertraglichen Vereinbarung möglich[106]. In diesem Zusammenhang ist zu beachten, dass sich nicht nur das DDZ, sondern auch der ursprüngliche Datenlieferant an das für ihn geltende Datenschutzrecht halten muss. Eine Universität ist z. B. an das geltende kantonale Datenschutzgesetz sowie ihre internen Richtlinien gebunden. Nach dem Datenschutzgesetz von Basel dürfte z. B. eine Universität einer Privatperson nur dann Zugriff auf Personendaten enthaltende Forschungsdaten geben, wenn diese sich vertraglich zur Einhaltung der in Art. 22 Abs. 4 IDG BS[107] genannten Bedingungen verpflichtet[108].

Das DDZ wird folglich Verträge aufsetzen müssen[109], die sich an den Anforderungen der Datenschutzgesetze von Bund und Kantonen orientieren. Mittels vertraglicher Absicherung können danach die Datensätze einzeln zugänglich gemacht werden, sofern die andere Partei die Anforderungen dazu erfüllt.

Fremde Inhalte

24. Allgemein

Unter fremden Inhalten wird in diesem Zusammenhang die Gesamtheit der Leistungen verstanden, die durch gesetzliche Bestimmungen einen gewissen Schutz geniessen. Dazu gehören der Markenschutz (MSchG), der De-

106 Notwendige Klauseln nach Art. 22 DSG: Verpflichtung zur Beibehaltung des Schutzniveaus, keine Weitergabe ohne Einverständnis, nur nicht personenbezogene Bearbeitung, veröffentlichte Ergebnisse dürfen keine Rückschlüsse zulassen.

107 Informations- und Datenschutzgesetz Basel. SG 153.260.

108 Die Voraussetzungen wären: Die Daten sind, sobald es der Bearbeitungszweck erlaubt, zu anonymisieren oder zu pseudonymisieren. Die Auswertungen der Daten dürfen keine Rückschlüsse auf die Personen mehr zulassen. Die Personendaten dürfen für keine anderen Zwecke als den angegebenen bearbeitet werden. Die Personendaten dürfen nicht Dritten bekanntgegeben werden. Die Informationssicherheit im Sinne von Art. 8 IDG ist sichergestellt.

109 Sowohl im Verhältnis DDZ & Datenlieferanten als auch DDZ & Datenbezüger muss darauf geachtet werden, dass den Voraussetzungen für die Weitergabe der jeweiligen Institution bzw. dem Vertragsinhalt mit den jeweiligen Probanden entsprochen wird.

signschutz (DesG), der Patentschutz (PatG), das Wettbewerbsgesetz (UWG) und das Urheberrecht (URG).
Der **Markenschutz** wird in den meisten Fällen unproblematisch bleiben, da sich der Schutzbereich auf eine *kennzeichenmässige Verwendung* beschränkt[110], ebenso das **Designrecht**, welches als gewerbliches Schutzrecht dem Rechteinhaber ermöglicht, eine Verwendung der geschützten Gestaltung für *gewerbliche Zwecke* zu verbieten.
Das **Patentrecht** wiederum hat nicht zum Ziel, die Unterlagen zur Entwicklung einer *technischen Lösung* zu schützen, sondern möchte die Verwendung in einem nicht autorisierten *System oder Produkt* verbieten (wissenschaftliche Unterlagen wie z. B. die technische Zeichnung eines Patentes können jedoch urheberrechtlich geschützt sein).
Das **Lauterkeitsrecht (UWG)**[111], auch Wettbewerbsrecht genannt, wirkt zu allen oben genannten Schutztiteln als eine Art Auffangtatbestand und möchte Verhaltensweisen ausschliessen, die für einen gesunden Wettbewerb innerhalb einer Volkswirtschaft schädlich sind. Da der Schutzgegenstand des Lauterkeitsrechts also der Wettbewerb an sich ist, wird es ebenfalls nur in Ausnahmefällen für ein Projekt wie das DDZ relevant werden.
Von besonderem Interesse bleibt also das **Urheberrecht.**

Das Urheberrecht

Das Urheberrecht entsteht automatisch beim Urheber mit der Schöpfung des Werks. Um in den Genuss der Schutzwirkung zu kommen, ist keine Anmeldung oder Registereintragung notwendig.
Schutzgegenstand des Urheberrechts sind **Werke und Leistungen** der Literatur und Kunst. Wobei beide Begriffe sehr weit zu verstehen sind. Es kann sich um einen Brief, ein Foto, eine wissenschaftliche Zeichnung oder auch eine Webseite handeln. Gemeinsames Merkmal ist ihr Ursprung in einer **geistigen Schöpfung** (womit z. B. Zeichnungen von Tieren ausgeschlossen werden) mit **individuellem Charakter.** Entscheidend für die Abgrenzung eines Werks von allen sonstigen menschlichen Leistungen, die keinen urheberrechtlichen Schutz geniessen, ist vor allem das Merkmal des individuellen Charakters. Gemeint ist damit eine Art Originalität im gegebenen Rahmen[112]. Die Anwendung dieser an sich neutral formulierten Merkmale auf die verschiedenen Werkkategorien zeigt jedoch in der Praxis gewisse Unterschiede. So wird z. B. bei Werken der angewandten Kunst (z. B. ein Stuhl, eine Gabel oder eine Uhr) ein urheberrechtlicher Schutz eher abgelehnt (teilweise besteht dann immer noch ein Designschutz, sofern er denn beantragt wurde)[113], als es bei einer Komposition oder einem Gemälde der Fall wäre. Auch in der Fotografie ist eine gewisse Strenge bei der Beurteilung des Werkcharakters spürbar. Von vornherein nicht schützbar sind nach herrschender Lehre **blosse Ideen** oder **Informationen.** Geschützt ist aber ihre konkrete Ausgestaltung oder Umsetzung.
Im Bereich der Behörden und des Staates sind diverse Werke trotz potenziellem Werkcharakter vom Urheberrechtsschutz ausgenommen. Dazu gehören Gesetze und amtliche Erlasse, Zahlungsmittel, Entscheidungen, Protokolle, Berichte von Behörden und öffentlichen Verwaltungen, sowie Patentschriften und veröffentlichte Patentgesuche[114].

25. Urheberrechtlicher Schutz einer Datenbank

Ebenfalls als Werke gelten sogenannte **Sammelwerke,** deren Eigenart sich aus der **kreativen Auswahl und Anordnung** verschiedener Elemente ergibt. Es ist theoretisch denkbar, dass eine Datenbank als Sammelwerk eingestuft wird, obwohl das Bundesgericht in dieser Hinsicht bisher eher zurückhaltend war[115].
Für den Fall, dass eine Datenbank keinen urheberrechtlichen Schutz beanspruchen kann, ist in der Schweiz höchstens noch das **Lauterkeitsrecht** (UWG) als Schutzmöglichkeit denkbar. Unlauter handelt jemand u. a. dann, wenn er *«das marktreife Arbeitsergebnis eines Dritten ohne angemessenen eigenen Aufwand mit technischen Reproduktionsverfahren übernimmt und verwertet»*. Es müsste dann jedes Merkmal dieser Formel erfüllt werden, damit eine Verletzung des UWG vorliegt.

EXKURS: Anders ist die Situation in der EU, wo ein **Datenbankrecht *sui generis*** besteht[116]. Man möchte Investitionen in Datenbanken fördern und gesteht ihnen deswegen einen gewissen Schutz zu, sofern beträchtliche ***Investitionen*** für die ***Übernahme, Aufbereitung*** oder ***Aufbewahrung*** der Daten aufgewendet wurden. Eine wichtige Abgrenzung zu den Datenbanken, die keinen Schutz geniessen, findet beim Begriff der Übernahme statt. So ist begrifflich die Übernahme der Erhebung nachgelagert, was bedeutet, dass die Kosten der Datenerhebung nicht zu den beträchtlichen Investitionen dazu gezählt werden, die für einen Schutz der Datenbank in der EU notwendig wären. Dies ist bedeutsam, weil in diversen Bereichen die höchsten Kosten beim Erheben der Daten selber anfallen.

110 D.h., das Recht an der Marke kann nur verletzt werden, wenn die Marke zur Kennzeichnung von Waren oder Dienstleistungen verwendet wird. Die Arbeit eines Journalisten oder Wissenschaftlers ist davon nicht betroffen.

111 Bundesgesetz gegen den unlauteren Wettbewerb (UWG).

112 Barrelet/Egloff, Das neue Urheberrecht, S. 13 ff.

113 Es besteht auch eine gewisse Gefahr, dass versucht wird, über das Urheberrecht die im Vergleich dazu viel kürzeren Schutzfristen des Designrechts auszuhebeln.

114 Siehe Art. 5 URG.

115 Z. B. BGE 134 III 166 «Fall Documed».

116 Richtlinie 96/9/EG.

Diese Daten wären also nur dann geschützt, wenn zu einem späteren Zeitpunkt beträchtliche Investitionen in die Aufbereitung oder Aufbewahrung notwendig waren.
Ein urheberrechtlicher Schutz für eine Datenbank kommt auch in der EU nur in wenigen Fällen in Frage[117].

26. Urheberrechtlich relevante Handlungen

Der Inhalt des Urheberrechts ist sehr umfangreich und umfasst sowohl diverse Vermögensrechte als auch Persönlichkeitsrechte.

(1) Zu den **Urhebervermögensrechten** gehört u. a. die ausschliessliche Berechtigung, ein Werk zu *vervielfältigen,* zu *verbreiten, wahrnehmbar zu machen* (dazu gehört auch das «*on demand»-Recht*[118]) zu *senden* oder *weiterzusenden*[119]. Diese Rechte können vom Urheber einzeln oder im Bündel, exklusiv oder nicht exklusiv verwertet werden.

(2) Zu den **Urheberpersönlichkeitsrechten**[120] gehört das Recht, selber zu bestimmen, wann und in welcher Form das eigene Werk einem unkontrollierten Personenkreis zugänglich gemacht wird (*Erstveröffentlichungsrecht*). Dies führt z. B. dazu, dass ein Manuskript, das in einem nicht öffentlich zugänglichen Archiv liegt, nicht zitiert werden darf, da es als nicht veröffentlicht gilt[121].

Das Recht auf *Anerkennung der Urheberschaft* gibt dem Urheber das Recht zu bestimmen, unter welcher Bezeichnung das Werk der Öffentlichkeit bekannt gemacht wird. Das Recht ist auch dann verletzt, wenn ein neues Werk sich nur so geringfügig vom Original unterscheidet, dass dessen Schutzbereich noch immer betroffen ist.
Im Grunde ist die Urheberschaft eine Tatsachenfrage und kann somit nicht abgetreten werden. Davon zu unterscheiden ist eine Abmachung, welche den Verzicht auf die Geltendmachung der Urheberschaft zum Inhalt hat *(Ghostwriterabrede).*
Ebenfalls Teil des Urheberpersönlichkeitsrechts sind die praktisch sehr relevanten *Änderungs- und Bearbeitungsrechte*[122]. Das Recht zur Änderung bedeutet die Möglichkeit, nicht schöpferische Veränderungen am betreffenden Werk vornehmen zu können. Das Bearbeitungsrecht meint die Berechtigung zur schöpferischen Umarbeitung des Werks zu einem Werk zweiter Hand (z. B. die Verfilmung eines Buchs) oder auch die Aufnahme in ein Sammelwerk[123].
Der **Werkgenuss** an sich stellt aber keine urheberrechtlich relevante Handlung dar[124].

27. Folgen eines urheberrechtlichen Schutzes

Für die Vornahme einer Handlung, die eines der oben genannten Urheberrechte berührt, ist das Einverständnis des Rechteinhabers notwendig.
Angenommen, dass die übernommenen Forschungsdaten oder ein darin enthaltenes Element **Werkcharakter** hätten, dann müssten die betroffenen Rechte beim Rechteinhaber eingeholt werden.
Angenommen, dass die Forschungsdaten **keinen Werkcharakter** hätten, aber durch das DDZ genügend originell und schöpferisch umgearbeitet würden, so dass bei der Aufbereitung der Daten ein Werk entstehen würde, dann müssten diese Urheberrechte wieder über eine Lizenz (z. B. Creative Commons) abgestossen werden, um eine freie Verwendung für die Benutzer zu ermöglichen.

Verwandte Schutzrechte

28. Sonstige Leistungen, die durch das Urheberrecht geschützt sind

Ebenfalls im Urheberrecht geregelt ist der Schutz für Leistungen, die zwar im Normalfall keine Werke darstellen, aber trotzdem einen vergleichbaren Schutz geniessen (z. B. eine Konzertaufnahme oder Fernsehsendung). Diese sogenannten **Leistungsrechte**, auch verwandte Schutzrechte genannt, sind eine Art kleines Urheberrecht für ausübende Künstler, Hersteller von Ton-/Tonbildträgern[125] und Sendeunternehmen[126]. Der Schutzumfang, d.h. die Handlungen, die nur mit Einverständnis des Rechteinhabers vorgenommen werden dürfen, ist enger als beim Urheberrecht, umfasst aber ebenfalls die Rechte zur *Verbreitung, Vervielfältigung, Wahrnehmbarmachung, Sendung* und *Weitersendung*[127].
Das Leistungsschutzrecht **besteht parallel zum Urheberrecht**, so dass es notwendig sein kann, sich sowohl die Urheber- als auch die Leistungsrechte bei den jeweiligen Rechteinhabern einzuholen.

117 Weitere Informationen siehe: Guibault/Wiebe, Safe to be open.

118 Etwas über ein Intranet oder das Internet zu jeder Zeit und von jedem Ort zugänglich zu machen.

119 Art. 10 URG.

120 Nicht zu verwechseln mit dem allgemeinen Persönlichkeitsrecht, welches mit dem Tod des Urhebers endet.

121 Hilty, Urheberrecht, S. 166.

122 Art. 11 URG.

123 Hilty, Urheberrecht, S. 170.

124 Siehe z. B. BGE 133 III 473 «Pressespiegel».

125 Nach Art. 36 URG z. B. sind Naturgeräusche und Tierstimmen keine Werke, doch der Hersteller der CD (kann auch eine juristische Person sein) erhält trotzdem einen Schutztitel.

126 Art. 37 URG.

127 Art. 33 URG.

Das Zitat

29. Zitate von geschützten Inhalten

Die Zitierfreiheit erfüllt eine wichtige Funktion beim Ausgleich zwischen dem Monopolrecht des Urhebers oder Leistungserbringer und dem Interesse der Allgemeinheit an einer Rezeption des Werks oder der Leistung. Sofern bei einem Zitat auf die **Quelle** verwiesen wird und eine **inhaltliche Auseinandersetzung** mit dem zitierten Werk stattfindet, können Werkteile oder auch ganze Werke (sofern noch durch den Zitatzweck gedeckt) in ein eigenes Werk übernommen werden. Dafür ist keine Erlaubnis notwendig, und es muss auch keine Entschädigung geleistet werden. Der in Frage stehende Verwendungszweck darf aber die wirtschaftliche Verwertung des Werks nicht beeinträchtigen [128].

Umstritten ist, ob das Zitatrecht auch für **Werke der bildenden Kunst** [129], der **Fotografie**, des **Films** oder der **Musik** gilt. Mittlerweile wird dies von einem grossen Teil der Lehre bejaht. Entscheidend ist auch hier, dass eine Auseinandersetzung mit dem Werk selber stattfindet und nicht nur mit der dargestellten Thematik oder Sache. Zur Ausschmückung dürfen Werke der bildenden Kunst nicht verwendet werden. So ist z. B. die Verwendung eines Fotos vom 11. September erlaubt, wenn dieses Foto besprochen wird, nicht jedoch zur Illustration eines Artikels über Terrorismus [130].

30. Das Plagiat und die Verletzung der wissenschaftlichen Ethik

Im Falle eines Zitats, bei dem der Kontext nicht gegeben war oder die Quelle falsch oder gar nicht angegeben wurde, kann ein Verstoss gegen das Urheberrecht vorliegen. Dies ist aber nicht zwingend. So kann z. B. die Schutzfrist des Werks bereits abgelaufen sein oder dem übernommenen Teil der Werkcharakter fehlen. Ein besonderer Fall einer solchen Übernahme fremder Inhalte ist das Plagiat. Das **Plagiat** ist eine Täuschung über den Urheber. Indem Inhalte von einem Dritten übernommen werden, ohne sie als solche zu kennzeichnen, masst man sich selber die Urheberschaft an und verletzt das Recht auf Anerkennung der Urheberschaft des wahren Autors.
Kein Plagiat liegt vor, wenn ein Urheber das eigene Werk als dasjenige eines anderen ausgibt (z. B. indem ein eigenes Bild mit Picasso unterschrieben wird). In diesem Fall liegt ein Verstoss gegen die Persönlichkeitsrechte nach Art. 28 oder 29 ZGB (Recht am eigenen Namen) vor [131].

Im Falle eines *«Ghostwriters»* wurde mit dem wahren Autor eine Abmachung getroffen, auf die Ausübung seines Rechts auf Namensnennung zu verzichten.

Schranken des Urheberrechts

31. Die Schutzdauer eines Werks oder einer Leistung

Der Schutz für **Urheberrechte** endet **70 Jahre** nach dem Tod des Urhebers (bei Computerprogrammen 50 Jahre). Wobei die Frist immer erst ab dem 31. Dezember des Todesjahres zu laufen beginnt. Falls ein gemeinsames Werk vorliegt, also mehrere Personen als Urheber zusammengewirkt haben, so beginnt die Frist erst mit dem Tod des letzten Miturhebers zu laufen [132].
Bei unbekannter Urheberschaft (nicht zu verwechseln mit der Verwendung eines Pseudonyms) erlischt das Recht 70 Jahre nach der Veröffentlichung des Werks [133].
Für **Leistungsschutzrechte** erlischt der Schutz **50 Jahre** nach dem Zeitpunkt der Darbietung, der Veröffentlichung des Ton-/Bildträgers oder der Erstausstrahlung der Sendung. In den Fällen, wo keine Veröffentlichung stattgefunden hat, ist das Datum der Herstellung entscheidend. Je älter ein Werk oder eine Leistung ist, umso grösser ist natürlich die Wahrscheinlichkeit, dass das Urheberrecht abgelaufen ist. Die Schweizerische Nationalbibliothek z. B. schätzt alle Dokumente, die vor mehr als 110 Jahren veröffentlicht wurden, als unbedenklich und frei verwendbar ein [134].
Keine besondere Regelung hat in der Schweiz die Situation von **verwaisten Werken** erfahren. Das Problem wird im Urheberrecht lediglich in Bezug auf Ton-/Tonbildträger (z. B. CDs, DVDs) angesprochen und bei Vorliegen gewisser Voraussetzungen einem Regime der kollektiven Verwertung unterworfen [135]. Eine kollektive Verwertung bedeutet für die Nutzer eine deutliche Vereinfachung des Zugangs.

32. Die Beziehung der Urheberrechte zum Werkexemplar (Erschöpfung)

Wenn ein Urheber ein Werk veräussert, hat er in Bezug auf das betreffende Werkexemplar sein Urheberrecht erschöpft. Erschöpfung bedeutet, dass der Urheber nicht weiter über die Verbreitung des betreffenden Werkexemplars bestimmen kann. Dieses erworbene Exemplar darf weiterverkauft, verliehen, vermietet oder ausgestellt

128 Büren/Marbach/Ducrey, Immaterialgüter- und Wettbewerbsrecht, Rn. 360 ff.

129 Z. B. Gemälde, Zeichnungen oder Grafiken.

130 Müller/Oertli, Urheberrechtsgesetz (URG), S. 340 ff.

131 Müller Oertli, Urheberrechtsgesetz (URG).

132 Art. 29 URG.

133 Art. 31 URG.

134 Schweizerische Nationalbibliothek (NB), Digitalisierungsleitlinie Juli 2014, S. 5.

135 Der Begriff kollektive Verwertung bedeutet, dass die Schutzrechte nicht mehr durch den Rechteinhaber selber ausgeübt werden können, sondern durch eine in der Schweiz anerkannte Verwertungsgesellschaft wahrgenommen werden müssen.

werden. Wobei für die **Vermietung** eine Vergütungspflicht besteht. Der Begriff der Vermietung meint aber nur die entgeltliche Überlassung, was bei geringen Mitgliedschaftsbeiträgen noch nicht angenommen wird, wenn damit lediglich ein Teil der Betriebskosten einer gemeinnützigen Institution gedeckt wird [136]. In diesem Fall wird die Überlassung als einfache Verleihung eingestuft, welche vergütungsfrei möglich ist. Ein Folgerecht (eine Beteiligung am Weiterverkauf von bereits verkauften Exemplaren) für die Urheber oder deren Rechtsnachfolger wurde regelmässig abgelehnt [137].

Von der Erschöpfung nicht betroffen sind die Urheberrechte an sich. Die Erschöpfung bezieht sich auch nicht auf Kopien, die im Rahmen des Eigengebrauchs [138] hergestellt werden.

33. Die gesetzlichen Schranken des Urheberrechts

Urheberrechte sind **Monopolrechte** auf Zeit, welche dem Urheber für sein Werk zugestanden werden. Das Werk selber wird aber oft Teil einer öffentlichen oder privaten Diskussion. Es wird also versucht, eine Balance zwischen den Interessen der Gemeinschaft und dem Recht des Urhebers auf sein geistiges Eigentum zu finden.
Wenn der urheberrechtliche Schutz noch nicht abgelaufen ist und auch kein gültiges Zitat vorliegt, gibt es gewisse Bereiche und Situationen, in denen der Gesetzgeber urheberrechtlich relevante Handlungen auch ohne Einverständnis der Rechteinhaber erlaubt. Diese Ausnahmen lassen sich unter dem Oberbegriff der urheberrechtlichen Schranken zusammenfassen. Es sind jedoch immer einzelne Konstellationen, die mehr oder weniger präzise benannt werden. Eine Generalklausel für die Schutzausnahmen, wie es das amerikanische Recht mit dem ***«fair use»***-Konzept kennt, gibt es in der Schweiz nicht.
Die im vorliegenden Fall wichtigsten Schranken sind in Art. 19 URG zu finden und betreffen die verschiedenen Formen des Eigengebrauchs.
Alle nachfolgend genannten Privilegierungen sind technologieneutral zu verstehen, d.h., es spielt keine Rolle, ob z. B. eine Handlung manuell oder mit technischen Mitteln (z. B. mit Hilfe eines Netzwerks) vorgenommen wird.

34. Der Privatgebrauch

Der **Privatgebrauch** in Art. 19 Abs. 1 lit. a URG erlaubt Privatpersonen ***jede Verwendung eines geschützten Werks,*** solange sie sich nur im persönlichen Bereich zwischen eng verbundenen Personen abspielt [139]. Juristische Personen sind davon ausgeschlossen, weshalb sich das DDZ nicht auf den Privatgebrauch berufen kann. Als Eigengebrauch gilt jedoch auch die Vervielfältigung zu beruflichen Zwecken [140], weshalb der einzelne Wissenschaftler davon profitieren kann (und das DDZ evtl. die Handlung als Drittperson für den Berechtigten vornehmen könnte, siehe Ziff. 38). Für den Privatgebrauch wird **keine Vergütung** geschuldet. Ein Privatgebrauch ist auch an offensichtlich rechtswidrig hergestellten Vorlagen (Raubkopien) möglich, sofern der Zugang selber rechtmässig erfolgt ist (so ist z. B. der Download eines Films von einer Streaming-Seite erlaubt, nicht aber das Kopieren eines gestohlenen Buchs) [141].

35. Der Schulgebrauch

Für den Unterricht in der Klasse ist aufgrund von Art. 19 Abs. 1 lit. b URG ebenfalls ***jede Werkverwendung*** gestattet. Diese auch **Schulgebrauch** genannte Ausnahme begünstigt sowohl die Lehrer als auch die Schüler, solange mit der Werkverwendung der Unterrichtszweck verfolgt wird. Dies gilt auch für Studenten an einer Universität oder Fachhochschule, jedoch nur innerhalb der Vorlesungsgruppe.
Durch die Technologieneutralität der Schranke ist auch das Zugänglichmachen von geschützten Inhalten über das Internet erlaubt, sofern der Zugriff (z. B. über ein Passwort) auf den berechtigten Kreis beschränkt wird.
Der Schulgebrauch ist nicht wie der Privatgebrauch vergütungsfrei, doch die Lizenzgebühren müssen nicht individuell verhandelt werden, sondern können kollektiv über die zuständige Verwertungsgesellschaft als Scharnier zwischen Urheber und Institut auf Grundlage der geltenden Tarife abgerechnet werden.
Im Vergleich zum Privatgebrauch besteht sowohl beim Schulgebrauch als auch beim in Ziff. 37 behandelten Betriebsgebrauch eine Einschränkung durch Art. 19 Abs. 3 URG. Darin wird festgelegt, dass im *Handel erhältliche Werke* nicht vollständig *vervielfältigt* werden dürfen. Für *Werke der bildenden Kunst* und *Partituren* ist sogar ein generelles Vervielfältigungsverbot vorgesehen. Da eine solche Regelung praktisch nicht durchsetzbar war, haben die Verwertungsgesellschaften in den **gemeinsamen Tarifen,** in welchen der Vergütungsansatz für die verschiedenen Werkkategorien und Verwendungsarten festgelegt wird, auch Vergütungen für in Art. 19 URG nicht vorgesehene Nutzungsarten vorgesehen und eingezogen. Damit wurde praktisch der Umfang des Schulgebrauchs auch auf vollständige Werke der bildenden Künste und Partituren erweitert.

136 Cherbuin/Dengg/Regamey, Digitale Bibliotheken und Recht, S. 17.

137 Für den EU-Raum ist ein dementsprechendes Recht mittlerweile verbindlich.

138 Zum Eigengebrauch, siehe Ziff. 34–38.

139 Durch den Verweis in Art. 39 URG gelten die Schranken auch für Schutztitel nach dem Leistungsrecht.

140 David Rüetschi in: Cherbuin/Dengg/Regamey, Digitale Bibliotheken und Recht, S. 18.

141 Müller/Oertli, Urheberrechtsgesetz (URG).

36. Der Betriebsgebrauch

Der **Betriebsgebrauch** nach Art. 19 Abs. 1 lit. c URG ist noch einmal enger ausgestaltet als der Privat- und Schulgebrauch. Die Änderungs-, Bearbeitungs- und Vorführrechte sind nicht mehr enthalten. Erlaubt ist lediglich das ***Vervielfältigen zur internen Dokumentation und Information***. Nicht genannt, aber trotzdem miterfasst, ist nach Ansicht der Lehre und des Bundesgerichts[142] das ***Verbreiten*** und ***Zugänglichmachen*** von geschützten Inhalten, solange sie auf den betriebsinternen Bereich beschränkt bleiben. Begünstigt sind alle privaten Unternehmen, öffentlichen Verwaltungen, Institute oder ähnlichen Einrichtungen. Somit wäre auch das DDZ durch diese Regelung begünstigt, sofern die geschützten Inhalte lediglich der internen Dokumentation und Information dienen würden und keine Änderungen oder Bearbeitungen derselben notwendig wären.

37. Die Möglichkeit, den Eigengebrauch durch Dritte wahrnehmen zu lassen

Die Werknutzung für den Eigengebrauch wird in der Schweiz noch einmal deutlich attraktiver durch den Umstand, dass es allen drei privilegierten Nutzergruppen erlaubt ist, die Vervielfältigung, zu der sie berechtigt sind, durch Dritte vornehmen zu lassen. Der klassische Fall ist der Pressespiegel, der durch einen «Copy Shop» für ein bestimmtes Unternehmen erstellt wird. Entscheidend für die Legalität ist laut Bundesgericht, dass die Vervielfältigungshandlung **auf Weisung** des zum Eigengebrauch Berechtigten vorgenommen wird. Wie aus diesem Beispiel ersichtlich ist, genügte es für das Bundesgericht, dass der zum Eigengebrauch Berechtigte (das Unternehmen) genaue Vorgaben über die zu speichernden Inhalte machte. Die konkrete Selektion und Auswahl dürfte der Drittperson überlassen werden.

Beim Beizug eines Dritten zum Eigengebrauch wird der **Dritte vergütungspflichtig**[143], selbst wenn die Vervielfältigung für einen zum Privatgebrauch Berechtigten vorgenommen wird, der selber nicht vergütungspflichtig wäre (siehe oben Ziff. 34). Somit ist also für jede Vervielfältigungshandlung, die durch einen Dritten durchgeführt wird (z. B. das Kopieren in der Bibliothek[144] oder Pressespiegel), wie auch für die Verwendung im Unterricht oder im Betrieb eine Vergütung an die zuständige Verwertungsgesellschaft geschuldet. Nur der selber vorgenommene Privatgebrauch ist vergütungsfrei.

38. Die Bedeutung des Eigengebrauchs für die Forschung

Für die wissenschaftliche Forschung bedeutet dies, dass sie immer dann von den Privilegierungen des Eigengebrauchs profitieren kann, wenn sie sich **im privaten Bereich** oder **innerhalb einer Einrichtung** abspielt (soweit keine betriebsfremden bzw. institutsfremden Personen miteinbezogen werden). Eine Nutzung von Werken über Institutions- oder Betriebsgrenzen hinweg bedarf hingegen einer auf den Einzelfall bezogenen Erlaubnis und behindert dadurch die Kooperation in der Forschung[145].

Die Schweiz kennt im Gegensatz zur EU[146] keine spezifischen Schutzausnahmen für den Bereich der wissenschaftlichen Forschung, hat aber durch die oben beschriebenen Schranken ein in etwa vergleichbares Schutzniveau, da die Schutzausnahmen der EU relativ eng gehalten sind[147].

Trotzdem wird in der Lehre, sofern sie sich denn dazu äussert, auch für die Schweiz ein zusätzliches eigenständiges **Forschungsprivileg** gefordert oder zumindest als sinnvoll erachtet, «zumal etwa die Reichweite der Berechtigung von Universitäten, **Forschungsdaten** im Sinne einer Zweitverwertung mittels Open Access anzubieten, gesetzgeberisch geklärt werden könnte»[148].

Im Patentrecht der Schweiz existiert bereits ein beschränktes Forschungsprivileg für die Verwendung von patentierten Lösungen, soweit sie selber Forschungsgegenstand sind oder weiterentwickelt werden[149]. Bei der unbefriedigenden Situation im Urheberrecht ist es unklar, ob es sich um ein Versehen des Gesetzgebers handelt und das Problem einfach nicht erkannt wurde, oder ob ein qualifiziertes Schweigen vorliegt. In der Lehre wird eher von einer Lücke ausgegangen, weswegen teilweise auch von einem **ungeschriebenen Forschungsprivileg** auch im Urheberrecht gesprochen wird[150].

Im Urheberrechtsbereich verbleiben also diverse Unsicherheiten, und es zeigt sich gleichzeitig auch ein gewisser Rechtssetzungsbedarf, wenn sich ein Teil der Forschung, wie sie heute eigentlich sinnvoll wäre, nicht weiterhin im Graubereich bewegen soll.

142 BGE 133 III 478.

143 Art. 20 Abs. 2 URG.

144 Das Bereitstellen eines Kopierers in einer Bibliothek wird als Beteiligungshandlung eines Dritten angesehen, welche ebenfalls der in Art. 20 URG vorgesehenen Vergütungspflicht untersteht.

145 Willi Egloff, Wissenschaftliche Forschung und Urheberrecht, S. 7 ff.

146 Die EU-Richtlinie 2001/29/EG, welche speziell für die Forschung urheberrechtliche Schranken vorsieht, wurde in den Mitgliedstaaten sehr unterschiedlich umgesetzt.

147 Das EU-Recht ist hier insofern von Interesse, als seine Beachtung die Voraussetzung für eine eventuell gewünschte Rechtskompatibilität darstellt. Zudem ist anzunehmen, dass die Schweiz mittelfristig eine Situation der mindestens gleich langen Spiesse für den Forschungsstandort Schweiz anstrebt.

148 Hilty, Urheberrecht, Rn. 226.

149 Art. 9 Abs. 1 lit. b PatG.

150 Müller/Oertli, Urheberrechtsgesetz (URG), S. 196.

39. Weitere Schranken des Urheberrechts (Archivexemplare / Bestandssicherung)

Es gibt noch eine Reihe weiterer Schranken, die das Projekt DDZ aber höchstens am Rande betreffen dürften. So ist es nach **Art. 24 Abs. 1 URG** ohne Einholen der Rechte erlaubt, eine Sicherungskopie herzustellen, wenn eines der Exemplare als Archivexemplar bezeichnet wird und an einem nicht öffentlich zugänglichen Ort aufbewahrt wird. Der Artikel bezieht sich aber nur auf Werke, die nicht mehr zu einem vernünftigen Preis besorgt werden können. Es wäre also z. B. nicht erlaubt, die gebrauchsbedingte Abnützung eines Standardwerkes zu minimieren[151]. Dies kann für Bibliotheken interessant sein, ermöglicht aber im Zusammenhang mit dem DDZ nicht den beabsichtigten Zweck.
Ebenso der **Art. 24 Abs. 1bis URG,** durch den Bibliotheken, Bildungseinrichtungen und Archive die zur Sicherung und Erhaltung ihrer Bestände notwendigen Digitalisierungen und Vervielfältigungen vornehmen dürfen, sofern damit kein wirtschaftlicher Zweck verfolgt wird. Umformatierungen und generell eine Anpassung an technische Entwicklungen sind dabei erlaubt, nicht aber das Kopieren von im Handel erhältlichen Exemplaren, um sich weitere Anschaffungen zu sparen[152].

Lizenz

40. Eine Lizenz für Forschungsdaten?

Angenommen, ein Datensatz wäre von fremden Inhalten bereinigt und würde lediglich Forschungsdaten enthalten, so könnte in der Mehrzahl der Fälle angenommen werden, dass der Datensatz auch nach der Aufbereitung durch das DDZ keinen **Werkcharakter** hätte. Damit bestünde die Gefahr, die Kontrolle über den Datensatz zu verlieren. D. h., es bestünde aus rechtlicher Sicht keine Handhabe mehr, um z. B. die Zitiervorschriften, die von den Datenlieferanten erwünscht sind, sicherzustellen. Es sind keine Rechte vorhanden, die zumindest teilweise zurückbehalten werden könnten[153].
Sofern **Forschungsdaten** also kein geschütztes Werk darstellen, ist eine mit dem Datensatz verbundene Lizenz wirkungslos und hätte höchstens eine psychologische Wirkung. Um bei nicht geschützten Forschungsdaten eine Bindung zu erreichen, müsste mit den Nutzern ein **Vertrag,** unter Einbezug entsprechender **Bedingungen,** geschlossen werden. Ein Vertragsschluss ist auch automatisiert im Internet denkbar, z. B. indem der Datensatz erst nach ausdrücklicher Annahme einer AGB[154] freigegeben wird, oder über einen Loginbereich, der ebenfalls AGBs untersteht.

Haftungsfragen

41. Haftungsrisiken des DDZ

Die Haftung der **«Access Provider»** oder der **«Content Provider»** richtet sich nach den allgemeinen Haftungsgrundsätzen. Ein gesetzlich gesondert geregeltes Internetrecht gibt es in der Schweiz nicht. D.h., es kommt insbesondere eine ausservertragliche Haftung nach Art. 41 OR in Frage. Begründet wird eine ausservertragliche Haftung durch ein **widerrechtliches Tun** oder **Unterlassen,** wobei eine Haftung durch Unterlassen nur bei der Schaffung einer Gefahr (Garantenstellung) in Frage kommt.
Ein widerrechtliches Handeln kann sich aus der Verletzung einer Datenschutznorm, eines Urheberrechts, des Persönlichkeitsschutzes[155] (z. B. Rufschädigung), des Strafrechts (z. B. Rassendiskriminierung) und weiteren weniger relevanten Schutznormen ergeben. Das DDZ ist dabei nicht als reiner Access Provider einzustufen, sondern als Content Provider (Werkmittler), und trägt damit eine gewisse Verantwortung für die aufgeschalteten Inhalte. So birgt z. B. eine nicht ordentlich durchgeführte Anonymisierung Haftungsrisiken (widerrechtliche Persönlichkeitsverletzung).
Damit die widerrechtlich verletzte Person den Schaden erstattet bekommt, muss sie ihn aber erst beweisen, was im Urheber- und Persönlichkeitsrechtsbereich bekanntermassen schwierig ist. Es wird jedoch regelmässig die Pflicht bestehen, die Verletzung zu beseitigen und künftig zu unterlassen[156].
Denkbar ist auch die Pflicht zur Leistung einer Genugtuung (Art. 49 OR) oder die Annahme einer «ungerechtfertigten Bereicherung» (Art. 62 OR) durch das DDZ.
Die Verantwortung für diese ausservertraglichen Schädigungen könnten teilweise vertraglich auf den Datenlieferanten abgewälzt werden, was jedoch auf Kosten der Attraktivität des DDZ geht. Wenn keine Vereinbarung getroffen wurde, liegt die Aufteilung der Verantwortung für die Verletzung im Ermessen des Richters[157].

Ausland

42. Aus datenschutzrechtlicher Sicht

Ob eine Bekanntgabe ins Ausland möglich ist, hängt vor allem davon ab, ob das betreffende Land einen angemessenen Schutz gewährt. Damit gemeint ist ein **zur Schweiz vergleichbarer Schutz.** Kann ein Land keinen vergleichbaren

151 Müller/Oertli, Urheberrechtsgesetz (URG), S. 312.

152 Botschaft zur Änderung des Urheberrechts 2006, S. 3430.

153 Eine CC-BY-Creative-Commons-Lizenz beispielweise bedeutet zwar einen Verzicht auf einen grossen Teil des Rechtebündels, welches einem Urheber zusteht, behält aber das Recht des Urhebers auf Namensnennung zurück.

154 Allgemeine Geschäftsbedingungen (vorformulierte Vertragsbedingungen).

155 Art. 28 ZGB.

156 Für das Urheberrecht ausdrücklich in Art. 61 ff. URG.

157 Hilty/Seemann, S. 79.

Schutz gewähren, so müssen diverse Voraussetzungen eingehalten werden, um eine Bekanntgabe doch noch zu ermöglichen[158] (z. B. eine Meldung an den EDÖB[159]). Für EU-Länder wird ein vergleichbarer Schutz angenommen.

Trotzdem müssen natürlich auch bei einer Bekanntgabe ins Ausland die allgemeinen Grundsätze der Datenbearbeitung eingehalten werden (Zweckbindung, Verhältnismässigkeit usw.). Keine Bekanntgabe ins Ausland liegt bei der Bekanntgabe von anonymisierten oder pseudonymisierten Daten vor, sofern in letzterem Fall der «Schlüssel» zur Wiederherstellung des Personenbezugs in der Schweiz bleibt. Ebenfalls keine Bekanntgabe ins Ausland ist das Aufschalten von Daten im Internet (dieser Grundsatz entspricht mehr einer praktischen Notwendigkeit als der Realität).

43. Aus urheberrechtlicher Sicht

Im Urheberrecht gilt das **Schutzlandprinzip,** was so viel bedeutet wie: es kommt das Recht des Landes zur Anwendung, für das der Urheberrechtsschutz beansprucht wird[160].

Je nach Art und Weise des Zugriffs auf die Daten können sich jedoch auch vertragsrechtliche Aspekte in die Beziehung mischen, deren anwendbares Recht sich vom urheberrechtlichen unterscheiden kann.

Die Erkenntnis daraus ist, dass die Bekanntgabe von Forschungsdaten ins Ausland zu einer Bewertung dieser Daten nach fremdem Recht führen kann.

Zum gesamten Kapitel Ausland werden noch genauere Abklärungen durchgeführt.

Denkbare Situationen und ihre Folgen

Fall 1: Der Datensatz enthält keine Personendaten, und keine unerlaubten geschützten Inhalte

- Open Data und freies Zugänglichmachen im Internet (Annahme: Aufbereitung durch das DDZ hat Werkcharakter).
  => Eine offene Lizenz (z. B. CC-BY oder CC-BY-SA) mit dem Dokument verknüpfen.
- Open Data und Zugänglichmachen im Internet (Annahme: Aufbereitung durch das DDZ hat keinen Werkcharakter).
  => Kontrollverlust akzeptieren und die Datenlieferanten dementsprechend informieren. Oder: Vertraglich (z. B. mittels AGB) die Einhaltung gewisser Bedingungen sicherstellen.

Fall 2: Der Datensatz enthält keine Personendaten, aber urheberrechtlich geschützte Inhalte

- Ein Zugänglichmachen der Daten ist nach Art. 19 URG höchstens innerhalb der Institution (SAGW) erlaubt, aber nicht darüber hinaus. Es dürfen auch keine Änderungen oder Bearbeitungen an den geschützten Werken vorgenommen werden (alles, was über ein reines Einscannen oder Kopieren hinausgeht. Z.B. die Verschlagwortung und Indexierung in einer Datenbank wären bereits eine Bearbeitung).
  Eine Archivierung von Forschungsdaten (der Ingest der Daten durch das DDZ) wird gestützt auf Art. 19 Abs. 1 lit. a. oder c. i.V.m. Art. 19 Abs. 2 URG erlaubt sein, selbst wenn darin fremde Werke enthalten sind. Ein anschliessender interner Gebrauch davon durch das DDZ wäre bereits heikel. Dieser müsste auf die Verwendung innerhalb der Institution (SAGW) beschränkt bleiben und dürfte Personen von ausserhalb nicht zugänglich gemacht werden. Dies führt dazu, dass Forschern von ausserhalb kein rechtmässiger Zugang eröffnet werden kann. Damit ist auch keine Vervielfältigung im Einzelfall über den durch Dritte vorgenommenen Privatgebrauch möglich[161] .
  Anders wäre die Situation vermutlich, wenn das betroffene Werk ein mit Einwilligung des Urhebers veräussertes Exemplar (oder ein Teil davon) wäre (also z. B. der Originalausschnitt aus einer Zeitung). In diesem Fall wäre zwar eine Bearbeitung immer noch nicht erlaubt, doch es wäre durch die Erschöpfung am betroffenen Exemplar ein rechtmässiger Zugang zum Werk möglich, und in der Folge dürfte das DDZ einem einzelnen Forscher, gestützt auf seinen Eigengebrauch, eine Kopie (bzw. eine Digitalisation davon) zukommen lassen oder zugänglich machen.
  => Eine Verwendung höchstens im engen Rahmen oder mit grosszügiger Hilfe des ungeschriebenen Forschungsprivilegs.

Fall 3: Der Datensatz enthält Personendaten, aber keine geschützten Inhalte

- Nur zugänglich nach einer vertraglichen Verpflichtung zur Einhaltung der im kantonalen oder nationalen Datenschutzrecht genannten Voraussetzungen.
  => Datenschutzvertrag

Fall 4: Datensatz enthält sowohl Personendaten, als auch geschützte Inhalte

- Nur zugänglich nach einer vertraglichen Verpflichtung zur Einhaltung der im kantonalen oder nationalen Datenschutzrecht genannten Voraussetzungen, und unter Beachtung der Vorbehalte bei Fall 2.
  => Datenschutzvertrag und siehe Fall 2

158 Art. 6 Abs. 2 DSG.

159 Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter.

160 Davon zu unterscheiden ist der Gerichtsstand, d.h. die Frage, an welchem Ort der Fall verhandelt wird.

161 Gestützt auf das Recht von Dritten, für zum Privatgebrauch Berechtigte Vervielfältigungen vorzunehmen (Art. 19 Abs. 2 URG).

Vereinfachtes Schema zur Abklärung des **Urheberrechts**: NEIN JA 2014 Version 0.5

URG:	Besteht ein Urheber- oder Leistungsschutzrecht?	Ist das Recht abgelaufen?	Ist das Werk verwaist?	Wurden die notwendigen Rechte erworben?	Liegt eine Parodie vor	Liegt ein Zitat vor?	Liegt eine Verwendung zur Berichterstattung über aktuelle Ereignisse vor?	Es gibt noch andere Schranken, die relevant werden könnten. Falls eines der folgenden Stichworte auf den abzuklärenden Sachverhalt zutrifft, lohnt sich eine nähere Abklärung: – Die Verwendung in einem Museums- oder Messekatalog – Das Erstellen von Archivierungs- und Sicherungsexemplaren, die an einem nicht öffentlichen Ort aufbewahrt werden. – Verwendung durch Menschen mit Behinderungen. – Das Nutzen von Archivwerken der Sendeunternehmen. – Das Werk befindet sich bleibend auf allgemein zugänglichem Grund.	Liegt eine begünstigte Verwendung zum Eigengebrauch vor?	Der Inhalt ist mit grosser Wahrscheinlichkeit geschützt und eine Veröffentlichung nicht erlaubt. Er darf höchstens im Umfang des jeweiligen Eigengebrauchs benutzt werden. **Kategorie 3**
VSS	Theoretisch kann jeder menschliche Ausdruck, egal in welcher Form, ein **Werk** darstellen. Entscheidend ist das Vorliegen einer bestimmten Eigenart (individueller Charakter). Es ist anzunehmen, dass wissenschaftliche Artikel, Bücher oder Werke der bildenden Künste, Werke im Sinne des Urheberrechts darstellen. Ebenso Musikstücke oder Filmwerke. Ein Werkcharakter kann sich auch in einer kreativen Anordnung oder Auswahl einer Sammlung zeigen (Sammelwerke) Auch Datenbanken können darunter subsumiert werden, sofern sie diese Anforderungen erfüllen. Der **Leistungsschutz** (das kleine Urheberrecht) schützt die ausübenden Künstler und Produzenten. Es ist also zu unterscheiden zwischen dem Werk an sich und der Sendung oder Aufnahme eines Werks.	Der Urheberrechtliche Schutz wirkt bis 70 Jahre über den Tod des Urhebers hinaus. Wenn ein Werk mehrere Urheber hat, so ist der Tod des letzten massgeblich (sofern der Werkbeitrag nicht abtrennbar ist). Die 70 Jahre werden ab dem 31.12. des Todesjahres gezählt. Ein leistungsrechtlicher Schutz dauert 50 Jahre ab dem Zeitpunkt der Veröffentlichung.	Die Verweisung ist in der Schweiz nur für Ton und Tonbildungsträger geregelt. Ein Werk gilt dann als verwaist, wenn der Rechteinhaber unbekannt ist oder nicht mehr ermittelt werden kann. Sofern die verwaiste Produktion mindestens 10 Jahre alt ist, kann das Werk über eine der zugelassenen Verwertungsgesellschaften lizenziert werden.	– Wird dem DHLab direkt vom Urheber eine (nicht exklusive) Lizenz eingeräumt oder erwirbt es gar gewisse Rechte? – Ist die Partei zu Einräumung bzw. Übertragung dieser Rechte berechtigt?	Eine Parodie dart eine bestehendes Wek benutzen, sofern die Nutzung **«transformativ»** ist, d.h. eine eigene Auseinandersetzung mit dem Ursprungsmaterial darstellt. Achtung: Eine Parodie erlaubt zwar die Benutzung des zugrunde liegenden Werks, stellt dann aber selber wieder ein eigenständiges Werk dar, dessen rechtlicher Status geprüft werden muss.	Ein Werk darf teilweise und in Ausnahmefällen sogar gesamthaft zitiert werden, sofern es als solches gekennzeichnet und die Quelle angegeben wird. Entscheidend ist der Kontext. Das Zitat muss in einem gewissen Zusammenhang zur Eigenleistung stehen und darf nicht einen Selbstzweck darstellen (eine reine Zitatsammlung wäre z.B. nicht erlaubt).	Werke, die bei der Berichterstattung über aktuelle Ereignisse wahrgenommen werden dürfen verwendet werden. Auch dürfen Berichte zu aktuellen Fragen, kleine Ausschnitte aus Fernseh- und Radioberichten sowie der Presse enthalten sein.	Unter den Eigengebrauch werden 3 Konstellationen subsumiert. Die **private Verwendung** (1), der **Schulgebrauch** (innerhalb einer Klasse) (2), sowie die Verwendung zur Information oder Dokumentation **innerhalb eines Betriebs** oder einer Institution (der Begriff ist zwar weit zu verstehen, erfasst aber keine Anstaltsübergreifende Verwendung.	
									Kategorie 1	
									Kategorie 2	

Kategorie 1: Veröffentlichung möglich. Zugänglichkeit für jedermann erlaubt.
Kategorie 2: Veröffentlichung nur unter bestimmten VSS / nach einer bestimmten Zeit / oder unter speziellen AGBs.
Kategorie 3: Gültiger Schutztitel. Nur innerhalb der Institution verwendbar. Kollektivverwertungspflichtig.

Vertrag zur Datenübernahme durch das DDZ [162]

1. Beschrieb und Name der zu übergebenden Daten
(Name der Datensammlung, der Projektleitung und sofern vorhanden der Institution. Zitiervorschlag für Sekundärarbeiten.)

2. Die übergebende Partei

a. Bestätigt, Inhaberin der Rechte an den zu übergebenden Daten, Dokumentationen und Materialien zu sein (nachfolgend «die Daten»).

b. Ist damit einverstanden, dass die Daten unter den in diesem Dokument genannten Bedingungen Dritten zugänglich gemacht werden [163].

c. Bestätigt, dass die Daten keine geschützten oder illegalen Inhalte enthalten, welche einer Nutzung nach Ziff. 3 entgegenstehen würden.

d. Bestätigt, dass die Daten im Einklang mit den geltenden Datenschutzbestimmungen erhoben und bearbeitet worden sind. Insbesondere wird bestätigt, dass die Einwilligung der Betroffenen auch eine Nutzung, wie sie das DDZ vorsieht, miteinschliesst.

e. Das Eigentum an den Daten verbleibt bei den Autoren. Die übergebende Partei erklärt sich bereit, die zu übergebenden Daten, sofern Urheber- oder Leistungsschutzrechte daran bestehen, über eine CC-BY- oder CC-BY-SA-Lizenz der Version 4 abzustossen. Für eventuell enthaltene Datenbanken kann, falls erwünscht, auch eine andere offene Lizenz gewählt werden [164]. Die Bearbeitung, Veröffentlichung, Archivierung usw. der Daten durch das DDZ erfolgt erst anschliessend an die «Öffnung» durch die CC-Lizenz und wird durch sie ermöglicht.

3. Verwendungszweck der erhaltenen Daten

a. Das DDZ ermöglicht die Archivierung, Pflege und weitere Nutzung von ausgewählten Datensammlungen.

b. Die erhaltenen Daten werden teilweise oder insgesamt aufbereitet (d.h. sofern notwendig bereinigt, digitalisiert, anonymisiert, transformiert usw.) und für die weitere wissenschaftliche Forschung (Sekundäranalysen) nutzbar gemacht.

c. Die Daten werden vom DDZ unter der Creative-Commons-(CC)-Lizenz veröffentlicht, welche von der übergebenden Partei gewählt wurde. Die gewählte CC-Lizenz kann vom DDZ bei einer Änderung der Empfehlungen für eine Open-Data-Publikation, im Sinne dieser Empfehlungen, angepasst werden [165].

4. Rechte des DDZ

a. Die übergebende Partei räumt dem DDZ alle Rechte ein, welche zur Erfüllung der unter Ziff. 3 genannten Verwendungszwecke notwendig sind.

b. Das DDZ ist ein Pilotprojekt, d.h., es kann keine Garantie gegeben werden, dass die Archivierung über eine bestimmte Dauer sichergestellt ist. Sollte es zur vorzeitigen Auflösung kommen, werden die Daten auf Wunsch im «as is»-Zustand zurückgegeben. Ebenso wird die Haftung für Datenverluste und eventuell dadurch entstandene Schäden auf grobe Fahrlässigkeit beschränkt.

5. Pflichten des DDZ

a. Das DDZ verpflichtet Drittparteien, welche Zugriff auf die Daten erhalten,
– die Datenquelle (die übergebende Partei) nach wissenschaftlichen Standards zu zitieren;
– die geltenden Datenschutzbestimmungen einzuhalten [166]
usw.

b. Das DDZ verpflichtet sich, die geltenden Datenschutzbestimmungen einzuhalten. Insbesondere wird vor einer «Open Data»-Publikation die Anonymisierung oder Pseudonymisierung von eventuell vorhandenen Personendaten vorgenommen (deren Bestehen muss bei der Übergabe kommuniziert werden).

c. Die Archivierung nach aktuellen Standards, solange das Projekt existiert und den Zielen nach Ziff. 3 dient [167].

6. Besondere Bestimmungen
(z.B. gewisse Klarstellungen, je nach Wunsch der übergebenden Partei.)

Für die übergebende Partei: Für das DDZ:

Ort und Datum:

.....

162 Daten- und Dienstleistungszentrum.

163 Im Verhältnis zu den Nutzern wird dann wiederum durch die CC-Lizenz die Verantwortlichkeit des DHLab für den Inhalt der Daten soweit zulässig beschränkt. Voraussetzung für die Gültigkeit einer solchen Klausel ist, ob sich das DHLab, aus der Perspektive des Nutzers betrachtet, den Inhalt der Daten zu eigen gemacht hat, oder ob der Schnitt-stellencharakter erkennbar blieb.

164 Z.B. die «Open Database License» (ODbL) der «Open Knowledge Foundation». Eine «Share Alike»-Lizenz speziell für Datenbanken.

165 Momentan könnte zwischen CC-BY und CC-BY-SA gewählt werden.

166 D.h., eventuell vorhandene Personendaten nach den geltenden Datenschutzbestimmungen zu bearbeiten.

167 Ausschlaggebend für diesen Entscheid ist das Gremium, welches die Auswahl von Projekten für das DDZ vornimmt.

Vertrag zur Bereitstellung für eine nicht personenbezogene Nutzung

1. Name und Beschrieb der bekanntzugebenden Daten:

.....

.....

.....

2. Titel des Forschungsvorhabens:

.....

3. Zweck der Bearbeitung:

.....

4. Verantwortliche Person(en):

.....

5. Verpflichtungserklärung:

Die verantwortliche Person verpflichtet sich in Bezug auf die oben erwähnten Personendaten zur Einhaltung folgender Auflagen:

a. Die bekanntgegebenen Daten dürfen zu keinem andern als dem im Gesuch genannten Zweck bearbeitet werden.
b. Die Personendaten dürfen nicht an Dritte weitergegeben oder diesen zugänglich gemacht werden.
c. Die Personendaten müssen, sobald es der Bearbeitungszweck erlaubt, anonymisiert verwendet werden, so dass keine schutzwürdigen Interessen von Dritten verletzt werden können.
d. Die Ergebnisse der Datenbearbeitung dürfen nur so bekanntgegeben werden, dass die betroffenen Personen nicht mehr bestimmbar sind.
e. Sämtliche Personen, die Zugang zu den Daten erhalten, müssen zuvor eine mindestens gleichwertige Verpflichtungserklärung abgeben. Die unterzeichneten Verpflichtungserklärungen sind aufzubewahren und auf Verlangen vorzuweisen.
f. Das DDZ kann verlangen, dass vor einer Publikation das Manuskript unterbreitet wird. Sofern die Publikation nach Auffassung des DDZ schutzwürdige Interessen von Personen verletzt, darf nicht publiziert werden.
g. Nach Beendigung der Auswertung der Daten, spätestens aber zum Zeitpunkt der Publikation, sind sämtliche Daten die einen Personenbezug aufweisen unwiderruflich zu löschen.
h. Die Daten sind entsprechend den im jeweiligen Dokument festgelegten Regeln zu zitieren. Falls keine solchen Regeln mit dem Dokument verknüpft sind, muss die Zitierung entsprechend der geltenden wissenschaftlichen Praxis erfolgen.

Ort und Datum: Unterschrift:

.....