

Tanja Wissik, Leon Wessels, and Frank Fischer

# The DH Course Registry: A Piece of the Puzzle in CLARIN's Technical and Knowledge Infrastructure

**Abstract:** This chapter presents the Digital Humanities Course Registry (DHCR) as part of CLARIN's Technical and Knowledge Infrastructure. The DHCR is an initiative started to provide an overview of the growing number of training activities in the field of Digital Humanities around the world. The goal of the registry, which is run jointly with DARIAH, is to collect information on the rich offerings of different courses with the help of the community and to delineate an up-to-date picture of the teaching and training opportunities in the field. First, we will introduce the DHCR, its goals, genesis, and main features. Then we will elaborate on the DHCR's position within CLARIN's Technical Infrastructure and how it helps to address CLARIN's agenda and strategic goals in terms of Technical Infrastructure, Open Science, and especially the FAIR Principles. Particular attention will be paid to the results of a cross-national hackathon using data and metadata from the DHCR. Furthermore, we will examine the position of the DHCR within CLARIN's Knowledge Infrastructure, which includes training and education.

**Keywords:** Training and Education, DH Course Registry, community-driven platform, API, Knowledge Infrastructure

## 1 Introduction to the DHCR

It is part of the grassroots history of the Digital Humanities (DH) that the first courses, workshops, and hackathons had to be organized outside established academic teaching formats because there was simply no place for them in the curriculum of higher education yet. Since information on offered courses was scattered across the internet, it soon became difficult to keep sight of the overall picture.

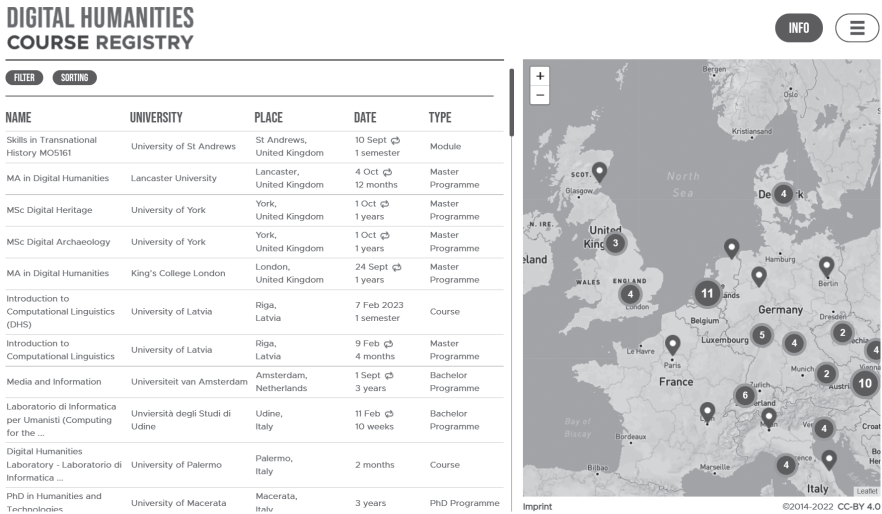
---

**Tanja Wissik**, Austrian Academy of Sciences, Vienna, Austria, e-mail: tanja.wissik@oeaw.ac.at

**Leon Wessels**, CLARIN ERIC and Utrecht University, Utrecht, the Netherlands,  
e-mail: leon@clarin.eu

**Frank Fischer**, Freie Universität Berlin, EXC 2020 Temporal Communities, Berlin, Germany,  
e-mail: fr.fischer@fu-berlin.de

To increase visibility, the Digital Humanities Course Registry (DHCR), originally a German initiative, was founded. As early as 2011, Manfred Thaller and Patrick Sahle published a list of courses related to digital methods in the Humanities (Sahle, Puhle, and Rau 2011). Gradually, this community-driven information collection has been internationalized and supplemented by an interactive map of Europe (and eventually, since 2018, a world map) showing the various locations of institutions offering DH-relevant courses (Figure 1).



**Figure 1:** DH Course Registry list and map view.

The revamp of the frontend in 2019 also included a technical change, the introduction of the API (Application Programming Interface), which is the prerequisite for a truly community-driven initiative and for versatile access to the data.

Many DH courses are now firmly established at universities and other institutions (for specific examples of courses and training events from South Africa and Lithuania, see Hennelly et al. (2022) and Petrauskaitė et al. (2022)). The formats covered include Bachelor, Master, and PhD courses with different focuses, as well as summer schools and workshop series. Courses can be held in-person or online and have to be recurring to be featured in the registry. We are especially targeting the following groups with the DHCR (cf. Wissik, Edmond, Fischer, et al. 2020):

- students (who want to join a university programme in Digital Humanities or related fields or want to find an opportunity for student exchange);
- lecturers (who wish to promote their teaching activities);

- programme administrators (who wish to promote and facilitate student and staff exchanges);
- researchers (who are interested in the proliferation of DH education);
- decision makers (who need quantitative evidence to drive funding for DH activities).

Since 2016, the DHCR has been jointly operated by the two infrastructures CLARIN and DARIAH. This solution guarantees the necessary stability, as it is a big challenge to keep a globally active and growing community-driven initiative like this alive, both technically and socially. This cannot be mastered by temporary research or infrastructure projects with a national or regional scope. Data curation within the DHCR follows a community-driven approach. The DARIAH working group “DH Course Registry” plays a key role,<sup>1</sup> taking care of the coordination, user administration and technical maintenance of the registry. Lecturers can upload their own courses, but there is also a group of national moderators mainly responsible for content maintenance and consistency; they are joined by volunteers from the CLARIN and DARIAH communities. For each country represented in the course registry, a national moderator is appointed (sometimes more than one for bigger countries) to monitor, curate, and update the database entries. To support this, the system also sends notifications – for example, when a new course has been submitted or when information is out of date – whereupon the owner of a record (the course maintainer) can be contacted to update an entry. Course metadata is collected in English. The TaDiRAH taxonomy, another DH community-driven project, is also used (Borek et al. 2016). Its integration makes it possible to search the course data for specific activities, objects, and techniques. The proof of the relevance and success of the course register is, of course, in its use and up-to-dateness. As of April 2021, the DH Course Registry contains 234 active courses and programmes in 29 countries. The collaborative collection of information is valuable both for individuals seeking to find or promote DH training opportunities and for those seeking to understand the evolution of DH over time and on an international scale. The rich data contained in the DHCR is now also explorable through an open API, a facility that makes the inherent knowledge of the database more accessible, as the following chapters will show. The remainder of this chapter is structured as follows: in Section 2 we will elaborate on the position of the DHCR within CLARIN’s Technical Infrastructure and on how the DHCR addresses CLARIN’s strategic goals, again in terms of Technical Infrastructure. In Section 3 we will discuss how the DHCR aligns with CLARIN’s Knowledge Infrastructure.

---

<sup>1</sup> See <https://www.dariah.eu/activities/working-groups/dh-course-registry/>

## 2 DHCR as part of CLARIN’s Technical Infrastructure

Since 2016 the DH Course Registry, as a joint initiative from the research infrastructures CLARIN and DARIAH, has been part of the CLARIN ecosystem and Technical Infrastructure. CLARIN’s mission is to create and maintain an infrastructure to support the sharing, use, and sustainable availability of language data and tools for research in the humanities and social sciences (SSH) (cf. de Jong et al. 2020). Over the last few years, CLARIN has fulfilled its mission by creating “a sound and robust technical basis to enable the sharing and reuse of language data and tools across institutional, disciplinary and international borders” (CLARIN ERIC Strategy 2021–2023). Part of the CLARIN strategic goals in terms of Technical Infrastructure are the FAIR Principles (findability, accessibility, interoperability, and reusability) and the initiative “CLARIN for Programmers”. In this section we will elaborate on the position of the DHCR within CLARIN’s Technical Infrastructure and on how the DHCR addresses CLARIN’s strategic goals in terms of technical infrastructure. Furthermore, we will discuss how various initiatives (e.g., provision of an API and organization of hackathons) contribute to the fact that the DHCR is not only a community-driven initiative when it comes to data collection, as described in Section 1, but also when it comes to data use and functionality improvement.

### 2.1 Open Science and FAIR Principles

One of the strategic goals of CLARIN in terms of Technical Infrastructure is to support the Open Science and FAIR Principles agenda. CLARIN has taken a leading role in the Open Science and FAIR agendas (cf. Rossi et al. 2020) as an early adopter of the FAIR Principles (de Jong et al. 2018). For example, all the design decisions of all data services are guided by the FAIR Principles (CLARIN ERIC Strategy 2021–2023, de Jong et al. 2018). Furthermore, CLARIN has a well-defined access policy and policies for data protection, as well as a template for Terms of Service (cf. Rossi et al. 2020).

Regarding two of the principles – interoperability and reusability – APIs can play an important role. APIs are services that allow direct and structured access to data, without having to download entire data sets. In Section 2.3 we describe the API of the DH Course Registry and illustrate how it contributes to the Open Science and FAIR agenda and the CLARIN ERIC Strategy.

## 2.2 “CLARIN for Programmers”

Another strategic goal in terms of Technical Infrastructure is to disseminate Natural Language Processing services more prominently to programming scientists, for example, with well-documented application programming interfaces and example snippets in popular development environments (CLARIN ERIC Strategy 2021–2023). In order to reach this goal, the initiative “CLARIN for Programmers” was created. As a study by Edmond and Garnett (2015) found that researchers “only cared about the data, and had no specific opinions about how that data was accessed and no particular need for some of the special functionality an API could offer”, it makes sense to promote APIs especially to programmers or programming scientists. We will show in Sections 2.4 and 2.5 how, for example, the ACDH-CH virtual hackathon series – a CLARIAH-AT initiative<sup>2</sup> – and its outcomes helped to support the CLARIN ERIC Strategy in relation to the Open Data agenda (Hanneschläger and Wissik 2020) and in terms of “CLARIN for Programmers”. In addition, it helped to improve the infrastructure.

## 2.3 DHCR API

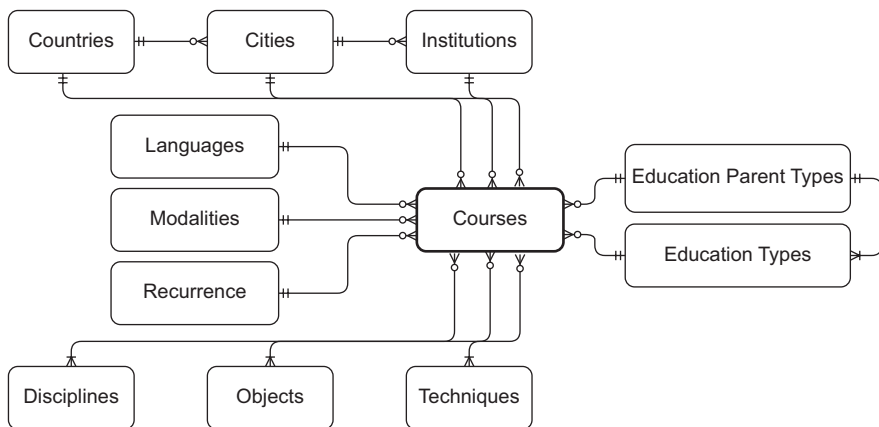
Until 2019 the course data in the DH Course Registry – e.g., name of the course, name of the institution offering the course (see also Figure 2 for the data model) – was only accessible via the search interface and it was not possible to download the data. In addition, only recent data was visible on the platform; the historical records were hidden in the backend and not publicly accessible. The solution to this problem was the implementation of an Application Programming Interface (API) in the framework of the DH Course Registry Sustain Project.<sup>3</sup> As noted by Tasovac et al. (2016), APIs have the potential to be “powerful, practical building blocks of digital humanities infrastructures. On the technical level, they let heterogeneous agents dynamically access and reuse the same sets of data and standardized workflows. On the social level, they help overcome the problem of ‘shy

---

<sup>2</sup> CLARIAH-AT is the national counterpart of CLARIN and DARIAH in Austria. See <https://digital-humanities.at/en/dha/clariah-at>

<sup>3</sup> Within the DHCR Sustain Project (<https://www.oeaw.ac.at/acdh/projects/dhcr-sustain/>), funded within the DARIAH Theme funding call 2018/2019, the API and its documentation was improved. The general technical development work and maintenance was funded by PAR-THENOS (H2020 Grant Agreement n. 654119) and CLARIN-ERIC. The API can be accessed here <https://dhcr.clarin-dariah.eu/api/v1/> and the documentation is available here <https://app.swaggerhub.com/apis/hashmich/DHCR-API/1.2>

data’, i.e., data you can ‘meet in public places but you can’t take home with you’ (Cooper 2010).” The DH Course Registry provides a public JSON data API for data export, custom analysis visualizations, and so on. Since one of the main purposes is the export of data, the following explains the data model behind the DH Course Registry, which revolves around courses as key entities (Figure 2). All metadata related to the courses can be grouped into two types of metadata: metadata related to the course content (education type, education parent type, language, modality, recurrence, disciplines, objects, and techniques) and metadata related to the provider of the courses (institutions, cities, countries). For education types, we have elaborated our own classification: Bachelor Programme, Master Programme, Research Master, and PhD Programme as whole degree programmes offered at higher education institutions; modules and courses as part of degree programmes; and summer schools and continuous education as education activities outside of degree programmes (cf. Wissik, Edmond, Fischer, et al. 2020). For modalities, we have online or on-site training activities and for recurrence, there is the choice between training activities that occur once (e.g., a specific course that is only offered for one semester) or training activities that are recurring, such as Bachelor Programmes.



**Figure 2:** DH Course Registry Data Model (Adapted from Wissik, Edmond, Fischer, et al. 2020).

The entities, objects, and techniques come from the TaDiRAH Taxonomy (Borek et al. 2016) and the disciplines are based on the disciplinary categorization as applied by the Dutch Scientific Council for Academic Research (NWO) or NARCIS (Safradin and de Jong 2017), respectively, but have been modified and enriched based on the needs of the growing DH Course Registry (cf. Wissik, Edmond, Fischer, et al. 2020). Figure 2 shows the data model of the DH Course Registry.

## 2.4 ACDH-CH Virtual Hackathon Series

The Open Data movement is not only gaining momentum in the context of the Digital Humanities, but also in other research areas. As we saw in Section 2.1, CLARIN ERIC supports the Open Science agenda and FAIR Principles, and the same can be said for the Austrian manifestation of CLARIN and DARIAH, CLARIAH-AT. One of the fundamental concepts and principles of Open Science is Open Data. Therefore, in early 2019, CLARIAH-AT funded an initiative launched by the Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) of the Austrian Academy of Sciences: a virtual hackathon series to promote Open data. These virtual hackathons focused on different Open Data sets that are publicly available online, and the tasks to be performed on these data included the creation of open-source code. Typically, hackathons take place on-site, where participants have to solve tasks within a short period of time. This requires programmers to be flexible, available, and willing to travel. A virtual hackathon, on the other hand, offers people around the globe the opportunity to participate and contribute without having to travel. Moreover, by setting a longer timeframe (in this case two to four weeks, depending on the hack), people with fixed time schedules could also participate. Therefore, our approach enabled a much larger and more diverse community to participate while also promoting the benefits of Open Data (cf. Hanneschläger and Wissik 2020). In the virtual hackathon series, the ACDH-CH organized four hacks over the course of 2019 and 2020. In 2020 the chosen data set was the data and metadata from the DH Course Registry, not only to promote Open Data but also to point programming scientists in the direction of the API. The task for the DH Course Registry hack was to develop a creative way of visualizing data and metadata about teaching activities. It was possible to visualize the data itself or the results of statistical analysis with the data. Participants were free to choose the visualization method, except for the map-based visualization, as this visualization is already implemented in the official DH Course Registry.<sup>4</sup> The best contributions were selected by an international jury. The judging criteria were creativity and innovation, accessibility, reusability, and reproducibility, as well as elegance (cf. Hanneschläger and Wissik 2020). The winners received cash prizes and were published on GitHub<sup>5</sup> and shared via Twitter. In the following, we will present the outcomes of the ACDH-CH Hackathon and discuss how they contributed to the improvement of the DHCR.

---

<sup>4</sup> <https://github.com/acdh-oeaw/ACDHchHackathon2020/blob/master/README.md>

<sup>5</sup> <https://github.com/acdh-oeaw/ACDHchHackathon2020/blob/master/results.md>

## 2.5 Outcomes of the ACDH-CH Hackathon

Since for the virtual hackathon task there were no strict requirements, it is not surprising that the winning projects<sup>6</sup> took very different approaches. The Notlef Project by Philip Allfrey was the only submission that worked only with the data provided via the DH Course Registry API. The DH Education Knowledge Map project by Marta Palandri and Raphael Mitsch enriched the DH Course Registry data with Wikipedia data. The ACDH-2020 project by Francesca Giovannetti, Ivan Heibi, and Bruno Sartini used the DH Course Registry data as starting point and enriched it with DH-related publication data from Crossref and Microsoft Academic. And the CORIANDER (Course RegIstry sTatistics aNd aDditional matERial) project by Martina Trognitz and Lukas Gehrig enriched the DH Course Registry data with data from a Zotero bibliography, which also used the TaDiRAH taxonomy. In the following, we will describe the ACDH-CH Hackathon outcomes and the winning projects in more detail.

The CORIANDER project (Trognitz and Gehrig 2020) added further functionalities to the DH Course Registry in order to visually explore the metadata categories disciplines, TaDiRAH objects, and TaDiRAH techniques further; in the original platform these are only used as filter options. It is browser-based, using HTML, CSS, and JavaScript as well as common JavaScript libraries for visualization. Python3 is used for data processing. The prototype application<sup>7</sup> has two main visualization modes for the data, organized by courses or keywords. In the course view, the courses can be explored by keywords (i.e., discipline, TaDiRAH objects and TaDiRAH techniques, countries and years) which are then visualized in a bar chart. For each individual course, additional literature (from Zotero and Wikidata) is accessed by clicking on the respective course. In the keyword mode, the co-occurrence of keywords can be explored (see Figure 3).

In the DH Education Knowledge Map project (Palandri and Mitsch 2020), the DH Course Registry Data was not seen as a “set amount of information but a starting point for the creation of a web of knowledge that can help us make unexpected connections” (Palandri 2020). For this purpose, a layer of wikification was added on top of the given data set via Wikipedia’s API. The application was developed with Dash, a Python framework for building web analytic applications. The DH Education Knowledge map is divided into four panels, presenting the courses as a table and a scatterplot, and the Wikipedia information in the form of a graph with an accompanying paragraph from Wiki about the selected node from the

---

<sup>6</sup> All the winning projects, their description and their evaluation can be found on the following GitHub page <https://github.com/acdh-oeaw/ACDHchHackathon2020/blob/master/results.md>

<sup>7</sup> The prototype application can be accessed here <https://bellerophons-pegasus.github.io/CORIANDER/index.html>



## CORIANDER

Course Registry Statistics and Additional Material



Explore how keywords co-occur with each other in the chord diagram. Select keywords and click on 'Redraw with Selection' for customisation. The slider sets the number of most frequent connections to display. A fading arc indicates the connection is not in the counterpart's top x. Click on Courses to explore those. To know more go to GitHub for the full README.

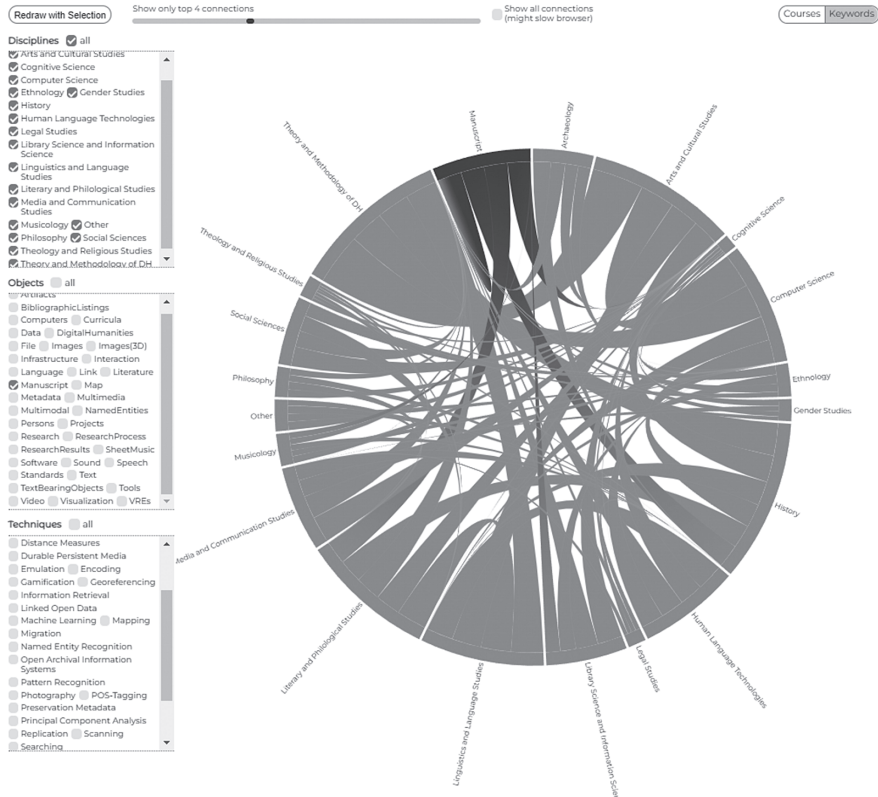


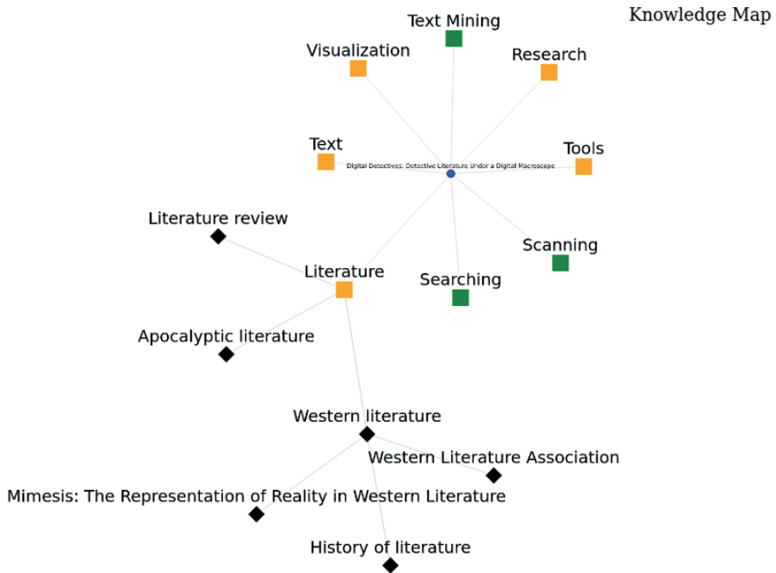
Figure 3: CORIANDER keywords view.

graph. The application connects the original data with Wikipedia data, allowing concept-based navigation and connections to related concepts (see Figure 4).

The Notlef Project (Allfrey 2020) is the only project that did not use additional external data resources to enrich the given DH Course Registry data set. The design for the Notlef visualization, as stated by the developer, was inspired by the 2009 Feltron Annual Report,<sup>8</sup> written by Nicholas Felton and implemented as an Angular 9 app. The Notlef app<sup>9</sup> consists of two pages – courses and places (see Figure 5) –

<sup>8</sup> <http://feltron.com/info.html>

<sup>9</sup> The app can be accessed here: <https://notlef-dhcr.web.app/overview>



#### Western literature

Western literature, also known as European literature, is the literature written in the context of Western culture in the languages of Europe, including the ones belonging to the Indo-European language family as well as several geographically or historically related languages such as Basque and Hungarian. Western literature is considered one of the defining elements of Western civilization. The best of Western literature is considered to be the Western canon. The list of works in the Western canon varies according to the critic's opinions on Western culture and the relative importance of its defining characteristics. Western literature includes written works in many languages:

**Figure 4:** Knowledge Map view (Palandri 2020).

wherein statistics of the DHCR data are visualized. A set of data (e.g., countries) is highlighted in colour and, by clicking on one of these values, the other data on that page will be filtered (e.g., to show information for a particular country).

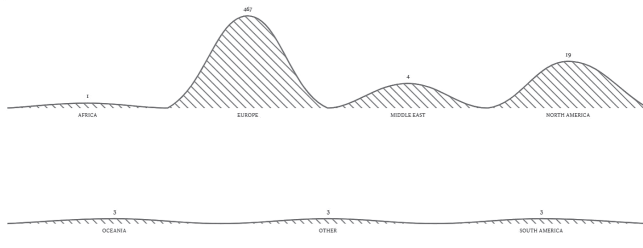
The goal of the ACDH-2020 project (Giovannetti, Heibi, and Sartini 2020) was to investigate which of the techniques taught by the different DH courses are most often applied in relevant publications and the collaborations in terms of academic publications between institutions offering these DH courses. For this purpose, the DH Course Registry data was enriched with DH related external data sets such as Microsoft Academics and Crossref. The jQuery base application<sup>10</sup> contains two different types of visualization: bar charts and networks (Figure 6). In the network visualization, the institutions are the nodes, and by clicking on the nodes the user can access information about the courses taught by these institutions and the publications associated with these institutions.

<sup>10</sup> The app can be accessed here: <https://br0ast.github.io/ACDH-2020/>

# Places

Countries, Cities, and Institutions

FIGURE 1 COURSES BY REGION



INSTITUTIONS WITH THE MOST COURSES

- University of Paris (7)
- Universitat Guel (6)
- Universiteit van Amsterdam (19)
- Universität Wien (6)
- Universität zu Köln (19)
- Université de Lausanne (19)
- Universität Utrecht (10)
- University College London (10)
- Universität Bonn (10)
- Universität von Lüneburg (6)
- Universität Basel (6)
- King's College London (6)
- Higher School of Economics (10)
- Johannes-Messias-Universität Würzburg (10)
- Friedrich-Alexander-Universität Erlangen-Nürnberg (10)
- Universität Zürich (10)
- Aalborg University (10)
- UNED | Universidad Nacional de Educación a Distancia (10)
- Paris Université de Berlin (6)
- Rijksuniversiteit Groningen (6)

COUNTRIES

**31**

ARGENTINA, AUSTRALIA, AUSTRIA, BELGIUM, BRASIL, CROATIA, CZECH REPUBLIC, DENMARK, ESTONIA, FINLAND, FRANCE, GERMANY, GREECE, HUNGARY, IRELAND, ISRAEL, ITALY, LATVIA, LITHUANIA, NETHERLANDS, INDIA, POLAND, PORTUGAL, RUSSIAN FEDERATION, SLOVENIA, SOUTH AFRICA, SPAIN, SWEDEN, SWITZERLAND, UNITED KINGDOM, UNITED STATES OF AMERICA.

NUMBER OF INSTITUTIONS

**166**

INSTITUTION WITH THE MOST COURSES

**University of Tartu**

CITY WITH THE MOST COURSES

**London**

19 COURSES

MOST INSTITUTIONS PER CITY

**Eight**

PARIS

CITIES

**142**

Figure 5: Places view mode in the Notlef app.

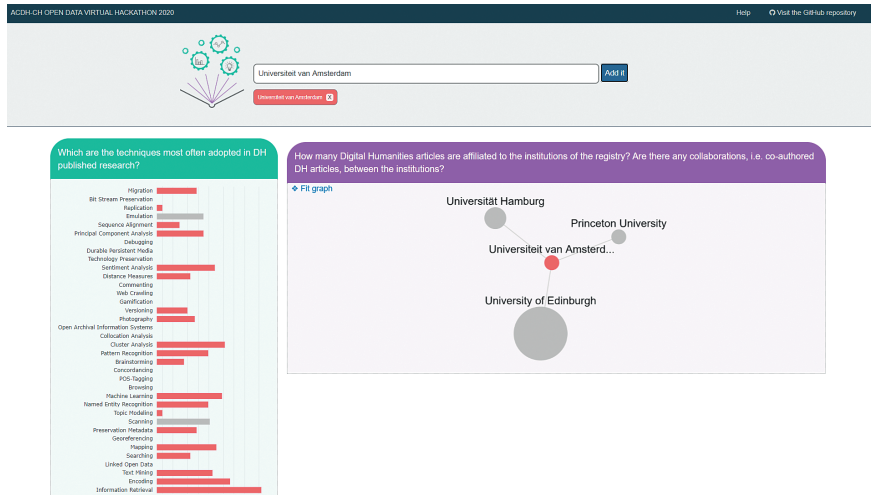


Figure 6: Network visualization publication collaboration for the search term “University of Amsterdam” in the ACDH-CH app.

The winning hackathon projects are a demonstration of what the community can do with the collected DHCR data if it is easily available through an API. The creators of these projects not only presented inspiring examples of how the data could be visualized and enriched and which questions could be answered with the data and visualization, but also addressed issues with the DHCR API

that would otherwise not have been discovered and thus fixed. For example, the creators of the CORIANDER project raised the issue that the distinction between current and historical data was not obvious. Therefore, after the hackathon the DHCR API was adapted and users can now view a list of current, maintained courses (as visible on the DHCR map) or a list of all courses including the historical ones, depending on the use case.

### **3 The DHCR as part of CLARIN’s Knowledge Infrastructure**

In this section we will discuss how the DHCR aligns with CLARIN’s Knowledge Infrastructure. The explicit and implicit aims of the DHCR will be compared to the strategic objectives of the overall Knowledge Infrastructure as expressed in several strategic documents. Over the years the DHCR has become an essential instrument of CLARIN’s Knowledge Infrastructure. The elements of its contribution can be divided into gathering and disseminating information, reaching out to user communities, and providing a forum to exchange thoughts and discuss best practices. First, however, we will briefly touch upon the need for a research infrastructure like CLARIN to have a knowledge infrastructure.

#### **3.1 Data literacy in the 21st century**

In recent decades Europe has fundamentally and irreversibly changed. The digital era has arrived and it is affecting the economy, governments, academia, and society at large. Language technology is an integral part of this digital transition. Researchers use language resources and tools to address a diverse range of research questions. Governments and industry apply text-mining algorithms to find valuable patterns in large amounts of language data and to discriminate between valid information and “fake news”. Citizens use applications like automatic speech recognition, machine translation, and autocomplete on a daily basis (CLARIN ERIC Strategy 2021–2023). In addition to a myriad of opportunities, however, the digital era comes with a number of challenges. Data and algorithms can and have been used as “weapons of mass destruction”, leading to, for instance, unwanted gender- and ethnicity-based discrimination and injustice (O’Neil 2017). Overcoming these difficulties requires increased levels of computational know-how.

The advance of the data economy has led to a growing need to train data professionals and to increase data literacy among citizens. Data literacy here is used according to Prado and Marzal's definition: the ability to "access, interpret, critically assess, manage, handle and ethically use data" (Prado and Marzal 2013: 126). The European Commission has acknowledged this view in its European Strategy for Data. According to this strategy, data analysis is among the most critical of skills shortages, resulting in about 500,000 unfilled positions in the EU. The EC believes that, if Europe's shortage in data professionals and lack of data literacy is not properly addressed, it will "affect the EU's capacity to master the challenges of the data economy and society" (European Commission 2020).

The digital era is affecting the academic world as well. Data-driven research methods allow researchers to address research questions that were previously considered too big to answer. Numerous authors have called upon linguists, historians, literary scientists, and other humanities scholars to adopt these new research methods (Borgman 2009; Guldi and Armitage 2014: 88–116; McGillivray et al. 2020). The emerging field of humanities scholars that have applied data-driven methods in their research and teaching is commonly known as the Digital Humanities. However, as stated in the Digital Humanities Manifesto, the Digital Humanities "is not a unified field but an array of convergent practices" (Schnapp et al. 2009). Some humanities disciplines have embraced the digital turn with more enthusiasm than others. Jensen and McGillivray, for example, note that corpus-based quantitative methods for historical linguistics have not gone mainstream. According to their diagnosis, there is a "chasm" dividing the early adopters from the majority of users. Jensen and McGillivray suggest this chasm is caused by, among other factors, a lack of training opportunities and educational practices (Jensen and McGillivray 2017: 22–25).

European Research Infrastructure Consortia (ERICs) such as CLARIN play an important role in increasing data literacy and advocating responsible data use among new generations of researchers, data professionals, and more generally among citizens. As such, ERICs enable researchers to apply data-driven methods, support EU countries to improve their position in the global economy, and to empower citizens to safely and efficiently interact with everyday technology (ERIC Forum 2020). Within CLARIN, increasing data literacy is one of the main objectives of what is called the Knowledge Infrastructure. The Knowledge Infrastructure serves as the "glue" for the various communities engaged with CLARIN and is the structure that aims to secure a continuous transfer of knowledge (CLARIN ERIC Strategy 2021–2023). It encapsulates a range of initiatives, including the DHCR. In what follows, we will outline three branches of activities that are part of the Knowledge Infrastructure and highlight the strategic role of the DHCR.

### 3.2 Disseminating information

CLARIN is a distributed research infrastructure. It consists of dozens of nodes spread across Europe and to a growing extent across the world. Making sure that researchers, teachers, citizen-scientists, commercial partners, policy makers, journalists, and other players involved can find the information they need, is one of the key objectives of the Knowledge Infrastructure. To achieve this goal, CLARIN has established a network of Knowledge Centres (abbreviated as K-centres). These K-centres offer information services, such as a help desk, online courses, best-practice documents, and guidance in gaining access to and using data and tools. Each K-centre has its own field of expertise, which could be an individual language, a type of language modality, a linguistic topic, a form of language processing, a type of data, or a generic method or issue.<sup>11</sup>

Much like CLARIN, the Digital Humanities are also characterized by their distributed nature. Researchers and teachers who apply and critically reflect on DH methods in their research and teaching, can be found in many humanities departments across the world. Though there are common platforms for them to exchange information – notably DH conferences organized at the national, supranational, and global level – information provided about DH courses is mostly limited to institutional catalogues. The DHCR offers an online platform with structured basic information about DH courses taught in and beyond Europe and with links to more detailed information. One of its strengths is that the DHCR centralizes information about a distributed field, while keeping its diverse nature intact. Consequently, the DHCR is specifically mentioned in the CLARIN ERIC Strategy 2021–2023 as an information platform for training opportunities (CLARIN ERIC Strategy 2021–2023).

### 3.3 Reaching out

Gathering and publishing information, however, is not enough. One cannot expect users to find the information they are looking for without actively promoting the available overviews. To increase awareness about the resources, tools, and services offered by CLARIN, a number of instruments are in place to actively reach out to existing and new communities of use. Both CLARIN ERIC and national CLARIN consortia organize User Involvement events to inform potential users about how they might benefit from the CLARIN infrastructure. In addition,

---

<sup>11</sup> See <https://www.clarin.eu/content/knowledge-centres>

there is a programme of CLARIN Ambassadors, who present CLARIN at scientific conferences and events.<sup>12</sup>

Another way of reaching out is to integrate CLARIN in existing structures. In 2019 CLARIN's Knowledge Infrastructure Committee organized a workshop on teaching CLARIN at universities. Prior to the workshop, participants were asked to complete a survey on the integration of CLARIN in university curricula. The results show that out of 22 participating CLARIN member and observer countries, CLARIN was integrated into courses taught at 31 different universities in 20 different countries. These courses are attended by nearly 6,000 students each year (Fišer and Krauwer 2019). To further encourage lecturers to integrate CLARIN within their courses, CLARIN has been involved in the development of teaching and training material. Currently these and other materials are being brought together and put in the spotlight through an initiative called the CLARIN Training Suite. This Training Suite will be enriched with material resulting from a continuous call for contributions.<sup>13</sup>

The DHCR is a virtual platform connecting students, teachers, and policy makers on a global scale and allowing them to exchange valuable information. Since Marshall McLuhan coined the term “global village” in the early 1960s (McLuhan 1962), the world has become ever more interconnected. As a result, spending a semester or more abroad, previously considered a luxury reserved for the upper class, has become increasingly mainstream. Though the COVID-19 pandemic temporarily limited opportunities for student exchanges, the overall trend shows that over the last few decades the number of internationally mobile students has been steadily growing (UNESCO Science Report 2015). Countless students spend part of their education abroad, selecting courses that best fit their needs and learning more about the world's cultural diversity along the way. The DHCR displays how the various areas of expertise within the Digital Humanities are distributed over Europe and to an increasing extent over the world, allowing those interested in the intersection of culture and the digital to find the courses they seek.

To make the DHCR's target groups aware of the platform, the DHCR has been presented at various DH conferences worldwide and in articles published in scientific journals (Wessels, Gheldof, and Wissik 2019; Wissik, Edmond, Fischer, et al. 2020; Wissik, Schmeer, Fischer, et al. 2020). On a more informal level, promotional material, including professionally designed postcards, has been distributed at numerous events. More recently the outreach strategy has been explicated in a dissemination plan. As part of this plan, among other initiatives, the DHCR has

---

<sup>12</sup> See <https://www.clarin.eu/content/clarin-ambassadors-programme>

<sup>13</sup> See <https://www.clarin.eu/content/call-contributions-clarin-training-suite>

started its own Instagram account in September 2020.<sup>14</sup> Through this social platform the DHCR will be promoted among target groups that do not regularly attend conferences and other formal academic events, in particular students (Woldrich, Strnadl, and Wissik 2020). Those interested can also sign up to receive updates related to newly published DH courses (e.g., in a particular country).<sup>15</sup>

### 3.4 Providing a forum

The final branch of activities discussed here is to provide a forum for researchers, teachers, and other users to exchange thoughts and discuss best practices. CLARIN facilitates such a forum by organizing online and physical events and making funding available through calls for proposals for workshops. As part of these calls, among others a series of workshops on oral history have been organized. Researchers interested in oral history archives and speech technology specialists gathered to discuss the challenges of using digitized oral history records. These workshops resulted in the development of a “chain” of tools to make oral history archives machine actionable.<sup>16</sup> Another instrument to connect users is the Mobility Grant. It is designed to fund short-term stays at a CLARIN node, used for sharing technical expertise and strengthen collaborations.<sup>17</sup>

As of 2020 the CLARIN Annual Conference has a dedicated session for teachers to present and discuss how they have integrated CLARIN in their courses. This session is titled CLARIN in the Classroom. During the CLARIN Annual Conference 2020, 11 showcases were presented, demonstrating the use of specific corpora, tools, and services. These showcases display the breadth and depth of CLARIN’s integration in the university curricula. In addition to these showcases, the UPSKILLS project was presented, which had just been accepted for funding through the Erasmus+ programme. This project aims to better prepare linguistics and language students pursuing a career in research or industry, by identifying and tackling the gaps in digital skills taught at universities. As one of the project partners, it is part of CLARIN’s role to integrate research infrastructures into teaching.<sup>18</sup>

The DHCR is a community-driven initiative. The national moderators involved are not just responsible for curating courses uploaded in the DHCR, but also func-

---

<sup>14</sup> See <https://www.instagram.com/dhcourseregistry/>

<sup>15</sup> See <https://dhcr.clarin-dariah.eu/subscriptions/add>

<sup>16</sup> See <https://oralhistory.eu/>

<sup>17</sup> See <https://www.clarin.eu/content/clarin-mobility-grants>

<sup>18</sup> See <https://www.clarin.eu/content/factsheet-clarin-upskills>



tion as a sounding board for the DHCR’s central management. By default, each national moderator is invited to participate in meetings of the DHCR Working Group. This Working Group is part of the DARIAH infrastructure. It assembles at least once a year to discuss progress and future improvements of the DHCR. As such, the DHCR Working Group functions as a forum to discuss initiatives of research infrastructures related to teaching and training.

## 4 Conclusion

In this chapter we have presented the DHCR as part of CLARIN’s ecosystem. We have introduced the DHCR, its genesis, main features, and goals, and have elaborated on the DHCR’s position within CLARIN’s Technical and Knowledge Infrastructures. We have shown that the DHCR is a true community-driven initiative, both on side of the data providers, that is, lecturers and programme directors who submit the course metadata, and the national moderators, who curate the data and on side of the data users. With the implementation of the DHCR API, the registry has opened up its treasure trove of data and contributed to CLARIN’s ecosystem in terms of Open Science and FAIR Principles and in terms of “CLARIN for Programmers”. The several dissemination activities such as virtual hackathons, screencasts, and social media campaigns have helped to reach out to the different user communities. In addition, the events (co-)organized by CLARIN and/or the DH Course Registry and the DHCR Working Group, in which the DHCR is embedded, provide a forum to discuss research infrastructure initiatives related to DH teaching and training, since education and training are important for research infrastructures to train the future generations of researchers in data literacy in the digital era.

## Bibliography

- Allfrey, Philip. 2020. *Notlef*. GitHub Repository. <https://github.com/philipallfrey/notlef> (accessed 14 April 2021).
- Borek, Luise, Quinn Dombrowski, Jody Perkins & Christof Schöch. 2016. TaDiRAH: A case study in pragmatic classification. *Digital Humanities Quarterly* 10 (1). <http://www.digitalhumanities.org/dhq/vol/10/1/000235/000235.html> (accessed 14 April 2021).
- Borgman, Christine L. 2009. The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly* 3 (4). <http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html> (accessed 14 April 2021).

- Cimiano, Philipp, Chiarcos, Christian, McCrae, John P. & Jorge Gracia. 2020. *Linguistic linked data: Representation, generation and applications*. Berlin: Springer.
- CLARIN ERIC Strategy 2021–2023. CLARIN ERIC <https://www.clarin.eu/content/vision-and-strategy> (accessed 14 April 2021).
- Common Language Resources and Technology Infrastructure. Strategy 2021–2023 at a glance. <https://office.clarin.eu/v/CE-2020-1709-CLARIN-Strategy-Summary-2021-2023-two-pager.pdf> (accessed 14 April 2021).
- Cooper, D. (2010). When nice people won't share: Shy data, web APIs, and beyond, *Second International Conference on Global Interoperability for Language Resources (ICGL 2010)*, n.p.
- Dissemination Strategy (1/2). *Reaching out: Dissemination strategy and planning for the DH Course Registry Project*. Internal report.
- Edmond, Jennifer, Frank Fischer, Laurent Romary & Toma Tasovac. 2020. Springing the floor for a different kind of dance: Building DARIAH as a twenty-first-century research infrastructure for the arts and humanities. In Jennifer Edmond (ed.), *Digital technology and the practices of humanities research*, 207–234. Cambridge: Open Book Publishers. <https://doi.org/10.11647/OBP.0192.09>.
- Edmond, Jennifer & Vicky Garnett. 2015. APIs and researchers: The emperor's new clothes? *International Journal of Digital Curation* 10 (1). 287–297.
- ERIC Forum. 2020. The ERIC Community and Horizon Europe mission areas, June 2020. [https://www.eric-forum.eu/wp-content/uploads/2020/06/ERIC-Forum\\_Horizon-Europe-Missions\\_Position-Paper.pdf](https://www.eric-forum.eu/wp-content/uploads/2020/06/ERIC-Forum_Horizon-Europe-Missions_Position-Paper.pdf) (accessed 14 April 2021).
- European Commission. 2020. A European strategy for data, COM/2020/66 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066> (accessed 14 April 2021).
- Fišer, Darja & Steven Krauwer. 2019. *Report on the survey for the CLARIN@Universities Workshop*. Internal report.
- Giovanetti, Francesca, Ivan Heibi & Bruno Sartini. (2020). ACDH-2020. GitHub Repository. <https://github.com/br0ast/ACDH-2020> (accessed 14 April 2021).
- Guldi, Jo & David Armitage. 2014. *The history manifesto*. Cambridge: Cambridge University Press.
- Hanneschläger, Vanessa & Tanja Wissik. 2020. *Opening up open data: Strategies and success stories*. Abstracts of the DH2020. [https://dh2020.adho.org/wp-content/uploads/2020/07/112\\_OpeningupOpenDataStrategiessuccessstories.html](https://dh2020.adho.org/wp-content/uploads/2020/07/112_OpeningupOpenDataStrategiessuccessstories.html) (accessed 14 April 2021).
- Hennelly, Martin, Langa Khumalo, Juan Steyn & Menno van Zaanen. 2022. Training of digital language resources skills in South Africa. In Darja Fišer & Andreas Witt (eds.), *CLARIN. The infrastructure for language resources*. Berlin: deGruyter.
- Jenset, Gard B. & Barbara McGillivray. 2017. *Quantitative historical linguistics: A corpus framework*. Oxford: Oxford University Press.
- Jong, Franciska de, Bente Maegaard, Koenraad De Smedt, Darja Fišer & Dieter Van Uytvanck. 2018. CLARIN: Towards FAIR and responsible data science using language resources. *Proceedings of the International Conference on Language Resources and Evaluation (LREC)* 11. 3259–3264.
- Jong, Franciska de, Bente Maegaard, Darja Fišer, Dieter Van Uytvanck & Andreas Witt. 2020. Interoperability in an infrastructure enabling multidisciplinary research: The case of CLARIN. *Proceedings of the International Conference on Language Resources and Evaluation (LREC)* 12. 3406–3413.

- McGillivray, Barbara, Beatrice Alex, Sarah Ames, Guyda Armstrong, David Beavan, Arianna Ciula, Giovanni Colavizza, James Cummings, David De Roure, Adam Farquhar, Simon Hengchen, Anouk Lang, James Loxley, Eirini Goudarouli, Federico Nanni, Andrea Nini, Julianne Nyhan, Nicola Osborne, Thierry Poibeau, Mia Ridge, Sonia Ranade, James Smithies, Melissa Terras, Andreas Vlachidis & Pip Willcox. 2020. The challenges and prospects of the intersection of humanities and data science: A white paper from The Alan Turing Institute. <https://ndownloader.figshare.com/files/24463460> (accessed 14 April 2021).
- McLuhan, Marshall, 1962. *The Gutenberg galaxy: The making of typographic man*. Toronto: University of Toronto Press.
- O'Neil, Cathy. 2017. *Weapons of math destruction: How Big Data increases inequality and threatens democracy*. London: Penguin.
- Palandri, Marta. 2020. DH Education Knowledge Map – creating knowledge maps via hypertext. Blog post. <https://medium.com/@marta.p/dh-education-knowledge-map-creating-knowledge-webs-via-hypertext-cfb6cc094c17> (accessed 23 March 2021).
- Palandri, Marta & Raphael Mitsch. 2020. DH Education Knowledge Map – creating knowledge maps via hypertext. GitHub Repository. <https://github.com/rmitsch/dh-knowledge-map> (accessed 14 April 2021).
- Petruskaitė, Rūta, Darius Amilevičius, Virginijus Dadurkevičius, Tomas Krilavičius, Gailius Raškinis, Andrius Utka & Jurgita Vaičėnienė. 2022. CLARIN-LT: Home for Lithuanian language resources. In Darja Fišer & Andreas Witt (eds.), *CLARIN. The infrastructure for language resources*. Berlin: deGruyter.
- Prado, Javier Calzada & Miguel Ángel Marzal. 2013. Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri* 63 (2) 123–134.
- Rossi, Giorgio, Jelena Angelis, Filipa Borrego, Joy Davidson, Elena Hoffert, Richard Wade & Patricia Postigo McLaughlin (eds.). 2020. Supporting the transformative impact of research infrastructures on European research. Report of the high-level expert group to assess the progress of ESFRI and other world class research infrastructures towards implementation and long-term sustainability. European Commission. [https://ec.europa.eu/info/sites/info/files/research\\_and\\_innovation/strategy\\_on\\_research\\_and\\_innovation/documents/ec\\_rtd\\_transformative-impact-ris-on-euro-research.pdf](https://ec.europa.eu/info/sites/info/files/research_and_innovation/strategy_on_research_and_innovation/documents/ec_rtd_transformative-impact-ris-on-euro-research.pdf) (accessed 14 April 2021).
- Safradin, Barbara & Franciska de Jong. 2017. CLARIN-PLUS D5.2 Operational course and education material registry. [https://office.clarin.eu/v/CE-2017-0986-CLARINPLUS-D5\\_2.pdf](https://office.clarin.eu/v/CE-2017-0986-CLARINPLUS-D5_2.pdf) (accessed 14 April 2021).
- Sahle, Patrick, Johanna Puhle & Lisa Rau. 2011. Digitale Geisteswissenschaften. [https://dig-hum.de/sites/dig-hum.de/files/cceh\\_broschuereweb.pdf](https://dig-hum.de/sites/dig-hum.de/files/cceh_broschuereweb.pdf) (accessed 14 April 2021).
- Schnapp, Jeffrey, Peter Lunenfeld & Todd Presner. 2009. The digital humanities manifesto 2.0. <https://www.digitalmanifesto.net/manifestos/17/> (accessed 14 April 2021).
- Tasovac, Toma, Adrien Barbaresi, Thibault Clérice, Jennifer Edmond, Natalie Ermolaev, Vicky Garnett & Clifford Wulfman. 2016. APIs in digital humanities: The infrastructural turn. Digital Humanities 2016, Kraków, Poland, 11–16 July. 93–96. <https://hal.archives-ouvertes.fr/hal-01348706/document>
- Tognitz, Martina & Gehrig, Lukas. 2020. CORIANDEr. GitHub Repository. <https://github.com/bellerophons-pegasus/CORIANDEr> (accessed 14 April 2021).
- UNESCO Science Report. 2015. Towards 2030. <https://unesdoc.unesco.org/ark:/48223/pf0000235406> (accessed 14 April 2021).

- Wessels, Leon, Tom Gheldof & Tanja Wissik. 2019. The DH Course Registry: An international platform for finding and promoting DH courses. Poster presented at the DH Benelux Conference 2019, University of Liège, 11–13 September.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons. 2016. The fair guiding principles for scientific data management and stewardship. *Scientific Data* 3. 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wissik, Tanja, Jennifer Edmond, Frank Fischer, Franciska de Jong, Stefania Scagliola, Andrea Scharnhorst, Hendrik Schmeer, Walter Scholger & Leon Wessels. 2020. Teaching digital humanities around the world: An infrastructural approach to a community-driven DH Course Registry. *Library Tribune* (6). 1–27.
- Wissik, Tanja, Hendrik Schmeer, Frank Fischer, Franciska de Jong, Walter Scholger & Leon Wessels. 2020. DH teaching activities as a resource for research: Accessing the DH Course Registry data via an API. Poster presented at the DH Conference 2020, virtual.
- Woldrich, Anna, Katharina Strnadl & Tanja Wissik. 2020. *Deliverable: Digital Humanities Course Registry. Dissemination Strategy*. Unpublished project deliverable.