Information Research - Vol. 28 No. 4 (2023)

More than data repositories: perceived information needs for the development of social sciences and humanities research infrastructures

Anna Sendra, Elina Late, and Sanna Kumpulainen
DOI: https://doi.org/10.47989/ir284598

Abstract

Introduction. The digitalization of social sciences and humanities research necessitates research infrastructures. However, this transformation is still incipient, highlighting the need to better understand how to successfully support data-intensive research.

Method. Starting from a case study of building a national infrastructure for conducting data-intensive research, this study aims to understand the information needs of digital researchers regarding the facility and explore the importance of evaluation in its development.

Analysis. Thirteen semi-structured interviews with social sciences and humanities scholars and computer and data scientists processed through a thematic analysis revealed three themes (developing a research infrastructure, needs and expectations of the research infrastructure, and an approach to user feedback and user interactions).

Results. Findings reveal that developing an infrastructure for conducting data-intensive research is a complicated task influenced by contrasting information needs between social sciences and humanities scholars and computer and data scientists, such as the demand for increased support of the former. Findings also highlight the limited role of evaluation in its creation.

Conclusions. The development of infrastructures for conducting data-intensive research requires further discussion that particularly considers the disciplinary differences between social sciences and humanities scholars and computer and data scientists. Suggestions on how to better design this kind of facilities are also raised.

Introduction

The advent and utilization of digital tools and materials (Nelimarkka, 2023) have transformed the fields of humanities into digital humanities and social sciences into computational social sciences. This process of digitalisation has allowed scholars in social sciences and humanities to conduct research with new approaches. For example, thanks to the use of digital tools and materials, a study that compares the coverage of a certain topic in different newspapers can be conducted faster and include more periodicals in the analysis. However, to better cope with the process of digitalisation, digital humanities and computational social sciences need better research infrastructures (hereafter, infrastructures). This study conceives infrastructures as information services created to support the research processes of the people who will use its digital tools and materials. Examples of international infrastructures for the digital humanities and computational social sciences include Huma-Num (https://www.huma-num.fr/about-us/) in France and CLaDA-BG (https://clada-bg.eu/en/about/organization.html) in Bulgaria. Nevertheless, the challenge is that the development of infrastructures continues to be mostly focused on its contents rather than on the desires of its end-users (Late and Kumpulainen, 2022). Considering that social sciences and humanities scholars may still have doubts about the use of digital tools and materials (Craig, 2015), there is a need to create these facilities from socio-technical perspectives (Foka et al., 2018). Adopting this approach will allow a better understanding of their information needs (Wilson, 1981) to improve these infrastructures.

In line with previous research (e.g., Fabre et al., 2021), it is also necessary to determine if these information services work and are successful in delivering desired outputs. However, assessment frameworks for infrastructures in the fields of digital humanities and computational social sciences are currently scarce. Accordingly, the aims of this study are to (1) determine the desires of digital researchers regarding an infrastructure for the digital humanities and computational social sciences; and (2) contribute to the formative evaluation of this kind of facility already during its development. While we acknowledge the distinct scholarly practices within social sciences and humanities disciplines (e.g., in terms of methods, data sources, etc.), our focus is on the common computational turn of these fields of study (Borgman, 2007). The case examined is a researcher driven national infrastructure for the digital humanities and computational social sciences that is under early stages of development and using a bottom-up approach in its creation. Specifically, three research questions are raised:

RQ1: What service offerings are relevant to the creation of this type of infrastructure?
RQ2: What are the common desires of digital researchers regarding this type of infrastructure?
RQ3: What is the role of user feedback and user interactions in the development of this type of infrastructure?

The research is based on qualitative, semi-structured interviews. In the following, we first provide an overview of previous research that analyses the creation of infrastructures for the digital humanities and computational social sciences. We then introduce the case study, including details about how the facility is being developed and how this research fits in the process. Next, we report the research setting and present our results, including both technical and non-technical perspectives. The study closes with a discussion and conclusion about our findings, while providing recommendations to improve the development of infrastructures.

Literature review

Although the definition of infrastructure is not consolidated (Caliari et al., 2020), this paper follows the explanation provided by Golub et al. (2020, p. 549), where these information services also '*include human networks and scholarly organizations as crucial types of resources and services that a DH research community uses to conduct research and promote innovation in the area*'. This is because

infrastructures are more than just data archives and tool repositories. These facilities have the potential to promote open science (Fabre et al., 2021), increase multi- and inter-disciplinary research (Golub et al., 2020) and support new ways of conducting scientific research (Dallmann et al., 2015; Ribes, 2014). A recent example is related to the COVID-19 pandemic, where thanks to the infrastructures in place, it was possible to develop a vaccine to confront the virus within a record time (Zakaria et al. 2021). Therefore, these information services are also capable of creating communities that better satisfy the information needs of researchers. Nevertheless, while infrastructures have a long tradition in fields like clinical research, the development of these facilities in the social sciences and humanities remains at its infancy (Kumpulainen and Late, 2022).

Several reasons may respond to this shortage of infrastructures for conducting data-intensive social sciences and humanities research. On the one hand, Foka et al. (2018) argue that the use of computational methods represents a cultural transformation for social sciences and humanities scholars, as these approaches are fundamentally changing the way research is usually conducted within these disciplines. On the other hand, previous research highlights that developing these information services for these fields is a complex process (Craig, 2015; Foka et al., 2018). A main obstacle is heterogeneity (Henrich and Gradl, 2013; Kálmán et al., 2015). As Koolen et al. (2020) explain, studies within these fields involve a diversity of research tasks. The problem remains that most existing digital tools only cover a certain type of work (Liu and Wang, 2020), thus limiting the potential of the infrastructure (Waters, 2022). Similarly, there is also heterogeneity in relation to both the data (Matres et al., 2018; Parkoła et al., 2019) and the end-user groups of these facilities (Bermúdez-Sabel et al., 2022), which can become a challenge in terms of standardization (Bermúdez-Sabel et al., 2022; Kaltenbrunner, 2017; Mariani, 2009).

Another barrier concerning infrastructures for conducting data-intensive social sciences and humanities research is uncertainty (Matres et al., 2018). Here uncertainty should be understood from a broad perspective. Firstly, these information services may face uncertainty in terms of policies, both at the institutional (Kálmán et al., 2015; Kaltenbrunner, 2017) and the national level (Kaltenbrunner, 2017). This means, for example, that not all scholars may be disposed to recognize the infrastructure as the point of reference for conducting data-intensive research (Almas, 2017). Secondly, another point of instability is related to how these facilities are funded (Mariani, 2009). It is often the case that infrastructures are developed within '*temporary research projects that are only supported for a limited period of time*' (Buddenbohm et al., 2017, p. 235), thus challenging their long-term continuity (Barbuti, 2018). Thirdly, these information services may also face uncertainty in terms of the resources that are going to be part of the infrastructure. As traditional social sciences and humanities approaches heavily rely on intuition, developing these facilities in a successful manner involves identifying what can and cannot be computerised (Orlandi, 2021). That is, they need to be in harmony with the scholarly practices of those who engage in data-intensive research.

These obstacles add to the challenges faced by existing infrastructures. In terms of the resources provided by these information services, Waters (2022) explained that most solutions are often difficult to use. This could be related to a shortage of cohesiveness among members of the infrastructure, especially in terms of training (Perkins et al., 2022). There is also a need for more flexible (Bermúdez-Sabel et al., 2022) and interoperable (Waters, 2022) resources. Regarding sustainability, previous research highlights that existing facilities are far from sustainable for reasons such as insufficient documentation (Buddenbohm et al., 2017). From the viewpoint of governance, the main problem of existing infrastructures for the digital humanities and computational social sciences is a lack of coordination and communication (Kálmán et al., 2015; Kaltenbrunner, 2017), which could lead these information services to *reinventing the wheel*

(Kaltenbrunner, 2017) or even failure (Oberbichler et al., 2022). The latter is precisely one of the elements behind the collapse of SyBIT (SystemsX.ch Biology IT - http://www.systemsx.ch/projects/systems-biology-it/), which aimed to develop a common data repository for a systems biology infrastructure in Switzerland (Kaufmann et al., 2020).

All these issues point us to the participants in these facilities. One obvious interested party are the funders and policy makers (Kaltenbrunner, 2017). Another is the service developers, which can include cultural heritage institutions (e.g., libraries, archives) and researchers from different organizations, both computer and data scientists and social sciences and humanities scholars. Finally, there are the end-users, including those researchers not involved in the creation of these information services, as well as other individuals and professionals (e.g., journalists). Each of these interested parties has different information needs that should be considered if they are needed to support social sciences and humanities research processes. In addition to adopting mainly techno-centric designs (Late and Kumpulainen, 2022), the problem with developing these kinds of facilities remains that '*networks are not always anchored all the way down to individual scholars within the included disciplines*' (Golub et al., 2020, p. 553). Considering that continuous discussion with different interested parties is important for developing an infrastructure (Perkins et al., 2022), it becomes crucial to include all of them in the conversations around the creation of these information services.

Research setting

This study examines the infrastructure for the digital humanities and computational social sciences in Finland, also known as DARIAH-FI (https://www.dariah.fi/what-is-dariah-fi/).
In Finland, interest in these fields has been increasing in the past few years (Matres et al., 2018). For example, social sciences and humanities scholars are currently using digital materials '*more extensively than printed or otherwise analogue data*' (Matres et al., 2018, p. 40). However, the community remains scattered (Matres et al., 2018), and some of the existing resources are more developed than others (Matres, 2016). The idea behind DARIAH-FI is to bring both the community and the resources under the same roof, to increase the efficiency (through coordination) and sustainability (through long-term building) of the resources used to conduct data-intensive research. The creation of this facility is a joint effort between several organizations, including the University of Helsinki, Aalto University, University of Turku, Tampere University, Jyväskylä University, and the University of Eastern Finland, cultural heritage organizations (National Library of Finland) and service centres (Language Bank of Finland, CSC – IT Center for Science).

Other relevant institutions, such as the National Archives of Finland, also take part in the process as collaborators. The overall aim of DARIAH-FI, which is still under development, is to provide tools, datasets, and workflows for enhancing data-intensive research in the social sciences and humanities. These service offerings will be centrally provided through the CSC – IT Center for Science to ensure accessibility. The infrastructure, which is expected to integrate into DARIAH-EU in the years ahead, covers three main areas, including ingestion, pre-processing and enrichment, and analysis. To ensure the creation of all these resources from an evidence-based perspective, semi-structured interviews with digital researchers were collected. All interviewed researchers are members of the DARIAH-FI development project, where their tasks are mostly focused on building a specific tool, dataset, or workflow for the information service. However, at the same time our informants are the end-users of the infrastructure (see e.g., Costabile et al., 2008), where their tasks will mainly concentrate on using and/or maintaining the resources developed. Likewise, the sample includes researchers with background both in social sciences and humanities and computer and data science. Since the interviews were conducted in the early days of the development phase, data collected reveals their preliminary thoughts about the infrastructure.

Selection of participants

Semi-structured interviews (n=13) were conducted using Microsoft Teams. The interviews took place between 7–27 September 2022. The categorization presented below (Table 1) is based on interview data about their background and main tasks within the DARIAH-FI development project.

N	Profile	Background
13	Non-technical = 9	Applied linguistics = 2
		Academic publishing and cultural heritage = 1
		Digital cultural studies = 1
		Finnish history = 1
		History and digital humanities = 1
		History and political sciences = 1
		Political science = 1
		Social research = 1
	Technical = 4	Computer science = 2
		Computational linguistics = 1
		Data science = 1

Table 1. Profile of the participants

The first author was responsible for recruiting and interviewing the participants. The recruitment process consisted of three participant calls between August and September 2022. The first call was sent through the DARIAH-FI Slack's workspace (29.8.2022), which had between 70 and 80 members when the invitation was posted. The second and third calls were sent through e-mail (6.9.2022; 16.9.2022). In this case, between 5 and 10 people received the invitation, as the reminders were only directed to the leaders of the different work packages.

After expressing their interest in taking part in the study, participants were offered multiple time slots in several days for the interview. Once an agreement was reached, participants were given an information sheet before the day of the interview that explained the study in more depth. The document included details regarding the data management plan and the purpose of the research, as well as the possibility of withdrawing from the study. Participants were also given a link where they could either ask questions about the information sheet or give their consent to participate in the study. All participants gave their informed consent.

Data collection

Online interviews were conducted in English due to our international research team. Overall, the interviews had a twofold purpose: (1) understanding in detail the tools, datasets, and workflows that are going to be part of the infrastructure; and (2) determining the needs and expectations of the members of the DARIAH-FI development project regarding the information service, placing special emphasis on evaluation. The interviews included questions on (a) the development state of the tool, dataset, or workflow, (b) technical aspects, (c) information needs and expectations, and (d) user feedback and user interactions. The guide used in the interviews is available here. (https://doi.org/10.5281/zenodo.8318015).

Lasting between 40 and 100 minutes, the interviews were recorded on digital video files and transcribed verbatim using a professional service. The research team cross-checked the transcripts for consistency and then proceeded to pseudonymise the dataset to safeguard the

participants' personal data. The digital video files were deleted once these processes were finalized. Only data from topics (a), (c), and (d) listed above were considered for the analysis. Some relevant excerpts from topic (b) were also included in the analysis (e.g., extracts discussing the role of log data).

Data analysis

Data analysis was based on the conceptual framework of inductive (reflexive) thematic analysis proposed by Braun and Clarke (2006, 2021). After getting familiar with the data (phase 1), the first author generated a list of preliminary codes from extracts of the interviews, which were classified in a Microsoft Word document (phase 2). This process was repeated in several rounds. Then, the first author generated a list of preliminary themes and sub-themes in another Microsoft Word document (phase 3). The preliminary themes and sub-themes identified were discussed with the research team in multiple internal meetings. Afterwards, the first author reviewed (phase 4) and refined (phase 5) the preliminary themes and sub-themes identified. An updated version of the themes and sub-themes identified were once again discussed with the research team in multiple internal meetings until reaching agreement. Lastly, the first author generated the analysis report in a different Microsoft Word document (phase 6).

Findings

The results of the inductive (reflexive) thematic analysis are shown in Figure 1. A total of three themes were identified in the data, including (1) developing a research infrastructure, (2) needs and expectations of the research infrastructure, and (3) an approach to user feedback and user interactions. The analysis also revealed sixteen sub-themes. A detailed description of the findings is provided in the following.

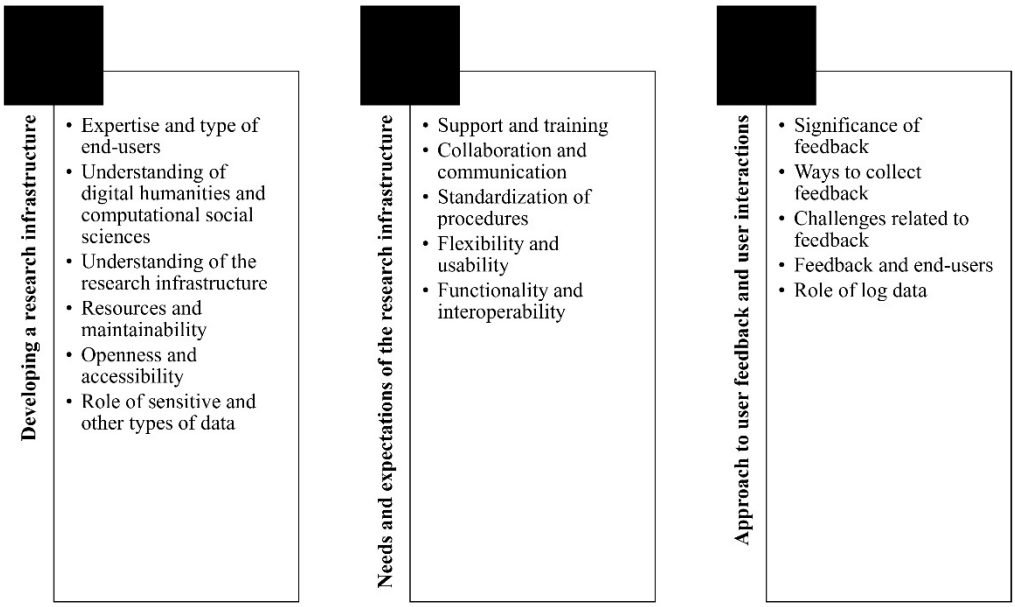

Figure 1. Thematic map

Developing an infrastructure is complex, but so is its continuity

When talking about the elements that affect the creation of an infrastructure for the digital humanities and computational social sciences, several factors were reported by the participants. A first one that can be inferred from the interview data is a lack of shared understanding of what this facility will be. However, data also reveals some views that can be used to start building this consensus. For example, both from the technical and the non-technical perspective, most interviewees see the infrastructure as a data hub that will ease

the access to digital materials, a desire probably related to the efficiency goal determined by the information service. As one computer and data scientist said:

> *The main contribution, I would say, is that they have a properly versioned, referenceable dataset close to a supercomputing facility. So the point indeed is that they don't need to download these 11 terabytes, but that these 11 terabytes are there for them to be used* (P6).

A similar view was echoed by a social sciences and humanities scholar: '*It will hopefully make easier to employ new datasets, new large datasets for the kind of research that we've already been doing*' (P8). Likewise, some interviewees suggested that the information service will increase collaboration opportunities. This perspective was also shared both within the technical and the non-technical profiles. Talking about this issue, P4 commented:

> *I think the primary goal is maybe to support the research that is already ongoing and support the connections between these teams, and also at the same time provide maybe some pilot cases, like flagship projects and very good demonstrations on how we can do this kind of data intensive … social sciences and humanities research so that others can then maybe pick up some ideas from how it's well done* (P4).

Furthermore, almost all social sciences and humanities scholars highlighted that the infrastructure will allow the conduct of research in a different way. As P10 put it:

> *If there would be tools, automated tools, I think that more data could be analysed for the purposes of research. And I think that that would have a big effect on research and what kind of things we can actually analyse, and yeah, get knowledge about* (P10).

Another element that impacts the development of this type of infrastructure is who are going to be its end-users. In this case, two divergent discourses emerged regarding who the interviewees think are going to be the end-users of the solutions they are developing. Respondents with a non-technical profile feel that using the resources will not require a lot of expertise from the end-user, as expressed by P13: '*We're hoping to create a tool that you do not need to see any lines of code to be able to use it, so just skills to be able to learn and use a program*' (P13). Conversely, most participants with a technical profile consider that their solutions will serve a more specific audience: '*We are not really building or using tools that are somehow general purpose, that somebody can go to computer and click something and then something nice happens with the data. It's not really like that*' (P4). Still, a common view among all interviewees regardless of their profile is that understanding how computational approaches work is essential to avoid the 'black box' effect when conducting digital humanities and computational social sciences research:

> *I think it's a kind of metaknowledge of how to enable trustworthy big data centric research, which is that you need to pay mind to the aspects of representativeness and biases in the material, and if you're doing computational enrichment to the material, whether that adds bias and whether that actually captures what you were interested in* (P1).

In this context, some participants raised the possibility of having two interfaces to serve these different skill levels. As P8 put it:

> *I think it's a sliding scale, … at least we are aiming for having a kind of, on one level having a web interface that you can employ pretty easily, and on the other level, there will be an* API *programming interface that you can plug into, also, if you are more versed in that* (P8).

This suggestion was brought up not only by social sciences and humanities scholars, but also by computer and data scientists. A third factor that can be inferred from the interview data is related to the maintainability of the infrastructure. Both within the technical and the non-technical profiles, most interviewees consider that the long-term preservation of the

information service needs to be discussed in further detail, especially in terms of costs and responsibilities. This is a crucial aspect for the sustainability of the infrastructure, clearly highlighted by the following excerpt:

> *I've learned over the years that it's possible to develop even quite interesting systems, but I mean, who maintains them over time, what is the institutional responsibilities and roles and institutional interest in maintaining them long term, because it can't be based on a research project* (P9).

A particular worry among some computer and data scientists is related to versioning, as several work packages involve data that varies over time. As P6 put it, there is a need for creating common plans at the consortium level: '*We need to have clear policies that describe where the data is, where … is the latest version*' (P6). Conversely, the main concern for several social sciences and humanities scholars is the location of the developed resources. This is a rather surprising result, as one of the goals is to centralise all tools, datasets, and workflows through the same service centre (i.e., CSC – IT Center for Science). For example, P3 said: '*We don't know actually where we are expected to store our data, so therefore we have also established an own data server at* UNIVERSITY-NAME, *simply to have some kind of data repository*' (P3). Nevertheless, it seems that only respondents with a technical profile have the maintainability of the infrastructure at the core of their work. As illustrated in the following, some of these participants are focused on making sustainable resources from the start:

> *That's what both of these development projects are actually targeting, because we already have the programmatic interfaces and the user interfaces, so both of these are actually trying to make these prototype tools more maintainable and extensible for the purpose of uptake in a permanent national infrastructure* (P1).

One final element that impacts the development of this type of infrastructure is its accessibility. A common view among social sciences and humanities scholars is that this requirement should be fundamental, so as the facility can reach as many of its (potential) end-users as possible. This result partly correlates with the doubts participants have regarding the deposition of the resources developed. As P13 explained:

> *In an ideal world you would of course have all the tools you need or could think you need available to you. But I know that's not the case. … I think that's one of the things that in an ideal infrastructure these would be accessible to all researchers and students at the university* (P13).

In this context, openness is not only related to the resources that an infrastructure provides, but it also concerns its training opportunities. As P5 commented: '*What I would like to see is that these teaching materials and also the teaching activities related to these tools would also be made open, … instead of tying these resources to certain institutions and certain institutions credentials*' (P5). However, it is not always possible for the service developers to subscribe to the principles of open science. Both within the technical and the non-technical profiles, respondents overall highlighted that licenses hinder the development of the information service, especially in terms of being able to offer more materials. For example, P2 said: '*There's a lot of copyright restrictions … so this is why we only can do this at this stage for copyright free data*' (P2). A related problem for some participants regardless of their profile is the management of sensitive and similar data. In all cases, interviewees agree that censorship should be avoided, but opinions differ as to whether how to best handle these types of data, especially between computer/data scientists. While most of them feel that existing mechanisms are enough, others consider that more measures will need to be taken in the future:

> *The ultimate thing might be that we have to just take the material away from the usage. … So far there hasn't been that kind of phase but somehow it waits on the background, I think that at some point that might come* (P12).

An infrastructure does not only provide data and tools

As for the information needs the informants have regarding the infrastructure, interview data also reveals several elements. A first common requirement relates to the training and support that the information service will offer. Particularly, the non-technical profile respondents highlight the importance of this need in understanding how to use the resources and the possibilities of what they can do with them. As P11 said, using the infrastructure in a satisfactory manner requires to provide community services: '*I think it needs actual persons to be there and help and find, if there is someone who is really at the beginning of their journey*' (P11). A similar view was echoed by another participant:

> *I don't think it's necessarily about how difficult the tools are, it's about social support and organisational support for using them. ... People have the skills and the attitude to be interested and are able to take this on, but there's just no support for this* (P13).

In this context, both social sciences and humanities scholars and computer and data scientists argued that providing good documentation and examples of use is a crucial aspect of training and support. For example, P5 commented:

> *Of course it's good practice to document everything and write user documentation ... It's good practice to at least write something that gives a starting point from where to begin and some resources to acquaintance with if that is necessary to understand the toolset* (P5).

An interviewee with a technical profile added that it is important to provide these materials also for those end-users with more skills: '*It would be good if there were libraries to interface with this. It would be good if the interfaces were better documented. That's something that would increase ease of use for them*' (P1).

Another prevalent need concerns the development of the infrastructure itself. Overall, there is a sense of a lack of internal communication amongst the respondents, especially within the non-technical profile. This shortage of information on what others are doing is affecting the potential of what the information service can offer, thus impacting its efficiency. As one respondent put it:

> *This is something I'm very much missing in this project, these very big discussions about interaction between the packages. As I listened to these presentations, ... there are connections, and we also mentioned that during these meetings, that it seems that work package A or work package C are a bit connected, and we should encourage the work packages also to interact* (P3).

In this context, P7 suggested that satisfying this need goes through generating more discussion opportunities, as current activities are insufficient:

> *We all have full calendars, we can have our general meetings once or twice a year, and a one-day meeting is not enough to share the information, there are five modules and some, how many, 15, 20 work packages, so, you do not have very much time to present your work to others* (P7).

As previously discussed, offering better infrastructures also require meeting the needs of other end-user groups. For a small number of participants, a way of identifying those requirements is by increasing communication at the external level, which in turn would help developing the information service further. This perspective is shared both by social sciences and humanities scholars and computer and data scientists. For example, one respondent said: '*It is really important to be open for feedback to develop our system, to get know what people are wanting from us, and different kind of user groups are also important*' (P2).

A third common requirement is related to standardisation. As an information service, an infrastructure is expected to provide datasets in line with certain standards. Nevertheless, some informants suggest that the creation of the facility is lacking standardisation

procedures, especially in terms of handling the data. Both within the technical and the non-technical profiles, various participants argue about the necessity of standardising processes on issues such as how to manage or harmonise a dataset. This result is probably related to the (perceived) lack of internal communication highlighted above, which in turn has consequences for the desired efficiency level of the infrastructure. As P6 put it:

> *One thing that I have to admit that now only comes to my mind ... is to what extent we describe the changes. Do we want to describe every little change? We want to make them into snapshots, that's for sure, but do we want to also describe every snapshot properly in metadata? Do we want to make this only for major changes?* (P6).

For both types of interviewees, solving this matter during the development phase could also be a way of generating best practices for the entire digital humanities and computational social sciences community, which in turn would meet the needs of other end-user groups:

> *If Helsinki University researcher has data and wants to do a study with a researcher from Jyväskylä University, the agreements are not aligned. ... It would be a lot easier, if there would be some kind of collaboration there, to make the process easier or standardise it in a way* (P11).

The problem remains that standardising procedures is not a clear-cut process for some respondents, especially for social sciences and humanities scholars. For example, a participant suggested that these disciplines may use different vocabulary for referring to the same concept, thus highlighting the need for developing a common lexicon: '*We need a shared glossary to understand terminology similarly, for instance, the term data or research data, they are understood differently in various institutes*' (P7).

One final prevalent need concerns the usability of the infrastructure. Two divergent and conflicting discourses emerged around this concept. Most interviewees, particularly social sciences and humanities scholars said that they would like to work with user-friendly solutions: '*People won't use tools that they cannot use, you know, like if they cannot have it on their own computer, and if it takes five hours trying to find it out, then yeah, it's not really efficient*' (P10). At the same time, almost all participants also highlighted the importance of utilizing flexible resources in their research:

> *That's a major problem in multiple of these systems, that they are not open enough ... They force you into either a user interface in which you should do all things but you cannot because it's not flexible enough, or the API, if it's programmatic, has been designed with a particular, quite limited use case in mind and it doesn't work* (P1).

This perspective was shared both within the technical and the non-technical profiles. The contradiction comes in that flexibility does not always correlate with user-friendliness; a concern mostly expressed by computer/data scientists:

> *There are so many different ways of sharing or manipulating data that I think we can't really imagine all the possible end use cases, so if we would have to choose that we develop something for some end users, we will also limit these possibilities a lot* (P4).

However, a social sciences and humanities scholar demonstrated how interoperability between systems could create both flexible and easy-to-use solutions:

> *The good thing with visone is also that there is a pipe to connect directly with R and communicate directly with R, so it is not necessary to export and import the data, but you can also use the data directly from R. ... This kind of interaction is very important also* (P3).

User feedback is crucial to keep moving the infrastructure forward

When talking about how the infrastructure should be evaluated, interviewees overall consider user feedback as something integral

and important, both within the technical and the non-technical profiles. As P7 put it:

> *I think if the service is new and there's a lot of new things to understand and to develop, it would be very useful if the researchers could tell us more what they need or how this system could be better* (P7).

Nevertheless, computer and data scientists are more inclined to collect this information in non-formalized ways (e.g., face-to-face encounters), whereas social sciences and humanities scholars talk more about formalised procedures for gathering feedback (e.g., questionnaires, interviews). In this context, some of the latter suggested that quantitative ways of collecting user feedback would be more useful once the information service is deployed:

> *If we get to a stage where this is really polished, then of course we might just have some kind of a support email or feedback form, sure, but as long as we are in development we actually need user feedback that's gathered from face-to-face situations* (P13).

Regardless of the method used, providing feedback should not be complex:

> *It could be just an e-mail system or whatever so it doesn't require that we have some huge info IT system in the background collecting the feedbacks for example. But something that is easily available and low threshold to use I would say* (P12).

This is a view shared by a few interviewees with both technical and non-technical profiles. In terms of handling the feedback, both social sciences and humanities scholars and computer and data scientists have the feeling that analysing and implementing feedback requires resources, especially in terms of costs and personnel. As P10 put it: '*Of course it requires someone and someone's time to analyse the feedback*' (P10). A related problem for all respondents, regardless of their profile, is that going through the feedback requires balance. For example, participants argued that not all the requests can be addressed: '*Some projects just get too many issues and pull requests, so that it becomes actually unfeasible for volunteer open-source developers to handle those*' (P5). Another issue concerns how to determine the relevance of the feedback:

> *If there is a lot of user feedback and it's all negative, ... how to see what is the line between somebody who just didn't want to actually read the instructions, and those who read the instructions but did not just understand it* (P11).

Furthermore, while age and gender are not seen as important user characteristics, both social sciences and humanities scholars and computer and data scientists agree that understanding information behaviour requires knowing the profile and skill level of the end-users. One interviewee commented: '*One attribute that does of course affect how people are able to use this kind of tools is their understanding of the digital humanities methods in general*' (P8). An interviewee with a non-technical profile highlighted the importance of collecting feedback with different end-user groups:

> *The end users are one side, they are giving us valuable information about how the portal in general is functioning and how it could be improved, but we should also have continuous debates and discussions with those mostly institutions providing us with data* (P3).

Lastly, interview data reveals that the difference between user feedback and log data remains unclear for most respondents, particularly social sciences and humanities scholars. Those aware of their distinctions see log data as something to understand information behaviour in further detail: '*If there is a kind of, feedback that is not given by the user, but by the actual tool, ... that is of course as valuable as the research, given feedback*' (P11). Both within the technical and the non-technical profiles, other interviewees also see log data as a way of identifying the (potential) userbase of the infrastructure: '*Of course we are very interested in who uses the data, and we probably should need statistics of it, to report to*

our funders that our data is used' (P7). However, computer/data scientists reported overall that there are no plans for logging transactions: '*We're not intending to track what users do with the data*' (P6). Another respondent with a technical profile argued that collecting log data is only worth it when the resource is used on a recurring basis: '*Even if they are openly available I think it will be quite marginal who will use this, so I don't know how much those statistics can tell*' (P4). For other participants, the problem is that logging transactions requires resources, as in the case of user feedback: '*If we were to do that, that would have to have some kind of permanent infrastructure, but now we're working on project money*' (P13). This issue was raised by some social sciences and humanities scholars and computer and data scientists.

Discussion

Our analysis shed light on the information needs of digital researchers regarding an infrastructure for the digital humanities and computational social sciences. It also sought to determine if the development of these information services considers any kind of assessment framework. To achieve this, three research questions, listed earlier, were raised. The discussion included both social sciences and humanities scholars and computer and data scientists. As Ribes (2014, p. 586) explained, '*new research infrastructure projects, i.e., cyberinfrastructure, will face a different and unique set of challenges than those of the past thirty years*'. It is expected that the discussion below will illuminate some of these difficulties, which in turn will help us establish how an infrastructure for the digital humanities and computational social sciences can be more structured and sustainable (Late and Kumpulainen, 2022).

Concerning the first research question (*What service offerings are relevant to the creation of this type of infrastructure?*), results show the difficulty of creating an infrastructure for the digital humanities and computational social sciences. This finding is consistent with the work of other studies in this area (Craig, 2015; Foka et al., 2018). Specifically, the main concern is related to the continuity of the information service, be it economic or otherwise. This is a challenge shared by other infrastructures for the digital humanities and computational social sciences (Buddenbohm et al., 2017). Still, results suggest that there are many, intertwined elements relevant to the creation of this type of infrastructure; hence the complexity of its development. Another interesting finding is that social sciences and humanities scholars seem to be more aware of what it entails to conduct data-intensive research within their disciplines (Foka et al., 2018). For example, regardless of their profile, most respondents consider that understanding how computational approaches work will be essential to use the resources that the infrastructure will offer. The problem remains that, unless more training opportunities are provided (Matres et al., 2018), this requirement might create a barrier for those who are at the beginning of their digital humanities and computational social sciences journey, thus limiting the potential userbase of the information service. However, opinions differ on the required expertise level of the facility. In line with previous research (Borgman, 2009), social sciences and humanities scholars are the ones who mainly expect the infrastructure to offer low threshold resources.

With respect to the second research question (*What are the common desires of digital researchers regarding this type of infrastructure?*), findings suggest differences in the information needs of digital researchers regarding an infrastructure for the digital humanities and computational social sciences according to their profile. For example, the main desire for social sciences and humanities scholars is related to training, which in turn affects their general understanding of the facility. By expecting increased collaborations and greater access to educational materials, the way the infrastructure is conceived is more in line with that of Golub et al. (2020), who consider that networks of people are an essential part of an information service. Conversely, computer and data scientists may not think about this information need because they already have the necessary expertise to use the resources being provided by the infrastructure. Still, the analysis reveals a list of

common requirements (e.g., standardization) to be contemplated in further stages of development. From a broader perspective, these results could help addressing the *retraining vs. collaboration* debate present in digital humanities and computational social sciences (Craig, 2015; Matres et al., 2018), especially for those social sciences and humanities scholars who are in later stages of their academic career. The findings of this study also raised questions in terms of the internal cohesiveness of the facility. Consistent with the literature (Kálmán et al., 2015; Kaltenbrunner, 2017), our analysis highlights that a lack of communication hinders the development of the infrastructure. As Oberbichler et al. (2022, p. 234) remind us, '*interdisciplinary collaboration can be difficult even when all participants have good intentions*'.

Regarding the third research question (*What is the role of user feedback and user interactions in the development of this type of infrastructure?*), results show that evaluation currently has a discreet role in the development of the infrastructure for the digital humanities and computational social sciences analysed in this study. These findings are likely to be related to the lack of resources in terms of costs and personnel also reported in the analysis. However, not engaging other end-user groups from early stages could be a misstep for the creation of the information service (Golub et al., 2020). For example, some of the critical development points for the infrastructure could remain hidden (Koolen et al., 2020). Understanding broader information behaviour could also determine future, relevant digital materials and tools for conducting social sciences and humanities research (Liu and Wang, 2020). Another interesting result concerns logging transactions. Findings reveal that service developers, particularly those with non-technical profiles, have difficulties to discern between user feedback and log data. This might be explained by the fact that user-related knowledge is '*often limited to simple methodological approaches to collecting and counting impact metrics*' (Mayernik et al., 2017, p. 1355). However, logging transactions could be an important issue for the development of the facility. In addition to provide detailed records about who is using the infrastructure, log data generates information on how it is being used (Jonkers et al., 2012). Therefore, enhancing the role of log data within the information service could contribute to the development of the infrastructure and its continuity.

Implications that might apply to almost any software
Resources for evaluation
Flexible and interoperable digital tools
Implications that might apply to infrastructures in general
Resources for long-term preservation
Standardization of internal procedures
Implications that might apply particularly to infrastructures for the digital humanities and computational social sciences
Improved communication within the community
Increased training

Table 2. Design implications

Based on these results, six design implications for an infrastructure for the digital humanities and computational social sciences can be drafted (Table 2). While some of these recommendations might apply to almost any software, to infrastructures in general, or specifically to infrastructures for the digital humanities and computational social sciences, in the following they are discussed considering the case study analysed in this research. Firstly,

regularly collecting user feedback and user interactions is important, even in early development stages. As shown by the results of this study, this is a clear area of improvement for the case discussed here. Secondly, resources should be as flexible and interoperable as possible to accommodate the desires of all types of end-users. While this finding is consistent with the literature (Bermúdez-Sabel et al., 2022; Waters, 2022), our results also highlight that meeting this information need is not an easy task. Therefore, exploring how this requirement can be fulfilled becomes an important area for further research. Thirdly, resources aimed at the continuity of the information service should be expanded. As identified elsewhere (Buddenbohm et al., 2017, p. 235), our findings further support the idea that relates sustained funding with long-term maintenance of infrastructures.

Fourthly, procedures should be standardized at the internal level. According to our results, this condition particularly applies to the practicalities of the facility (e.g., data versioning). Meeting this information need could also be a way of addressing the heterogeneity present in social sciences and humanities research (Henrich and Gradl, 2013; Kálmán et al., 2015). Fifthly, communication within end-user groups should be facilitated. While this challenge is not new (Kálmán et al., 2015; Kaltenbrunner, 2017), the information service discussed in this study is currently addressing this issue. Not only the infrastructure counts with a national coordinator that tries to ease this process, but different activities where digital researchers can talk about their views and desires regarding the facility are organised regularly. Finally, educational opportunities on, for example, how to use the resources offered by the facility should be increased. Particularly, our findings reveal that digital researchers are interested in guidance and documentation. This is an important result when considering that some end-users, especially those with non-technical backgrounds, might remain sceptical of the digital transformation of social sciences and humanities research (Craig, 2015).

Conclusion

This study adds to the body of research indicating that creating an infrastructure, whether for the digital humanities and computational social sciences or other disciplines, is indeed a complicated process (Oberbichler et al., 2022; Timotijevic et al., 2021). Not only do multiple, interrelated factors affect the development of the facility, but the information needs of the people who use digital tools and materials for conducting social sciences and humanities research are diverse and (sometimes) contradictory. In fact, many of the findings identified here have been tried to be solved for decades (Star and Ruhleder, 1994). What is the problem, then? A possible explanation may be related to the disciplinary differences between social sciences and humanities scholars and computer and data scientists. While computing increasingly mediates social sciences and humanities research (Berry, 2011), '*we also cannot expect people to abandon working practices instantly when they have suited them well over many years and, in some humanities fields, generations*' (Warwick, 2012, p. 18). Although our sample is relatively small, it provides enough hints of the differences in the information needs between social sciences and humanities scholars and computer and data scientists. Therefore, future studies should explore how these distinct academic cultures impact the creation of an infrastructure for the digital humanities and computational social sciences. Still, our findings allowed us to outline design implications to develop the facility in the most favourable way. Summarised in Table 2, these suggestions include working towards the internal standardisation of procedures and maintaining a constant dialogue with different end-user groups. Likewise, '*mutual understanding between disciplines emerges from communication*' (Oberbichler et al., 2022, p. 235). Therefore, working towards this meeting point requires of an information service like DARIAH-FI the creation of further opportunities for exchanging ideas both within and beyond the consortium.

The latter argument brings us to the limitations of this study. One source of weakness is the potential bias towards participants with non-technical profiles, which restricts the generalizability of the results. Another limitation is that this study only provides the views of members of the DARIAH-FI development project. In this context, we will complement the findings of this research with additional semi-structured interviews with other interested parties (e.g., social sciences and humanities scholars not related to the creation of the information service). This second study will allow us to compare if their desires match with the results reported in this research. Another aspect to explore in the future study is whether the different academic cultures affect the creation of an information service for the digital humanities and computational social sciences. Similarly, the results highlighted here are only a still picture from the early stages of the development phase. As needs and expectations may change over time (Ribes, 2014), further research needs to reassess in later stages the views of the individuals that are participating in the development of the infrastructure discussed in this study (e.g., during the maintenance phase). Despite these limitations, this research offers from a user-centred perspective valuable and practical recommendations to attain the best possible design of an information service for the digital humanities and computational social sciences.

Acknowledgements

This work was supported by the Research Council of Finland, grant numbers 345618 and 351247. We would also like to thank Jaakko Peltonen and Farid Alijani for helping with the development of the interview guide and the participants for their time and involvement in the study.

About the authors

Anna Sendra works as a Project Manager in the Faculty of Information Technology and Communication Sciences at Tampere University (Finland). She received her Ph.D. from Universitat Rovira i Virgili (Spain) and her research focuses on exploring the processes of digitization from the perspective of media and communication studies, especially in the context of health. She can be contacted at anna.sendratoset@tuni.fi.

Elina Late works as a Senior Research Fellow at Tampere University in the Faculty of Information Technology and Communication Sciences. She received her Ph.D. from Tampere University and her research interests include scholar's information practices, scholarly communication and open science. She can be contacted at elina.late@tuni.fi.

Sanna Kumpulainen holds the position of Associate Professor in Information Studies at Tampere University, Finland. Her research centers on exploring the dynamics of human interaction with information and developing strategies for its facilitation. Her research pursuits encompass interactive information retrieval, digital libraries, and the promotion of open science. To reach out to her, you can contact her via email at sanna.kumpulainen@tuni.fi.

Disclosure statement

A work-in-progress version of this study was submitted as a poster to the 86th Annual Meeting of the Association for Information Science and Technology (ASIS&T).

References

Almas, B. (2017). Perseids: experimenting with infrastructure for creating and sharing research data in the digital humanities. *Data Science Journal*, 16(19), 1-17. https://doi.org/10.5334/dsj-2017-019

Barbuti, N. (2018). From digital cultural heritage to digital culture: Evolution in digital humanities. In E. Reyes, S. Szoniecky, A. Mkadmi, G. Kembellec, R. Fournier-S'niehotta, F. Siala-Kallel, M. Ammi, & S. Labelle (Eds.), *Proceedings of the 1st International Conference on Digital Tools & Uses Congress*, Paris, France, October 3-5, 2018 (pp. 1-3). Association for Computing Machinery. https://doi.org/10.1145/3240117.3240142

Bermúdez-Sabel, H., Platas, M. L. D., Ros, S., & González-Blanco, E. (2022). Towards a common model for European poetry: Challenges and solutions. *Digital Scholarship in the Humanities*, 37(4), 921-933. https://doi.org/10.1093/llc/fqab106

Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, 12. https://web.archive.org/web/20231029143159/https://culturemachine.net/wp-content/uploads/2019/01/10-Computational-Turn-440-893-1-PB.pdf

Borgman, C. L. (2007). *Scholarship in the digital age: information, infrastructure, and the Internet*. MIT Press.

Borgman, C. L. (2009). The digital future is now: a call to action for the humanities. *Digital Humanities Quarterly*, 3(4). http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis?. *Qualitative Research in Psychology*, 18(3), 328-352. https://doi.org/10.1080/14780887.2020.1769238

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

Caliari, T., Rapini, M. S., & Chiarini, T. (2020). Research infrastructures in less developed countries: the Brazilian case. *Scientometrics*, 122(1), 451-475. https://doi.org/10.1007/s11192-019-03245-2

Costabile, M. F., Mussio, P., Provenza, L. P., & Piccinno, A. (2008). End users as unwitting software developers. In R. Abraham, M. Burnett, & M. Shaw (Eds.), *Proceedings of the 4th International Workshop on End-user Software Engineering*, Leipzig, Germany, May 12, 2008 (pp. 6-10). Association for Computing Machinery. https://doi.org/10.1145/1370847.1370849

Craig, A. B. (2015). Science gateways for humanities, arts, and social science. In G. D. Peterson (Ed.), *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, St. Louis (Missouri), United States, July 26-30, 2015 (pp. 1-3). Association for Computing Machinery. https://doi.org/10.1145/2792745.2792763

Dallmann, A., Lemmerich, F., Zoller, D., & Hotho, A. (2015). Media Bias in German Online Newspapers. In Y. Yesilada, R. Farzan, & G. Houben (Eds.), *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, Guzelyurt, Northern Cyprus, September 1-4, 2015 (pp. 133-137). Association for Computing Machinery. https://doi.org/10.1145/2700171.2791057

Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Qualitative Science Studies*, 2(1), 42-64. https://doi.org/10.1162/qss_a_00111

Foka, A., Misharina, A., Arvidsson, V., & Gelfgren, S. (2018). Beyond humanities qua digital: Spatial and material development for digital research infrastructures in HumlabX. *Digital Scholarship in the Humanities*, 33(2), 264-278. https://doi.org/10.1093/llc/fqx008

Golub, K., Göransson, E., Foka, A., & Huvila, I. (2020). Digital humanities in Sweden and its infrastructure: *Status quo* and the *sine qua non*. *Digital Scholarship in the Humanities*, 35(3), 547-556. https://doi.org/10.1093/llc/fqz042

Golub, K., Milrad, M., Huang, M. P., Tolonen, M., Matres, I., & Bergsland, A. (2017). Current efforts, perspectives and challenges related to Digital Humanities in Nordic Countries. In K. Golub, & M. Milrad (Eds.), *Extended Papers of the International Symposium on Digital Humanities*, Växjö, Sweden, November 7-8, 2016 (pp. 169-177). Linnaeus University. https://web.archive.org/web/20231029144539/https://ceur-ws.org/Vol-2021/paper12.pdf

Henrich, A., & Gradl, T. (2013). DARIAH-DE: Digital research infrastructure for the arts and humanities – concepts and perspectives. *International Journal of Humanities and Arts Computing*, 7(Supplement), 47-58. https://doi.org/10.3366/ijhac.2013.0059

Hormia-Poutanen, K., Kautonen, H., Lassila, A. (2013). The Finnish National Digital Library: a national service is developed in collaboration with a network of libraries, archives and museums. *Insights*, 26(1), 60-65. https://doi.org/10.1629/2048-7754.26.1.60

Jonkers, K., Anegon, F. d. M., Aguillo, I.F. (2012). Measuring the usage of e-research infrastructure as an indicator of research activity. *Journal of the American Society for Information Science and Technology*, 63(7), 1374-1382. https://doi.org/10.1002/asi.22681

Kálmán, T., Tonne, D., & Schmitt, O. (2015). Sustainable preservation for the arts and humanities. *New Review of Information Networking*, 20(1-2), 123-136. https://doi.org/10.1080/13614576.2015.1114831

Kaltenbrunner, W. (2017). Digital infrastructure for the humanities in Europe and the US: Governing scholarship through coordinated tool development. *Computer Supported Cooperative Work*, 26(3), 275-308. https://doi.org/10.1007/s10606-017-9272-2

Kaufmann, D., Kuenzler, J., & Sager, F. (2020). How (not) to design and implement a large-scale, interdisciplinary research infrastructure. *Science and Public Policy*, 47(6), 818-828. https://doi.org/10.1093/scipol/scaa042

Koolen, M., Kumpulainen, S., & Melgar-Estrada, L. (2020). A workflow analysis perspective to scholarly research tasks. In H. O'Brien, L. Freund, I. Arapakis, O. Hoeber, & I. Lopatovska (Eds.), *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, Vancouver, Canada, March 14-18, 2020 (pp. 183-192). Association for Computing Machinery. https://doi.org/10.1145/3343413.3377969

Kumpulainen, S., & Late, E. (2022). Struggling with digitized historical newspapers: Contextual barriers to information interaction in history research archives. *Journal of the Association for Information Science and Technology*, 73(7), 1012-1024. https://doi.org/10.1002/asi.24608

Late, E., & Kumpulainen, S. (2022). Interacting with digitised historical newspapers: understanding the use of digital surrogates as primary sources. *Journal of Documentation*, 78(7), 106-124. https://doi.org/10.1108/JD-04-2021-0078

Liu, S., & Wang, J. (2020). How to organize digital tools to help scholars in digital humanities research?. In R. Huang, D. Wu, G. Marchionini, D. He, S. J. Cunningham, & P. Hansen (Eds.), *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, Virtual Event, China,

August 1-5, 2020 (pp. 373-376). Association for Computing Machinery. https://doi.org/10.1145/3383583.3398615

Mariani, J. (2009). Research infrastructures for human language technologies: A vision from France. *Speech Communication*, 51(7), 569-584. https://doi.org/10.1016/j.specom.2007.12.001

Matres, I. (2016). Digital historical materials for academics, educators, hobbyists, creatives and browsers: the visitors evaluate digi.kansalliskirjasto.fi. *Informaatiotutkimus*, 35(3), 48-49. https://web.archive.org/web/20231029145341/https://journal.fi/inf/article/view/59436

Matres, I., Oiva, M., & Tolonen, M. (2018). In between research cultures – The state of digital humanities in Finland. *Informaatiotutkimus*, 37(2), 37-61. https://doi.org/10.23978/inf.71160

Mayernik, M. S., Hart, D. L., Maull, K. E., & Weber, N. M. (2017). Assessing and tracing the outcomes and impact of research infrastructures. *Journal of the Association for Information Science and Technology*, 68(6), 1341-1359. https://doi.org/10.1002/asi.23721

Nelimarkka, M. (2023). *Computational thinking and social science: combining programming, methodologies and fundamental concepts*. SAGE Publications.

Oberbichler, S., Boroş, E., Doucet, A., Marjanen, J., Pfanzelter, E., Rautiainen, J., Toivonen, H., & Tolonen, M. (2022). Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians. *Journal of the Association for Information Science and Technology*, 73(2), 225-239. https://doi.org/10.1002/asi.24565

Orlandi, T. (2021). Reflections on the development of digital humanities. *Digital Scholarship in the Humanities*, 36(Supplement 2), ii222-ii229. https://doi.org/10.1093/llc/fqaa048

Parkoła, T., Chudy, M., Łukasik, E., Jackowski, J., Kuśmierek, E., & Dahlig-Tyrek, E. (2019). MIRELA – Music Information Research Environment with dLibrA. In D. Rizo (Ed.), *6th International Conference on Digital Libraries for Musicology*, The Hague, Netherlands, November 9, 2019 (pp. 65-69). Association for Computing Machinery. https://doi.org/10.1145/3358664.3358674

Perkins, D. G., Knuth, S. L., Lindquist, T., Johnson, A. M., Eichmann-Kalwara, N., Dunn, T. (2022). Challenges and Lessons Learned of Formalizing the Partnership Between Libraries and Research Computing Groups to Support research: the Center for Research Data and Digital Scholarship. In J. Wernert, A. Chalker, S. Smallen, T. Samuel, & J. P. Navarro (Eds.), *Practice and Experience in Advanced Research Computing* (pp. 1-4). Association for Computing Machinery. https://doi.org/10.1145/3491418.3535165

Ribes, D. (2014). The kernel of a research infrastructure. In S. Fussell, W. Lutters, M. R. Morris, & M. Reddy (Eds.), *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, Baltimore (Maryland), United States, February 15-19, 2014 (pp. 574-587). Association for Computing Machinery. https://doi.org/10.1145/2531602.2531700

Star, S. L., & Ruhleder, K. (1994). Step Towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. In J. B. Smith, F. Don Smith, & T. W. Malone (Eds.), *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, Chapel Hill (North Carolina), United States, October 22-26, 1994 (pp. 253-264). Association for Computing Machinery. https://doi.org/10.1145/192844.193021

Timotijevic, L., Astley, S., Bogaardt, M.J., Bucher, T., Carr, I., Copani, G., De la Cueva, J., Eftimov, T., Finglas, P., Hieke, S., Hodgkins, C.E., Koroušić Seljak, B., Klepacz, N., Pasch, K., Maringer, M., Mikkelsen, B.E., Normann, A., Ofei, K.T., Poppe, K., … Zimmermann, K. (2021). Designing a research infrastructure (RI) on food behaviour and health: Balancing user needs, business model,

governance mechanisms and technology. *Trends in Food Science & Technology*, 116(October 2021), 405-414. https://doi.org/10.1016/j.tifs.2021.07.022

Warwick, C. (2012). Studying users in digital humanities. In C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital humanities in practice* (pp. 1-22). Facet Publishing.

Waters, D. J. (2022). The emerging digital infrastructure for research in the humanities. *International Journal on Digital Libraries*, 24, 87-102. https://doi.org/10.1007/s00799-022-00332-3

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37(1), 3-15. https://doi.org/10.1108/eb026702

Zakaria, S., Grant, J., & Luff, J. (2021). Fundamental challenges in assessing the impact of research infrastructure. *Health Research Policy and Systems*, 19(1), Article 119. https://doi.org/10.1186/s12961-021-00769-z