RESEARCH ARTICLE

A landscape of data – working with digital resources within and beyond DARIAH

Tibor Kálmán[1] · Matej Ďurčo[2] · Frank Fischer[3] · Nicolas Larrousse[4] · Claudio Leone[5] · Karlheinz Mörth[2] · Carsten Thiel[5]

Abstract

The way researchers in the arts and humanities disciplines work has changed significantly. Research can no longer be done in isolation as an increasing number of digital tools and certain types of knowledge are required to deal with research material. Research questions are scaled up and we see the emergence of new infrastructures to address this change. The DigitAl Research Infrastructure for the Arts and Humanities (DARIAH) is an open international network of researchers within the arts and humanities community, which revolves around the exchange of experiences and the sharing of expertise and resources. These resources comprise not only of digitised material, but also a wide variety of born-digital data, services and software, tools, learning and teaching materials. The sustaining, sharing and reuse of resources involves many different parties and stakeholders and is influenced by a multitude of factors in which research infrastructures play a pivotal role. This article describes how DARIAH tries to meet the requirements of researchers from a broad range of disciplines within the arts and humanities that work with (born-)digital research data. It details approaches situated in specific national contexts in an otherwise large heterogeneous international scenario and gives an overview of ongoing efforts towards a convergence of social and technical aspects.

Keywords Research infrastructure · Digital humanities · Arts and humanities · Sustainability · DARIAH · FAIR principles

1 Introduction

Funding agencies, on both the European and national levels, increasingly require that research data and publications produced in publicly funded research projects be

---

✉ Tibor Kálmán
tibor.kalman@gwdg.de

Extended author information available on the last page of the article

published in an open access format. Policy recommendations on research data management are being revised in the context of Open Science (European Commission 2018). It has become a common practice for researchers to publish their research data in an open-access fashion, using free or permissive licenses. In the arts and humanities in particular, however, data sharing and reuse among researchers is not a commonly established practice. Even if researchers in these disciplines published their data in European repositories and archives, this data is often hard to find, access, or reuse. Even if there were an increased awareness of the need and benefit of sharing resources within the disciplines of the arts and humanities, much needs to be done to make it an integral part of an everyday research practice.

The sharing of resources is an inherently complex phenomenon that involves many different actors and is influenced by many factors. Challenges to the level of the data itself are well summarised by the FAIR principles, which comprise of stable identifiers, rich, broadly disseminated metadata, widely adopted formats, vocabularies and protocols (Wilkinson et al. 2016). These requirements need to be supported by an appropriate technical infrastructure: (a) stable *repositories for depositing* and publication of the data; (b) means for broad *dissemination* of metadata, most notably the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) in combination with large-scale aggregators; (c) *authentication and authorisation* infrastructure (AAI), allowing for fine-grained handling of permissions and (d) *interoperability* between tools, i.e., support for established formats and availability of well-defined APIs and import/export functionality to ensure permeability and an easy data flow within the research process. These technical requirements need to be underpinned by policy measures: *promotion of standards* and permissive intellectual property rights (IPR) for research seconded by clear *licensing*. It is also important to establish academic gratification for the creation and publication of research data and software, as well as to appreciate its value as research output and enable a proper *academic contribution*. The latter point is particularly crucial: while the other aspects could be considered as, primarily, enabling factors, the gratification aspect constitutes a strong incentive for researchers to willingly share their work.

All of these measures need to be accompanied by appropriate training and outreach campaigns, raising awareness and ensuring the transfer of this kind of knowledge. Both scholars and students and the interested public need to have the opportunity to acquaint themselves with digital methods, technologies, formats and best practices. Ideally, this should take place in intensive, small-scale, hands-on settings, which focus on individual aspects, up-to-date online training material, comprehensive documentation, and opportunities for on-demand personal consultations with experts.

The sharing of resources should not be seen as a mere handover of data, but rather as an integral aspect of working with digital resources, interwoven with all the various stages of the research data lifecycle, from creation and curation to dissemination of digital resources for reuse and knowledge acquisition. It naturally affects and is affected by all stakeholders in the research area. While the decision of individual scholars to share the resources they created is the *conditio sine qua non*, it is crucial to embed the resource in a fruitful, supportive broader environment that ensures all the above-mentioned enabling factors. The traditional institutional context might be the home organisation of the scholar, but given the global challenge to increase the accessibility of research data, the issue at stake cannot be addressed by individual institutions

anymore and requires joint efforts on many levels, involving entities from the individual research groups up to European and global institutions. Research infrastructure consortia feature a multi-layered structure, ranging from topic-specific working groups and national consortia to the governing bodies on a European level. They are in an ideal position to tackle these multifaceted challenges. Not only do they represent their respective community, but they are also an integral part of it, possessing a deep understanding of research practices in the field.

This article gives an overview of the ongoing developments and reflects on the current discourse within and beyond the DARIAH research infrastructure. It is structured as follows: First, we present the DARIAH initiative in detail, including the reasons for its initiation and its unique position in the European context. We then shift our focus to describe different national chapters of DARIAH and their take on dealing with (born-)digital research data collections in a heterogeneous research environment. By helping to moderate the change of scientific practices in the humanities, we aim to make it easier to integrate digital and technical aspects into research workflows in disciplines that were previously rather 'untechnical'. Some remarks on our work towards a convergence of social and technical aspects of this endeavour will conclude the article.

2 DARIAH – A digital and distributed infrastructure for the arts and humanities

A research infrastructure can serve as the basis for offering services and resources for the sharing and management of data and for the management of associated legal and organisational issues. Developing such a sustainable research infrastructure, which integrates existing resources, tools and services to broaden the possibilities of a truly open science, and promotes the acceptance of digitally-enabled approaches is also the *raison d'être* of the DARIAH initiative.

DARIAH is short for Digital Research Infrastructure for the Arts and Humanities. This pan-European organisation aims at enabling and supporting digital research methods and teaching across the arts and humanities (DARIAH 2018). DARIAH-EU, as the umbrella organisation is called, was founded in the framework of the European Strategic Forum for Research Infrastructures (ESFRI) and first appeared on the ESFRI roadmap in 2006 as one of six projects for the humanities and social sciences (European Roadmap for Research Infrastructures 2006: 33). Within the ESFRI, the legal form of European Research Infrastructure Consortium (ERIC) has been developed to enable the funded European research alliances to operate on a stable, long-term basis. After a long preparation phase, the DARIAH-ERIC was established by the European Commission in August 2014. To date, 17 countries— Austria, Belgium, Croatia, Cyprus, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Poland, Portugal, The Netherlands, Serbia and Slovenia—have become DARIAH members, and the list of cooperating partners in these and other countries is growing. Six further candidate countries are expected to become members by 2020.

In practice, DARIAH is a vivid marketplace of ideas and know-how, where people from different countries and disciplines can meet and collaborate, help and learn from each other. It addresses the aforementioned challenges in many different ways. Mainly through its individual partners, DARIAH provides the necessary basic technical

infrastructure and specialised tooling to underpin the whole research process; be it virtual research environments (VRE) for co-creation and publication, repositories for long-term preservation and publication of research data, general publication platforms, or generic project-management solutions, allowing efficient communication in highly distributed collaboration setups. Around these technical efforts, DARIAH also organises numerous training and outreach events to raise awareness and transfer practical skills for digital methods to the scholarly community.

On the European level, DARIAH uses its unique position and capacity to push forward necessary policy work that makes the handling and especially sharing of research resources easier. It propagates the utilisation of standards to address the problem that large parts of the produced research data are neither visible, nor reusable (legally or technically). This is why DARIAH engages in the Open Science Policy Platform (OSPP) (Edmond 2018). In the framework of the ongoing project DESIR (DARIAH ERIC Sustainability Refined, see CORDIS 2018), DARIAH has identified six dimensions of sustainability that it seeks to strengthen: dissemination, growth, technology, robustness, trust, education. Up until the projected end of DESIR in December 2019, we will see international workshops and other types of dissemination events to initiate collaborations and further educational work, and the existing services will be enhanced with a focus on entity-based search, scholarly content management, visualisation and text-analytic services. Furthermore, DARIAH collaborates with other SSH infrastructures such as CESSDA (Consortium of European Social Science Data Archives, see CESSDA 2018), CLARIN (Common Language Resources and Technology Infrastructure [see CLARIN 2018]), and the emerging research software engineering community. The aim is to find a common understanding of how to sustain research software, to address specific challenges of research infrastructures, and to develop a unified technical reference (Kalman et al. 2018). It is a declared task in the DARIAH Strategic Action Plan, released in November 2017, to help developing sustainability models for Digital Humanities (DH) projects and their data collections, especially to ensure the longevity of such projects after the direct funding period has run out (DARIAH 2017).

In the future, DARIAH aims at working towards a more resilient, robust setup of the technical infrastructure, making datasets and services more independent from individual providers through stronger cooperation between partners of the consortium, and with e-Infrastructures like EGI (EGI 2018), EOSC (European Commission 2017) or EUDAT (EUDAT 2018), offering basic generic services. With concentrated expertise both on infrastructural aspects and on actual research in the Digital Humanities, DARIAH can act as a broker and mediate between the needs of individual research projects and the large-scale technical solutions offered by e-Infrastructures. Several initiatives were started to lay the technical and organisational groundwork for such collaboration between DARIAH and related e-Infrastructures. For instance, the EGI DARIAH Competence Centre (Harmsen et al. 2015) helped with pilot projects like Storing and Accessing DARIAH contents on EGI (Wandl-Vogt et al. 2017), to analyse, distinguish and meet DARIAH requirements within the EGI infrastructure. The EOSC-hub initiative, which consolidates and integrates access mechanisms to e-Infrastructure resources, recently initiated its DARIAH Thematic Service (Dumouchel 2017) to strengthen the collaboration. Through institutions that are active in both CLARIN and DARIAH, there is cooperation with EUDAT, with particular regard to topics related to preservation and access to long-term storage resources.

3 National Flavours of DARIAH

In this Section, we give an overview over different approaches and national flavours of DARIAH that are working with and sharing a wide variety of data and services through software and tools as well as accompanying learning and teaching material. We present three different examples of DARIAH member countries that demonstrate how national activities contribute to the overall goals.

A crucial characteristic of the DARIAH research infrastructure is its distributed nature as a federated network where most of the services are not offered by a central instance, but through the contributions of individual partners. There are various ways in which DH research communities, their data, and their supporting infrastructures are embedded in the national research landscapes.

3.1 DARIAH in Austria

3.1.1 National consortium CLARIAH-AT

Right from the start, the national group of humanities research infrastructures in the humanities was set-up as one joint organisational structure comprising of both CLARIN and DARIAH (Ďurčo and Mörth 2014). This approach proved to be very efficient and successful. Interestingly enough, dynamics aiming at a higher degree of interaction and cooperation can also be seen in other countries. In the Netherlands, two infrastructures run one big national project; in Denmark and France, the coordination of both RIs is placed with the same person or institution; in Germany, talks on greater interaction are ongoing, and in other countries similar tendencies can be discerned. The Austrian Centre for Digital Humanities at the Austrian Academy of Sciences (ACDH-OeAW 2015) is the coordinating national institution for both research infrastructures. The centre was founded with the intention to foster the change towards digital paradigms in the humanities and pursues a dual agenda of conducting digitally enabled research and providing technical expertise and support to the research communities at the Academy and in the Austrian research landscape.

ACDH-OeAW is not the only player in Austria offering services for the digital humanities community. In CLARIAH-AT, the national group of institutions involved in the two European Research Infrastructure Consortia CLARIN and DARIAH, 14 partner institutions work together to provide a common framework to improve the situation with respect to efficiency of dealing with research data. In 2015, numerous partners of the consortium contributed to a national strategy for Digital Humanities in Austria (Alram et al. 2015). One of the central goals of this strategy, which was fleshed out at the request of the then Ministry for Science, Research and Economy, was the creation of infrastructures to guarantee long-term preservation of research data. One of the measures proposed in the strategy to achieve this goal was the establishment of a national repository federation to ensure long-term access to research data hosted by exchanging expertise, sharing technologies, and interlinking repository resources. The long-term goal is to reach an agreement between individual partners of the federation making sure that partners would step in with their repositories as fall-back options in case one of the participating repositories ceases to exist. Implementation of the measures is part of the agenda for the CLARIAH-AT consortium for the upcoming three-year period.

3.1.2 Data services – One-stop shop for DH projects

In the following, we highlight one specific institution, the ACDH-OeAW, to exemplify how local centres support their respective communities, contributing their share to the common cause. ACDH-OeAW strives to cover the whole research process: project planning, data modelling, data curation and processing, digitisation, application development, service hosting and especially long-term preservation of data. All of this is accompanied by personal consulting and support for individual research endeavours and knowledge transfer, as well as outreach activities promoting the use of digital methods in the various fields of the humanities.

Stable, reliable, long-term preservation of research data being an essential precondition for sharing of resources, the ACDH-OeAW is running a repository called ARCHE (A Resource Centre for the HumanitiEs) (ARCHE 2017) as one of its core services offering stable hosting of digital research data—in particular, for the Austrian humanities community. ARCHE welcomes data from all researchers in the Austrian Academy of Sciences, but also from other institutions in and outside the country. While its predecessor, *CLARIN Centre Vienna / Language Resources Portal*, was dedicated to digital language resources, ARCHE is open to a broader range of disciplines. ARCHE is mainly meant to preserve resources related to Austria, which would include resources that were collected or created in Austria, or involve a geographical area or historical period of interest to Austrian scholars. The collection policy details the types of data the repository is ready to accept and store. ARCHE has been awarded the CLARIN B centre status and certified under the Core Trust Seal (CoreTrustSeal 2018), formerly Data Seal Approval.

Secure and robust long-term preservation of data hinges on many factors. Next to the technical level (bitstream preservation), a host of data-related aspects (metadata, established formats), and the institutional setting are to be considered. ARCHE explicitly states which formats it recommends and accepts for depositing. The categories are 'preferred' and 'accepted'. Preferred formats are expected to be stable and usable also in the long-term. Accepted formats are considered less reliable for the long-term and are converted to one of the preferred formats during the ingest process, both formats being stored. The preservation plan, which is currently being developed, will describe the workflow for format monitoring and migration, so as to ensure that data is preserved if formats become obsolete.

ARCHE pursues the principles of Open Access and Open Data. It encourages data depositors to use open licences, like CC-BY and CC-BY-SA, adhere to rules for good scientific practice, and apply the FAIR Data Principles. The repository itself supports the FAIR principles in various ways. Not only does it make the data *findable* by offering search and browse functionalities, but it also makes it available for harvesting through third-party aggregators, such as CLARIN's metadata catalogue Virtual Language Observatory (VLO) (Van Uytvanck et al. 2010), by means of publishing metadata via OAI-PMH. It makes the data *accessible* by assigning persistent identifiers and *interoperable* by promoting the use of recommended formats and offering direct access to the data and metadata for both human and machine interaction. And, finally, all of these measures contribute to the *reuse* of the data.

In addition to ACDH-OeAW, two other participating institutions have been providing stable hosting and publishing solutions for research data: the Centre for Information Modelling, with the ACDH at the University of Graz running the repository GAMS (Stigler and Steiner 2014) and the University of Vienna, with the PHAIDRA repository

(Budroni and Höckner 2010). All three repositories build on Fedora Commons (Fedora 2018), GAMS being an integrated system which comes with a specialised ingest tool and a Text Encoding Initiative (TEI) based publication framework. The common technical framework is a good basis for establishing a repository federation, where data could be transferred to and hosted by one of the other partners in case one of the services would shut down.

Although sustainable preservation of data is an indispensable part of up-to-date data management in research, there are a number of other components required to cover the whole range of workflow steps in digitally working projects. We refer specifically to tools for automatic processing of data and also solutions supporting the manual collaborative creation and curation of born-digital data (commonly referred to as virtual research environments). Confronted with a multitude of projects with at times very individual needs, ACDH-OeAW adopted a pragmatic approach, trying to use what is there and to provide the missing pieces. In practice this means, e.g., that data encountered in projects encoded in MS Word or Excel files are converted to formats better suited to the long term, like TEI or Simple Knowledge Organisation System (SKOS). Yet, in other cases, we develop project-specific web-based applications with custom-tailored data models, which allow the project teams to create and curate data collaboratively. While this may seem inefficient, we increasingly witness consolidation tendencies and economies of scale, as the colleagues supporting the projects gain more experience in generic frameworks, which allows us to develop new applications with considerably less effort, and re-integrate new functionalities required by new projects back into the common code-base.

For ACDH-OeAW, knowledge transfer and outreach are central pillars of the DH strategy. The team organises numerous training activities, most notably the two event series ACDH Lectures and ACDH ToolGallery. The latter being a one-day format, in which various practical tools are presented in a combination with a theoretical introduction on a given topic and a hands-on session, giving participants a chance to try out a particular tool with the support of a qualified expert. ACDH-OeAW also runs the platform Digital Humanities Austria (DHA 2015), which is the main national dissemination channel for DH in Austria; it is used to announce events and features a comprehensive exhibition of DH projects and a DH bibliography, which serves as an entry point for humanities scholars to delve into DH. An essential part of the community-building efforts is the annual DHA conference, which was organised by ACDH-OeAW in the first three years, before starting to move to other Austrian cities: in 2017, the conference was organised by the Research Centre Digital Humanities at the University of Innsbruck.

Part of the institute's strong commitment to training & education is also the provision of two specialised services for the DH community: #dariahTeach (DARIAH-TEACH 2017), an e-learning platform for teaching material for DH, and the DH Course Registry (DH-registry 2017), an online catalogue providing an overview of DH-related curricula in Europe being collaboratively maintained by CLARIN and DARIAH.

3.2 DARIAH in Germany

3.2.1 National consortium – DARIAH-DE

DARIAH-DE is the German national contribution to DARIAH. It currently consists of a consortium of 19 partners, comprising universities, academies of sciences and

independent research institutions, libraries, data centres, a non-governmental organization (NGO) and a commercial partner (DARIAH-DE 2018h). Now in its third project phase, DARIAH-DE receives funding from the German Federal Ministry of Education and Research. The project's current focus is the preparation of the operational phase in 2019, aimed at providing a permanent infrastructure for the arts and humanities in Germany, a process which DARIAH-DE and CLARIN-D are jointly advancing in close collaboration with the ministry, the academies of sciences and disciplinary stakeholders (Forschungsinfrastrukturen für die Geisteswissenschaften 2018).

The heterogeneous nature of the DARIAH-DE consortium enables the research project to address the multi-faceted challenges for research infrastructures. Two pillars of DARIAH-DE are its tight integration with research and teaching through its partners. Dedicated work packages focus on quantitative data analysis, visualisation and annotation with the two focal points addressed in each. Another work package researches the impact and reach of DH in the humanities community, while a strong collaboration with CLARIAH-AT under the umbrella of #dariahTeach focussed on curricular, educational and training materials on a wide variety of topics.

The third main aspect is the provision and operation of the technological infrastructure: from basic components such as servers, monitoring and user support through collaboration solutions and development toolchains to the layer of scholarly services. For these, DARIAH-DE's infrastructure partners, such as data and computing centres and libraries, provide existing and well-established components and services. This includes an authentication and authorisation infrastructure (AAI) that is part of the worldwide authentication network, built by the higher education and research institutions. Over the course of the DARIAH-DE project, the tight collaboration of the developers embedded in their fields and the service providers operating the services have been focused upon and sustainability solutions have been developed to ensure the basis for the long term operation of this infrastructure.

Finally, the pillar most relevant to the present article is dedicated to the processing and storing of research data, for which several tools and services are offered. Building on the TextGrid project, DARIAH-DE has continued the development of the TextGrid Repository, focussed on critical digital scholarly editions and optimised for XML-TEI encoded data, to build the DARIAH-DE Repository (cf. DARIAH-DE 2018g). The operation of the repository is institutionalised through the Humanities Data Centre (HDC), a joint venture of Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) and Göttingen State and University Library (SUB). Both institutions thus ensure the sustainability of all data stored in the repository. This repository is one component of the *Data Federation Architecture* (DFA, see (Gradl and Henrich 2016) for an overview of the underlying concepts and Fig. 1 for the underlying workflow) offered by DARIAH-DE to manage research data.

3.2.2 Data services – A federation architecture

The DFA consists of the DARIAH-DE Repository, the Collection Registry, the Generic Search and the Data Modeling Environment (DME). All components (services and applications) of the DFA are designed to interact with one another. They can be used all together or as standalone services depending on the individual needs of the researcher.
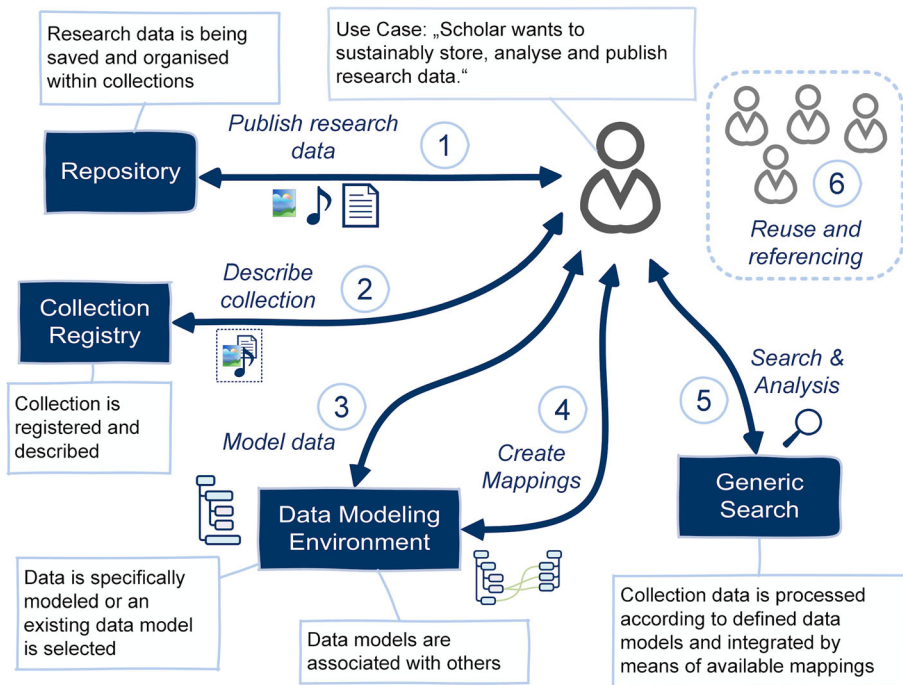
Fig. 1 DARIAH-DE Data Federation Architecture, Tobias Gradl (updated version from: Gradl et al., 2015, used with permission)

The *DARIAH-DE Repository* (DARIAH-DE 2018f) is a digital long-term archive for humanities and cultural scientific research data, enabling researchers to store and publish data in a secure and sustainable manner. At the entry point, the DARIAH-DE Publikator (DARIAH-DE 2018e) offers a user-friendly web interface for data management, description and ingest into the repository. The storage backend is divided into two areas: a restricted private storage area and a public area. All preparation for publication is done in the private storage area via the Publikator and involves three simple steps: First, a collection needs to be created; second, all associated data belonging to the collection has to be uploaded and, finally, all data has to be described by metadata. The repository uses the Dublin Core Simple (cf. Dublin Core Metadata Initiative 2013) metadata standard for description of data, only a few fields are mandatory, such as licence information. Furthermore, persistent identifiers for stable referencing are provided through the publication process – the collections as well as all associated objects get individual Digital Object Identifiers (DOIs). There is a dedicated PID-Service as part of the DFA for assigning unique identifiers and registering them at the DataCite DOI-network. Once published, all data is publicly available.

After publication, an optional but highly recommended possibility is the registration of the collection in the *Collection Registry* (DARIAH-DE 2018a). The Collection Registry enables researchers to make their published data even more visible and understandable and, therefore, more accessible. A draft entry with the metadata already mentioned is automatically created during the publication process and stored in the Collection Registry for further enrichment. For this, a dedicated metadata model for enhanced description of collections and associated data is provided: the DARIAH

Collection Description Data Model, DCDDM (see DARIAH-DE 2017), based on (Dublin Core Metadata Initiative 2007). Once the collection is registered, all data is searchable via the DARIAH Generic Search interface. Due to the modular design of DARIAH's Data Federation Architecture, all kinds of metadata—including such that describe data published outside the DARIAH-DE DFA—can be registered and made accessible for the Generic Search. Information on how to access data can be provided, including specification of interfaces and APIs. This includes data that originate in a digital form, but also non-digital data or collections of objects.

The design of the *Generic Search* (DARIAH-DE 2018c) is aimed at providing researchers in the Digital Humanities with an individually adjustable search facility for their research needs. The myCollections functionality enables them to compile their own query by preselecting the sources out of the Collection Registry, store and share them with research colleagues. This allows researchers to precisely query predefined metadata sets. Custom collections can be added at any time via the Collection Registry interface to enlarge the data set of their own query.

The Generic Search is accessible without registration and allows a combination of different search strategies and dynamic adjustment of the enquiry's granularity, e.g., by adjusting the faceted classification or the number of included collections.

If collections with different metadata schemes need to be integrated into the DFA, the *Data Modelling Environment (DME)* (DARIAH-DE 2018b), as a further component allows a web based user-friendly mapping and association of metadata fields. The web interface enables researchers to explicate their knowledge on the semantic description of their collections. This bottom-up approach allows for more flexibility when including additional external sources, without enforcing explicit standards. This is especially important for the arts and humanities disciplines with their variety of perspectives on collections, terminology and data models.

Besides the Data Federation Architecture, which is designed for research data management purposes of all disciplines within the arts and humanities, DARIAH-DE also offers tools and services that are used for specific project contexts or are related to specific research methods. There are general services for collaborative work and project management allowing collaboration across locations. Furthermore, tools for annotating, analysing and visualising data are provided. A prominent example is the *Geo-Browser* (DARIAH-DE 2018d), which allows the analysis of space-time relations of data and collections of source material, facilitating their representation and visualisation in a correlation of geographic spatial relations at corresponding points of time and sequences.

Additionally, a virtual research environment (VRE), especially designed for the creation of digital editions based on XML/TEI, offers open source tools and services to collaboratively edit and generate research data. The VRE TextGrid (TextGrid 2018) enables the editing, storing and publishing of data for scholars in the humanities in a protected environment.

DARIAH-DE is not only a digital research infrastructure, but also a social infrastructure. It fosters exchange of experiences and expertise and offers a variety of communication and training facilities, like user meetings, issue specific workshops with hands-on sessions, and regular events on the theme of Digital Humanities, spanning a broad range of topics. The information supply of DARIAH-DE is continuously being enhanced and provided through multiple channels and platforms, e.g. through a Digital Humanities blog (DHdBlog), a Twitter account with current news, a YouTube channel (DHd-Kanal) with

tutorials, a "Doing Digital Humanities" bibliography as well as many publications and presentations which have been created during the seven years of project lifetime so far.

DARIAH-DE creates a network of digital humanities services, expertise and communities to support research and cooperation in the humanities and cultural sciences, and promotes open access sharing of digital resources.

3.3 DARIAH in France

3.3.1 National consortium – DARIAH-FR

The CNRS (Centre National de la Recherche Scientifique – National Centre for Scientific Research) is a public organisation under the responsibility of the French Ministry of Education and Research. The CNRS, in connection with universities, has implemented an ecosystem aiming to cover the entire lifecycle of the production of scientific data and publications in the Humanities and Social Sciences. This ecosystem is based on the following infrastructures: Open Editions (2018), CCSD (Centre pour la Communication Scientifique Directe 2018), PERSEE (Portail de diffusion de publications scientifiques) and TGIR Huma-Num (Très Grande Infrastructure de Recherche Huma-Num 2018).

Huma-Num coordinates the participation in DARIAH and CLARIN of the above-mentioned organisations, as well as other potential contributors, such as Huma-Num's national consortia (see below). It is also involved in other European and international projects like OPERAS (OPERAS 2018). Huma-Num is an infrastructure that aims to facilitate the digital turn in Humanities and Social Sciences and is part of the national ESFRI roadmap, which is in turn aligned with the European Union's ESFRI framework. This allows good perspectives for recurrent funding.

To perform these missions, Huma-Num's organisation is based on both human and technological layer. It funds "groups of people", called *consortia*, working on common areas of interest (e.g., similar scientific objects) and also provides a technological infrastructure, offering a variety of platforms and tools to process, preserve and disseminate digital research data.

The main idea of a consortium is to organise multidisciplinary collective dialogue within research communities by bringing together different types of actors (researchers, technical staff, etc.) coming from different institutions, with the aim of creating synergies. In return, a consortium is expected to provide technological (or scientific) good practices and produce corpora, new standards, and tools.

Furthermore, Huma-Num provides a technological infrastructure on national scale, based on a large network of partners. Technically, the infrastructure itself is hosted in a big data centre built by and for physicists. A long-term preservation facility from another data centre (CINES – Centre Informatique National de l'Enseignement Supérieur) is also utilised. In addition, a group of correspondents in the "Maison des Sciences de l'Homme" network (MSH Network 2018) all over France is in charge of relaying information about Huma-Num's services and tools.

3.3.2 Data services throughout the data lifecycle

Huma-Num provides tools and services for each step in the research data lifecycle. It coordinates the production of digital data, while offering a variety of platforms and tools to

process, preserve and disseminate the data. It also provides research projects with a range of utilities to facilitate the interoperability of various types of digital raw data and metadata (see Fig. 2).

More specifically for digital collections, the aim is to foster the exchange and dissemination of metadata, and of the data itself, via standardised tools and lasting, open formats. These tools developed, by Huma-Num, are all based on semantic web technologies, mainly for their auto-descriptive features, and for the enrichment opportunities they enable. All our resources are, therefore, fully compatible with the Linked Open Data (LOD).

Three services have been designed and developed by Huma-Num to process, store and display research data, while preparing them for re-use and long-term preservation; to put it another way, the aim is to provide a chain of tools to make data FAIR. These complementary services embrace the research data lifecycle and are designed to meet the needs arising there from: constitute a coherent chain of research data tools. While they interact smoothly with one another, they are also open to external tools using the same technologies.

The scientific objective is to promote data sharing so that other researchers, communities, or disciplines, can reuse them, including from an interdisciplinary perspective and in different ways. A map, for example, may become a scientific object, which reflects both the point of view of a geographer and that of a historian. More generally, the principles and methods of the Semantic Web (RDF, SPARQL, SKOS, OWL), on which these services rely, enable data to be documented or re-documented for various uses without confining them to inaccessible silos. Another important point is to make the storage of data independent of the device used to disseminate the data. Another objective is to prevent the loss of data by preparing their long-term preservation. Documenting the use of appropriate formats, which are the basis of data interoperability, greatly facilitates the archiving process.

The workflow implemented by Huma-Num has been built on interoperability. The aim is to foster the exchange and dissemination of metadata, but also of the data themselves via standardised tools and lasting, open formats. Huma-Num uses different technologies for cold, warm and hot data. If the technology used for hot data was quite

Fig. 2 DARIAH-FR's Services for Data, Huma-Num

classical, for warm data, Huma-Num has established a mesh of distributed storage all over France (currently 9 nodes) using different storage technologies encapsulated. Thus, backup and versioning can be made on any node. Furthermore, the data center where Huma-Num's infrastructure is hosted provides a backup on tapes for cold data.

Huma-Num already provides a long-term preservation service based on the CINES (Centre Informatique National de l'Enseignement Supérieur, 2018) facility, a National Computer Center of Higher Education which is responsible for permanent archiving for scientific data in France. This is much more than the bit preservation done with the above-mentioned technologies. A long-term preservation project means that one needs to organise the data with a view to reuse by someone, who did not participate in its creation, that presupposes a lot of curation. In addition, the data should be expressed in a format accepted by the partner and additional information has to be provided to document the context of data production, metadata, etc. Huma-Num accompanies these projects by acting as go-between linking data producers, CINES, archivists and other actors.

After a detailed description of three national landscapes, we now shift our focus to the ongoing efforts towards a convergence on the European level in light of the heterogeneity of research data collections, of formats, tools and services.

4 Convergence of tools, methods and collections

It was always the vision of DARIAH to enable the DH research community to reuse and build on existing solutions, developed in and by the community. This includes both the social and the technical aspects of the convergence from individual solutions to a distributed infrastructure.

The social aspect builds around the idea of an Open Marketplace, which enables us to share and review existing services and solutions. From the technical side, DARIAH has identified the need to address the sustainability of the software, which provide some of the core parts of any digital infrastructure. In the following section, we describe how these are being addressed.

4.1 The open marketplace

The idea of developing DARIAH 'as a social marketplace for services' (Blanke et al. 2011) dates back as far as to the preparatory phase of the DARIAH initiative.

The long-term goal is to provide an Open Marketplace platform, which is planned as an easy-entry place where scholars can find solutions for the digital aspects of their daily research work, such as software, tools, (born-)digital data sets, repositories, services, learning and teaching material. The Marketplace targets all researchers from the broader SSH, not just those scholars who would regard themselves as *digital* humanists. Various approaches had been started in the past to provide collections and registries with similar goals. The most important difference between such approaches and the DARIAH Marketplace is that it will contextualise the tools and services offered, with user feedback, user stories, links to training material, showcases, contact addresses, ratings. It is going to be actively curated and sustained by the DARIAH community. The idea is not that these solutions would be produced by DARIAH itself, but that

the Marketplace creates visibility for them to help researchers do their work (DARIAH 2017) (Fig. 3).

There have been previous attempts at providing an active, community-backed registry of digital tools and services. While most of them did not always live up to their expectations (for a prominent example cf. Dombrowski 2014), one can still learn from them and reuse their highly curated data. Such an attempt was undertaken within the framework of the H2020 project "Humanities at Scale" coordinated by the DARIAH-ERIC. Building on TERESAH, the "Tools E-Registry for E-Social science, Arts and Humanities" originally developed within the FP7 project "Digital Services Infrastructure for Social Sciences and Humanities" (DASISH) until 2014, a demonstrator for a central registry with distributed data sources was created (Engelhardt et al. 2017).

While the DARIAH Marketplace is still being formed, it is the declared goal not to just add another list-based overview of digital tools, but to assemble and highlight DH knowledge. The platform will create a place addressing and involving the entire research community and also, eventually, the public and industry (bearing in mind EOSC and EU access policy guidelines for research infrastructures).

4.2 Sustainability of tools and software

The social aspect of the marketplace is built on the idea of sharing and reviewing existing services and solutions. In the case of software, providing some of the core technical parts of any digital infrastructure, DARIAH has identified the need to address its sustainability problems (cf. Thiel 2017). In the current status-quo, the construction of sustainable infrastructures is done through grant-based research projects, which has a number of problems. Software built to address specific research questions is often developed in an ad-hoc manner. This is not helped by the fact that software is not yet generally accepted as creditable research output in and of itself. Without a recognition of
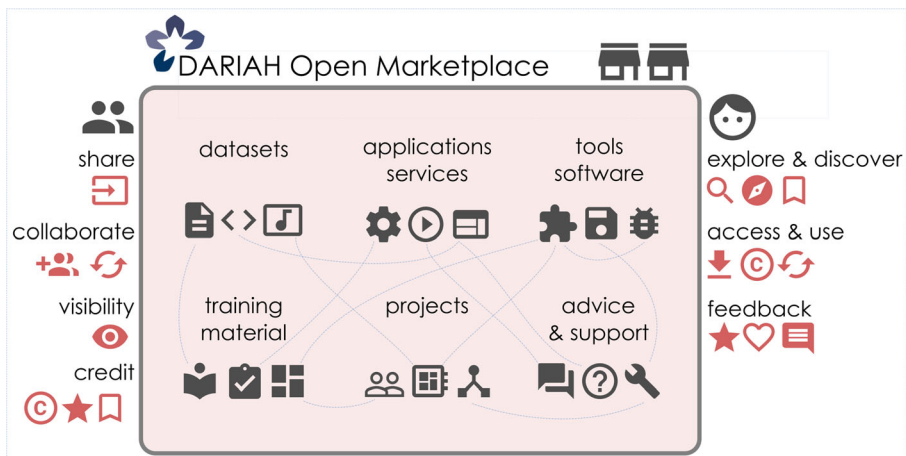

Fig. 3 Illustrative sketch of DARIAH Open Marketplace

the value of the software as a form of research, the individual researcher's willingness to invest additional time into improving the software in a way that does not directly impact the output will be minimal.

The requirement to provide data management plans as part of H2020 grants, which is implemented by national and other funders, sees source code as being identified as digital resources that need preservation. To address this, the UK's Software Sustainability Institute developed a solution to create a Software Management Plan through DMPonline (Software Sustainability Institute 2018) and GitHub and Zenodo have joined forces to add a simple possibility to publish GitHub releases in Zenodo, making software releases citable through DOIs (GitHub 2016). Archiving code is the first step in ensuring the availability for future re-use and reproducibility of research output generated with that software. The second step is making sure that the code can be processed and executed when needed, which goes beyond classical practices of data curation, (cf. Katz et al. 2016) for a discussion on the topic. In our context, two problems are most relevant. For reproducibility of results, access to the entire exact build environment is required and it must, therefore, be referenced in the archived software in a machine readable format. For re-use of the software, the adaptability to the constantly changing reality of information technology, such as changes to external libraries and dependencies, becomes relevant. As technology progresses, so do research questions and new applications not envisioned during the original development can emerge (cf. Harms, Grabowski 2011). For a future researcher to be able to actually adapt a given software product, sufficient documentation and code legibility must exist. While research thrives on innovative solutions with fast-paced development progress, the requirements for software maintainability for the long run are directly contrary (see Hettrick 2016, Chapter 3) for a more detailed discussion.

This is also a particular problem for infrastructures striving to sustain software developed within projects as services. To be able to do so, the infrastructure providers must make a judgement on the expected and unexpected cost that long-term software maintenance will incur. This can only be done if the software is of sufficiently good quality. To address this, infrastructures are developing guidelines and best practices for developers. At the same time, existing quality measures, such as ISO standards, can be one frame of reference (see e.g. Buddenbohm et al. 2017), while (Doorn et al. 2016) suggest establishing an independent certification, modelled on the Data Seal of Approval, now CoreTrustSeal (CoreTrustSeal 2018).

For an infrastructure to provide a valuable service to the scholarly community, the reliability and the trustworthiness of the services offered is a fundamental prerequisite. By improving the quality of the software and making this transparent to the end user of the technology through the Open Marketplace platform, DARIAH strives to address both. In particular, through DESIR work was started on a general Technical Reference (Moranville et al., 2018) as baseline for new development and the Marketplace will improve the findability and discoverability of research software. The combination of both supports and builds upon known recommendations for research software (Jiménez et al. 2017).

5 Conclusion

We have summarised ongoing developments and reflected current discussions within the research infrastructure DARIAH and within some of DARIAH's member states, which are creating and integrating solutions for challenges of heterogeneous research data, tools, services in the arts and humanities. We highlighted that the focus of DARIAH is not simply digitized analogue material of galleries, libraries, archives, and museums. As (digital) research produces born-digital materials (e.g. datasets, tools, softwares), which have to be managed, DARIAH's collection of data is much broader. The challenges, issues and factors of the heterogeneity of (born-)digital research data that DARIAH aims to address only become apparent in large international infrastructures willing to integrate heterogeneous research practices, data formats, tools and services from the wide range of DH disciplines. This article provided insights into this process, both on European and national levels, and reflected on discussions and solutions in the broader DARIAH network.

These discussions include the many factors and challenges that influence the sharing of resources in the arts and humanities. The DARIAH research infrastructure seeks to support the scholarly community to enable and foster the work with and sharing of digital resources in numerous ways. This includes the need to look at the activities on the European and national levels and is exemplified by the three examples from member countries, showcasing also the variety in the setups of the national consortia.

In order to support communities in reusing distributed existing resources in a coherent manner, a coordinated multi-faceted strategy is paramount. It has to involve technological provisions for robust services as well as sustainable software plans, work on policy level promoting use of standards and permissive licensing, all accompanied by training and outreach activities to raise awareness and convey practical skills on digital methods.

DARIAH also acknowledges its position in the general landscape of existing initiatives, infrastructures, as well as projects, and strives to promote exchange and leverage synergies with them. In addition to the collaborations with the initiatives of the SSH communities like CESSDA, CLARIN, EUROPEANA and OpenAIRE, the cooperations with e-Infrastructures like EGI, EOSC or EUDAT are intensified and expanded.

A central goal of this pan-European endeavour is to enable, promote, and simplify the discovery and access to the wealth of (born-)digital resources available in line with the FAIR principles. In order to achieve this, DARIAH has started developing a curated community-driven discovery platform, the DARIAH Open Marketplace. Once released, it will serve the researchers and broader audiences in finding data sets, tools and services that are applicable and reusable in their daily research. The key to success is to involve the communities, and in this regard, the Marketplace has a pivotal role for the future.

References

ACDH-OeAW (2015). *Austrian Centre for Digital Humanities at the Austrian Academy of Sciences.* Retrieved from https://www.oeaw.ac.at/acdh/. Accessed 26 Feb 2018.

Alram, M., Benda, Ch., Ďurčo, M., Mörth, K., Wentker, S., Wissik, T., Budin, G., et al. (2015). DHAUSTRIA-STRATEGIE. Sieben Leitlinien für die Zukunft der digitalen Geisteswissenschaften in Österreich. Wien. https://doi.org/10.1553/DH-AUSTRIA-STRATEGIE-2015.

ARCHE (2017). *A Resource Centre for the HumanitiEs.* Retrieved from https://arche.acdh.oeaw.ac.at/. Accessed 26 Feb 2018.

Blanke, T., Bryant, M., Hedges, M., Aschenbrenner, A. & Priddy, M. (2011). Preparing DARIAH. *IEEE 7th International Conference on E-Science.* IEE Digital Library: Stockholm (pp. 158–165). https://doi.org/10.1109/eScience.2011.30.

Buddenbohm, S., Matoni, M., Schmunk, S., & Thiel, C. (2017). Quality assessment for the sustainable provision of software components and digital research infrastructures for the arts and humanities. *Bibliothek Forschung und Praxis, 41*(2), 231–241. https://doi.org/10.1515/bfp-2017-0024.

Budroni, P., Höckner, M. (2010). Phaidra, a Repository Project of the University of Vienna; in: iPRES 2010, 7th International Conference on Preservation of Digital Objects, Vienna.

CCSD (Centre pour la Communication Scientifique Directe) (2018). *A center which offers a set of services for the management of open archives.* Retrieved from https://www.ccsd.cnrs.fr. Accessed 26 Feb 2018.

CESSDA (2018). *About CESSDA.* Retrieved from https://www.cessda.eu/About>. Accessed 26 Feb 2018.

CINES (Centre Informatique National de l'Enseignement Supérieur) (2018). *Digital archiving solutions for long term preservation.* Retrieved from https://www.cines.fr/en/long-term-preservation.

CLARIN (2018). *CLARIN in a Nutshell.* Retrieved from https://www.clarin.eu/content/clarin-in-a-nutshell. Accessed 26 Feb 2018.

CORDIS (2018). *DARIAH ERIC Sustainability Refined.* Retrieved from https://cordis.europa.eu/project/rcn/207190_en.html. Accessed 26 Feb 2018.

CoreTrustSeal (2018). *CoreTrustSeal Data Repository Certification.* Retrieved from https://www.coretrustseal.org/. Accessed 26 Feb 2018.

DARIAH (2017). *2020: 25 Key Actions for a Stronger DARIAH by 2020.* Retrieved from https://www.dariah.eu/wp-content/uploads/2017/02/DARIAH_STRAPL_v06112017.pdf. Accessed 26 Feb 2018.

DARIAH (2018). *Dariah in a Nutshell.* Retrieved from https://www.dariah.eu/about/dariah-in-nutshell/. Accessed 26 Feb 2018.

DARIAH-DE (2017). *DARIAH Collection Description Data Model DCDDM.* Retrieved from https://github.com/DARIAH-DE/DCDDM. Accessed 26 Feb 2018.

DARIAH-DE (2018a). *DARIAH-DE Collection Registry.* Retrieved from https://colreg.de.dariah.eu. Accessed 26 Feb 2018.

DARIAH-DE (2018b). *DARIAH-DE: Data Modelling Environment.* Retrieved from https://dme.de.dariah.eu/dme. Accessed 26 Feb 2018.

DARIAH-DE (2018c). *DARIAH-DE Generic Search.* Retrieved from https://search.de.dariah.eu/search/. Accessed 26 Feb 2018.

DARIAH-DE (2018d). *DARIAH-DE Geo-Browser.* Retrieved from https://geobrowser.de.dariah.eu/. Accessed 26 Feb 2018.

DARIAH-DE (2018e). *DARIAH-DE Publikator.* Retrieved from https://repository.de.dariah.eu/publikator. Accessed 26 Feb 2018.

DARIAH-DE (2018f). *DARIAH-DE Repository.* Retrieved from https://de.dariah.eu/repository. Accessed 26 Feb 2018.

DARIAH-DE (2018g). *Data Federation Architecture Technical Documentation.* Retrieved from https://repository.de.dariah.eu/doc/services/. Accessed 26 Feb 2018.

DARIAH-DE (2018h). *Der DARIAH-DE Forschungsverbund.* Retrieved from https://de.dariah.eu/der-forschungsverbund>. Accessed 26 Feb 2018.

DARIAH-TEACH (2017). *dariahTeach.* Retrieved from https://teach.dariah.eu/. Accessed 26 Feb 2018.

DHA (2015). *Digital Humanities Austria.* Retrieved from http://digital-humanities.at/. Accessed 26 Feb 2018.

DH-registry (2017). *DH Course Registry.* Retrieved from https://registries.clarin-dariah.eu/courses/. Accessed 26 Feb 2018.

Dombrowski, Q. (2014). What ever happened to project bamboo? *Literary and Linguistic Computing, 29*(3), 326–339. https://doi.org/10.1093/llc/fqu026.

Doorn, P., Aerts, P. and Lusher, S. (2016). Research software at the heart of discovery, DANS & NLeSC. Retrieved from https://www.esciencecenter.nl/pdf/Software_Sustainability_DANS_NLeSC_2016.pdf. Accessed 26 Feb 2018.

Dublin Core Metadata Initiative (2007). *Dublin Core Collection Description Application Profile.* Retrieved from http://dublincore.org/groups/collections/collection-application-profile/. Accessed 26 Feb 2018.

Dublin Core Metadata Initiative (2013) Dublin Core metadata element set, version 1.1: Reference description. Retrieved from http://www.dublincore.org/documents/dces/. Accessed 26 Feb 2018.

Dumouchel, S. (2017). *How the notion of access guides the organization of a European research infrastructure: the example of DARIAH.* Retrieved from https://dh2017.adho.org/abstracts/088/088.pdf> [Last accessed 17 May 2018].

Ďurčo, M. & Mörth, K. (2014). *CLARIN-DARIAH.AT – Weaving the network*, in: 9th Language Technologies Conference. Information Society – IS 2014, Ljubljana, Slovenia, pp. 14–18.

Edmond, J. (2018 Feb) Untangling Barriers: Director Jennifer Edmond on DARIAH's Commitment to Open Science. Retrieved from https://www.dariah.eu/?p=1997. Accessed 26 Feb 2018.

EGI (2018). *EGI: advanced computing for research. Retrieved from* https://www.egi.eu/about/. Accessed 26 Feb 2018.

Engelhardt, C., Leone, C., & Moranville, Y. (2017). Distributed Metadata Schema and Demonstrator for Open Humanities Methods. [Research Report] Göttingen State and University Library; DARIAH. 2017. Available at https://hal.archives-ouvertes.fr/hal-01637051v1.

EUDAT (2018). What is EUDAT? Retrieved from https://www.eudat.eu/what-eudat. Accessed 26 Feb 2018.

European Commission (2017). *EOSC Declaration. Retrieved from* https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf. Accessed 26 Feb 2018.

European Commission (2018). *Commission Recommendation of 25.4.2018 on access to and preservation of scientific information.*Retrieved from http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51636. Accessed 26 Feb 2018.

European Roadmap for Research Infrastructures. (2006). Report 2006. Luxembourg: Office for Official Publications of the European Communities. Retrieved from https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en.pdf. Accessed 26 Feb 2018.

Fedora (2018). Fedora Repository. Retrieved from http://fedorarepository.org/. Accessed 26 Feb 2018.

Forschungsinfrastrukturen für die Geisteswissenschaften (2018). *Wissenschaftsgeleitete Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Deutschland.* Retrieved from https://www.forschungsinfrastrukturen.de/. Accessed 26 Feb 2018.

GitHub (2016). *Making Your Code Citable. Retrieved from* https://guides.github.com/activities/citable-code/. Accessed 26 Feb 2018.

Gradl, T., & Henrich, A. (2016). Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen. *Bibliothek Forschung und Praxis, 40*(2), 222–228. https://doi.org/10.1515/bfp-2016-0027.

Gradl, T., Henrich, A., & Plutte, C. (2015). Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Föderation von Kollektionen. In Baum, C. & Stäcker, T.(eds.) G*renzen und Möglichkeiten der Digital Humanities. Zeitschrift für digitale Geisteswissenschaften*, 1. DOI: https://doi.org/10.17175/sb001_020.

Harms, P., & Grabowski, J. (2011). Usability of Generic Software in e-Research Infrastructures. *Journal of the Chicago Col loquium on Digital Humanities and Computer Science, 1*(3) 1–18. http://resolver.sub.uni-goettingen.de/purl?gs-1/9238.

Harmsen, H., Kalman, T. & Wandl-Vogt, E. (2015). DARIAH meets EGI. *Inspired newsletter – Issue* 19. Retrieved from https://www.egi.eu/news-and-media/newsletters/Inspired_Issue_19/dariah.html. Accessed 26 Feb 2018.

Hettrick, S. (2016). *Research Software Sustainability: Report on a Knowledge Exchange Workshop.* Retrieved from https://www.esciencecenter.nl/pdf/Research_Software_Sustainability_Report_on_KE_Workshop_Feb_2016_FINAL.PDF>. Accessed 26 Feb 2018.

Jiménez R.C., Kuzak M., Alhamdoosh M., et al. (2017). *Four simple recommendations to encourage best practices in research software* [version 1]. *F1000Research*, 6:876. https://doi.org/10.12688/f1000research.11407.1.

Kalman, T., Thiel, C., Van Uytvanck, D., Moranville, Y. (2018). Sustainable Research Software – Managing a Common Problem of SSH Infrastructures. Digital Infrastructures for Research 2018, Lisbon, Portugal Retrieved from https://indico.egi.eu/indico/event/3973/session/22/contribution/111

Katz, D. S., Niemeyer, K. E., Smith, A. M., Anderson, W. L., Boettiger, C., Hinsen, K., & Hooft, R. (2016). Software vs. Data in the Context of Citation. *PeerJ Preprints, 4.* https://doi.org/10.7287/peerj.preprints.2630v1.

Moranville, Y., Rodzis, M. & Thiel, C. (2018). *DARIAH Technical Reference.* Retrieved from <https://dariah-eric.github.io/technical-reference/. Accessed 26 Feb 2018.

MSH Network (2018). *Réseau National des Maisons des Sciences de l'Homme.* Retrieved from http://www.msh-reseau.fr. Accessed 26 Feb 2018.

Open Editions (2018). O*pen access to comprehensive services in journal publications, books, scientific blogs and scientific events.* Retrieved from https://www.openedition.org/. Accessed 26 Feb 2018.

OPERAS (2018). *An European research infrastructure for the development of open scholarly communication, particularly in the social sciences and humanities*. Retrieved from http://operas.hypotheses.org. Accessed 26 Feb 2018.

Software Sustainability Institute **(**2018**)**. *Software Management Plans*. Retrieved from https://www.software.ac.uk/software-management-plans. Accessed 26 Feb 2018.

Stigler, J. **&** Steiner, E. **(**2014). GAMS and Cirilo client. Policies, documentation and tutorial. Retrieved from http://gams.uni-graz.at/. Accessed 26 Feb 2018.

TextGrid (2018). *Virtual Research Environment for the Humanities*. Retrieved from https://textgrid.de/en/. Accessed 26 Feb 2018.

TGIR Huma-Num **(**2018). *An infrastructure for humanities which offers a range of services dedicated to the production and reuse of data*. Retrieved from https://huma-num.fr/. Accessed 26 Feb 2018.

Thiel, C. (2017). Workshop: Software sustainability: Quality and re-usability. DHd-Blog. Retrieved from http://dhd-blog.org/?p=8685. Accessed 26 Feb 2018.

Van Uytvanck, D.**,** Zinn, C.**,** Broeder, D.**,** Wittenburg, P. **&** Gardelleni, M. (2010). *Virtual language observatory: The portal to the language resources and technology universe*. In *Seventh conference on International Language Resources and Evaluation [LREC 2010]* (pp. 900-903). Tübingen: European language resources association (ELRA).

Wandl-Vogt, E., Barbera, R., La Rocca, G., Calanducci, A. & Kalman, T.**,** (2017). Brid[g]ing the GAP: 100 Jahre Dialektlexikographie als Cloud Service. Der SADE Use Case im DARIAH Competence Centre. Elisabeth Burr. DHd 2016. Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts. 2. überarbeitete und erweiterte Ausgabe. Universität Leipzig, 7. bis 12. März 2016. Duisburg.

Wilkinson, M. D., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18.

Affiliations

Tibor Kálmán[1] · **Matej Ďurčo**[2] · **Frank Fischer**[3] · **Nicolas Larrousse**[4] · **Claudio Leone**[5] · **Karlheinz Mörth**[2] · **Carsten Thiel**[5]

Matej Ďurčo
Matej.Durco@oeaw.ac.at

Frank Fischer
ffischer@hse.ru

Nicolas Larrousse
nicolas.larrousse@huma-num.fr

Claudio Leone
leone@sub.uni-goettingen.de

Karlheinz Mörth
Karlheinz.Moerth@oeaw.ac.at

[1]   Gesellschaft für wissenschaftliche Datenverarbeitung mbh Göttingen (GWDG), Göttingen, Germany

[2]   Austrian Academy of Sciences, Vienna, Austria

[3]   National Research University Higher School of Economics, Moscow, Russian Federation

[4]   CNRS, Paris, France

[5]   Göttingen State and University Library, Göttingen, Germany