DARIAH-EU
Digital Research Environment
for the Arts and Humanities

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Report on the DARIAH Digital Practices in the Arts and Humanities Web Survey 2016.

DARIAH VCC2
DiMPO

FINLAND

Report on the DARIAH Digital Practices in the Arts and Humanities Web Survey 2016.

Author: Inés Matres,
University of Helsinki

FINLAND

This chapter reports on the findings of the Finnish sample of the DARIAH web survey on digital practices in the Arts and Humanities 2016.

OVERVIEW OF CONTENTS

Introduction 3

1. Characteristics of the Finnish sample 4

    1.1 Disciplines 4

    1.2 Professional affiliation and status 6

    1.3 Years in research, age and gender. 7

2. Digital research practices in Finland 9

    2.1 Digital access to research materials 9

    2.2 Interest, use and purpose of digital research methods or tools 11

    2.3 Specific digital methods or tools used 12

    2.4 Collaboration and communication during the research process 18

    2.5 Publishing and dissemination of results 20

    2.6 Use of databases 22

3. Needs for digital research infrastructures in Finland 23

    3.1 Improved digital access to research resources or data 24

    3.2 Training on digital methods and use of tools 24

    3.3 Digital technical infrastructures and support 25

    3.4 Collaboration and networking 26

4. Conclusions 27

    4.1 Access to digital publications and research data 27

    4.2 Fit-for-purpose digital humanities methods and tools 28

    4.3 Importance of human infrastructures 28

5. References 30

6. Annex I Glossary of methods and tools 31

7. Annex II DARIAH questionnaire Finland 33

Introduction

This report is based in 239 responses to a questionnaire distributed by the University of Helsinki in October 2016. Students and researchers in eight Universities and several cultural heritage organisations in Finland were invited to have a say in digital research practices and needs for digital infrastructures.

The purpose of this questionnaire is twofold:

1) The University of Helsinki is currently hosting HELDIG – The Helsinki centre for digital humanities[1]. In the following years this centre will develop as a platform for addressing the challenges of digitalization, based on multi- and interdisciplinary research and collaboration in humanities, social sciences, and computer science. The results of this questionnaire will be used to inform HELDIG of current digital research practices and needs.

2) This Finnish questionnaire in digital research practices and needs for digital infrastructures in humanities and related fields, is based on a core survey designed and developed by the working group "Digital Methods and Practices Observatory[2] (DiMPO)" in the framework of DARIAH, the European network for digital research infrastructures in the arts and humanities. The results of this survey will be aggregated to a wider European initiative that has collected some 2.200 responses from around Europe (France, Germany, Austria, Poland, Switzerland, Ireland, Greece, Slovenia, Serbia and Lithuania).

The questionnaire[3] and this report are structured following the research data LifeCycle identified in digital humanities (Puhl et.al 2015: 43). In accordance with this LifeCyle, respondents report on use and interest to use digital methods and tools to: find, capture, organise, enrich, analyse, visualise, disseminate and store their research assets and outcomes. To complete this information, respondents define their research field and interests, their experience and current position, as well as affiliation to an institution for research.

[1] HELDIG homepage: https://www.helsinki.fi/en/researchgroups/helsinki-digital-humanities/ (accessed 12.12.2016)
[2] More information on this is found at the end of this report as Annex.
[3] The questionnaire is available at the end of this report as Annex.

1. Characteristics of the Finnish sample

1.1 Disciplines

In Finland this questionnaire was targeted at arts and humanities but was open as well to researchers from other disciplines, if they thought their research was related to the humanities. Based on the Faculty structure of the University of Helsinki, responses were first organized in broad disciplines as displayed in the figure below.

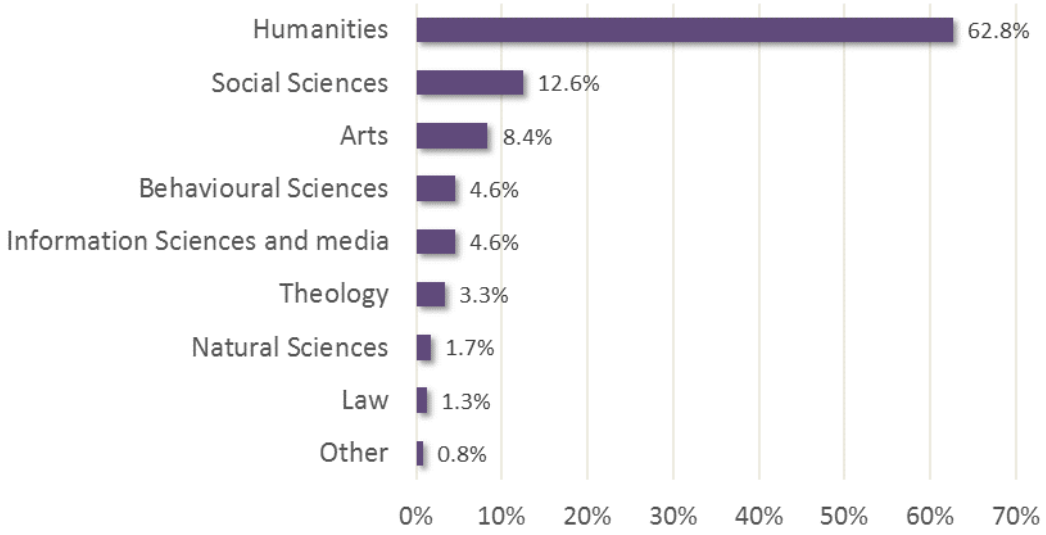


Figure 1 Finland sample, what is your broad research discipline? (N=239, 1.3%=3 responses)

Respondents in arts and humanities (including theology) sum 74.5% of respondents (r=178). This document reports on digital practice in humanities rather than any other scientific discipline. However, a significant amount of respondents (36%) fluctuate between fields, therefore all responses are included in this report.

To identify the specific field of research, respondents answered in free text. The figure below represents responses after the data was carefully analysed, that is, the numbers should be taken as approximation, as some respondents identify with one or several fields.

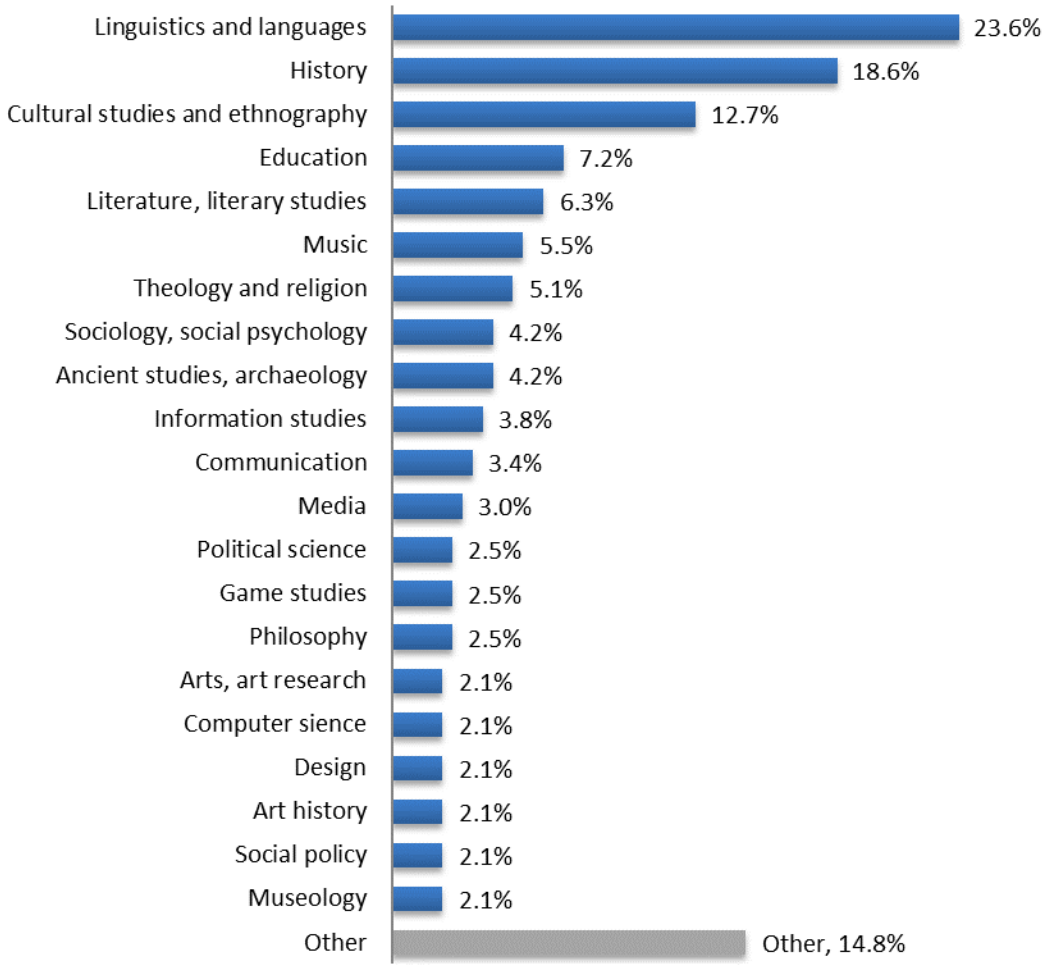


Figure 2 Finland sample, research fields, r= 237 (percent, 2.1% = 5 respondents)

A third of researchers in the arts and humanities (r=56) mention research interests in more than one field within the humanities, a few (r=16) are related to disciplines outside the traditional humanities. Combinations with linguistics, history, ethnography, education and media were the most common among humanists.

A fourth of researchers outside the humanities (r=14), mentioned research interest in more than one field. Here, information science and linguistics were mentioned above others.

Some interesting multidisciplinary research interests among respondents were:

- Text research, particularly the study of ideological text
- English, linguistics, historical sociolinguistics, corpus linguistics (and specifically, the development of visual and statistical methods in collaboration with computer scientists)
- Literature research collaborating with physicists to develop a technological framework for determining 'maleness' in big data.
- Museology, community heritage, heritage crime, metal detecting and archaeology
- Art pedagogies in photographic art
- Music, technology, and learning, learning in video games and social networks

1.2 Professional affiliation and status

Most respondents undertake research at a University institution as indicated in the figure below. The questionnaire was sent via email to researchers (master and doctoral students) in eight Universities in Finland as well as research personnel in three of these Universities in total. It was then distributed to museums, academic libraries and archive sector using sectorial mailing-lists and social media.

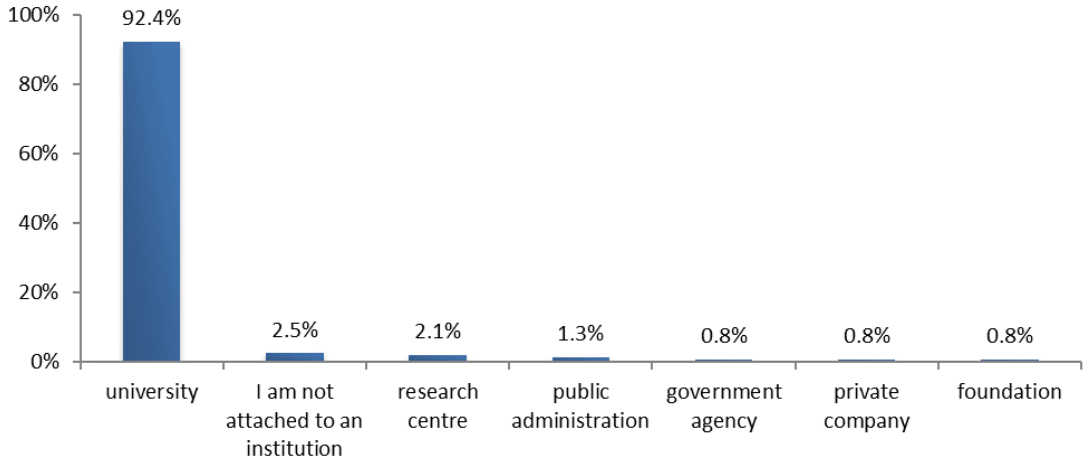


Figure 3 Finnish sample, Affiliation, (r= 237, 0.8% = 2 respondents)

Regarding occupation, the most numerous group of respondents were doctoral students (r=83) followed by a similar number of master students, professors or senior researchers, postdocs and assistant researchers or lecturers (each group r= 30-35). The other 22 respondents undertake research independently or are otherwise employed as displayed in the figure below:

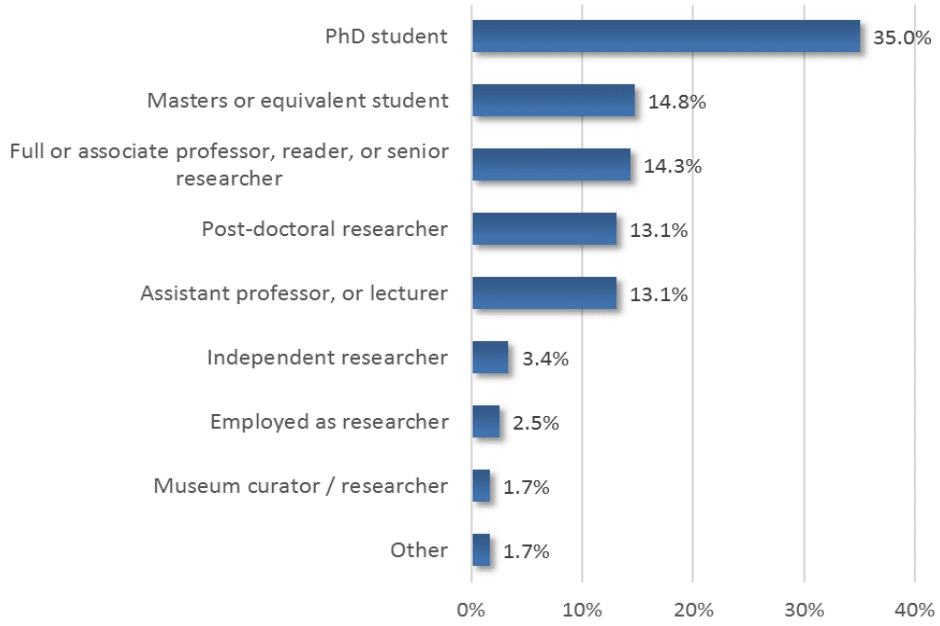


Figure 4 Finnish sample, Professional status. (r= 237, 1.7% = 4 respondents)

1.3 Years in research, age and gender.

In this section the trajectories in research of respondents, their age and gender are analysed. Here it is interesting to compare age-groups with overall trajectory and to identify possible gender gaps in research disciplines.

The largest group of respondents are researchers who have spent 3 years or less doing research, that is 4 out of 10. This group is followed by another large group of respondents with more than 10 years of experience (3,3 out of 10).

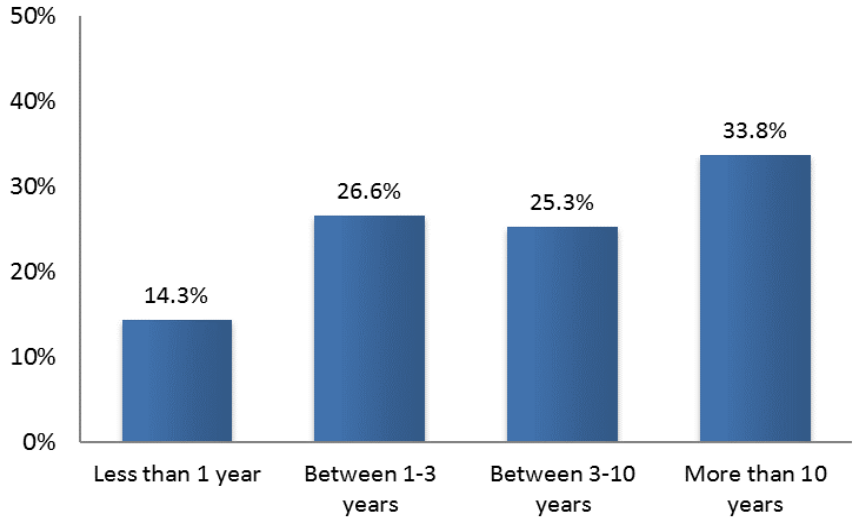


Figure 5 Finnish sample - Years in research, r= 237, percent (5% = 12 respondents)

In the figure below, the age distribution of respondents is displayed. Keeping in mind that the largest respondent groups were doctoral and master students, it is no surprise that the largest age group of respondents is in their 30s (3 out of 10).

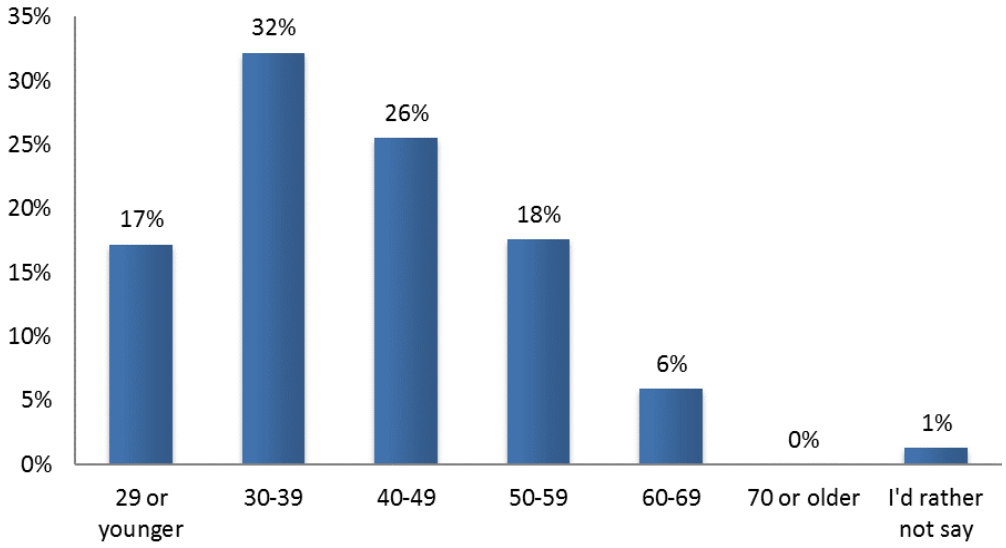


Figure 6 Finnish sample, Age, r=238 (percent, 1%= 3 respondents)

One could easily assume that the older researchers are, the more years of experience they have. This applies for most age groups, except for the 30 year-olds. In this group surprisingly, the same number of respondents have less than a year experience in research and more than 10. A

possible explanation for this heterogeneity is that some in this group might have considered the years in University as a whole when responding to this question.

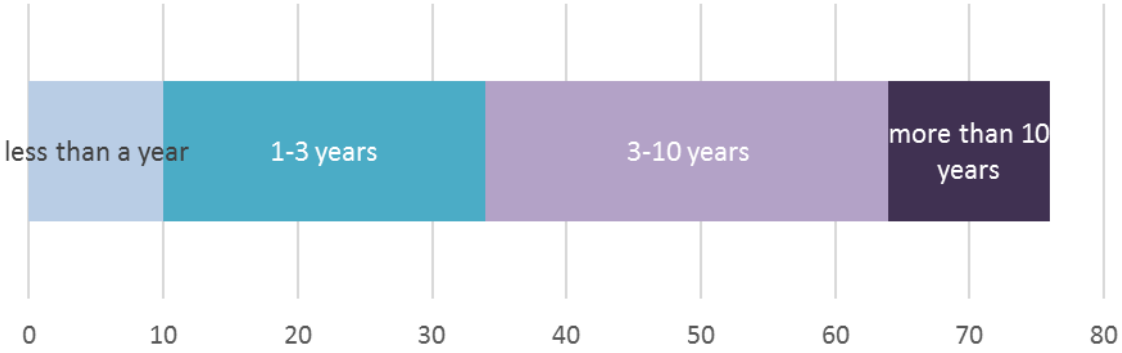


Figure 7 Finnish sample, segment ages 30-39, How long have you been working on research (r=77)

Some respondents criticized the international core survey, for only regarding a binary gender representation. Preliminary European data of this survey show that a slight majority (56%) of researchers in the arts and humanities are women. Respondents to the Finnish questionnaire evidence a slightly bigger majority (62%). In the university statistics collected by Statistic Finland, Humanities is not differentiated from Education, but is separated from cultural studies. Taking into consideration all three fields, also well represented by respondents of this questionnaire, the gender proportion is similar, as displayed in the figures below.

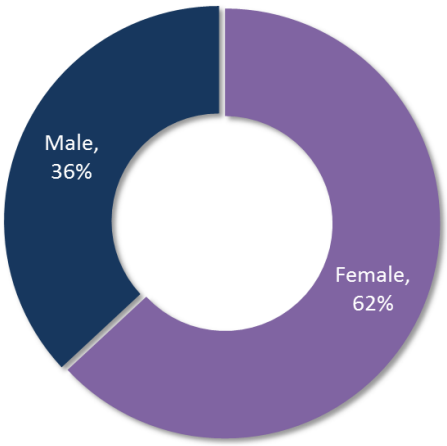


Figure 8 Finnish sample, Gender (r=233)

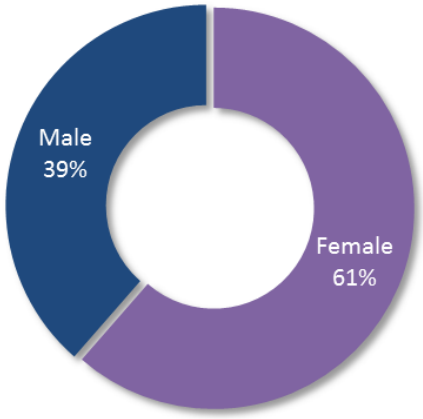


Figure 9 Gender, Enrolled university students in Humanities, Education, Cultural studies (Statistics Finland 2015).

2. Digital research practices in Finland

2.1 Digital access to research materials

Without exception, all research materials are accessed in digital format or using a digital device; though analogue access to books, journals and archival holdings is quite high. In the figure below the most frequent types of materials are listed (articles, books, images, video, audio, archival holdings and maps). Noteworthy from the responses, is that at least 3 out of 10 respondents use a mobile device to consult articles, image material and video material.

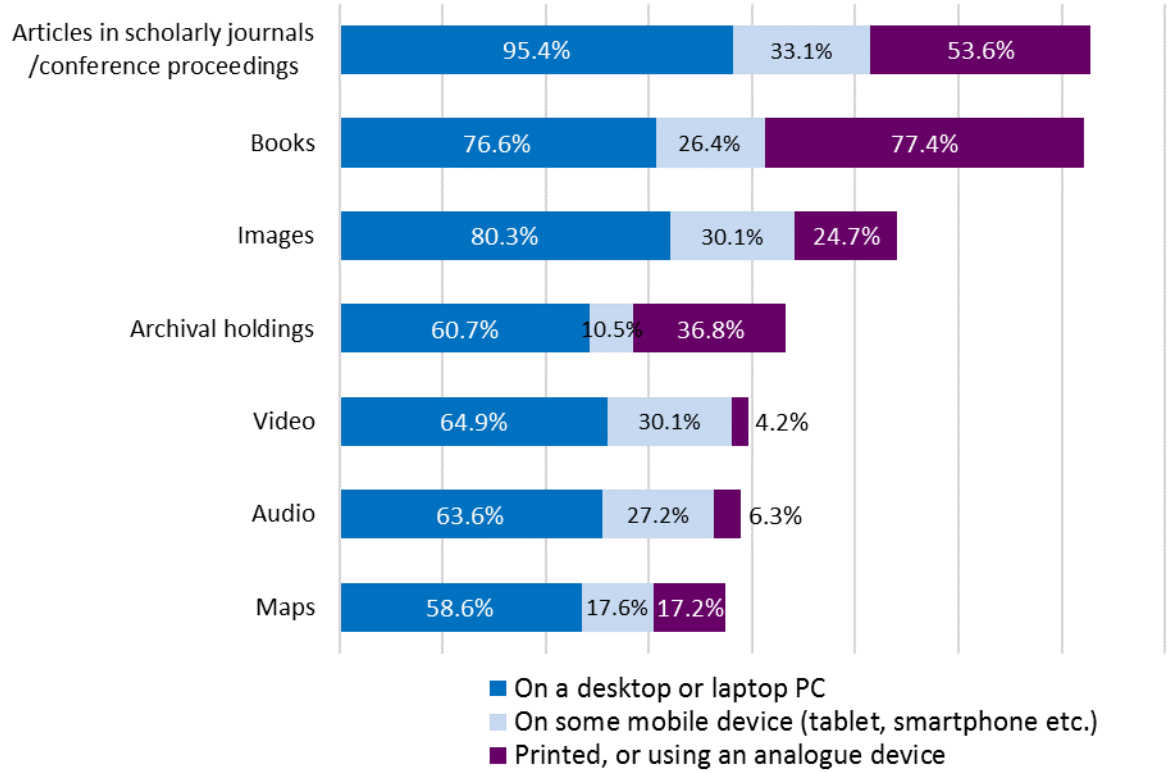


Figure 10 Finnish sample, Use of desktop PC/laptop, mobile devices and printed or analogue devices to consult research material. N=239

Respondents were asked to complete this list with other digital materials contained in their research. Things such as text corpora, databases, newspapers, social media content or survey data are as well considered research data. The illustration below is a representation of the types of data that Finnish respondents consider research material:

"Corpora (or is that included in 'archival holdings'?)"

Figure 11 Finnish sample, Tag cloud with digital research materials (r=66)

In addition to the materials, respondents also rated some statements regarding availability and licensing of digital materials used. From the responses, represented below, there is room for improvement on access to materials on the web, knowledge of the licenses and machine-readability of materials.

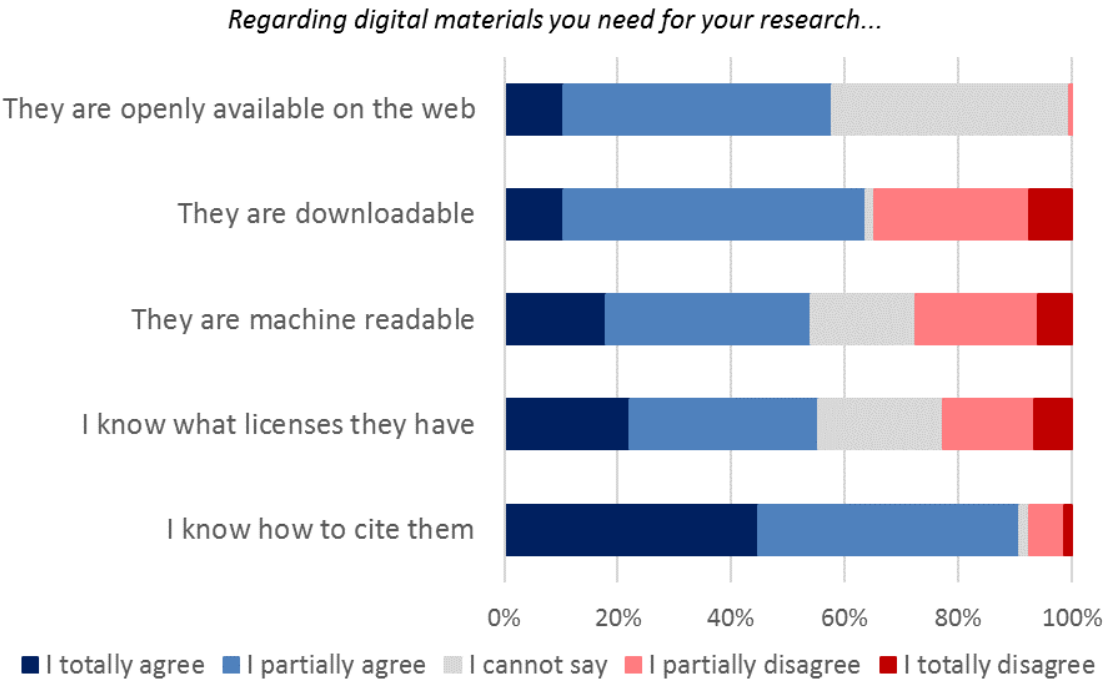


Figure 12 Finnish sample, access to digital research materials (percent, N=239)

2.2 Interest, use and purpose of digital research methods or tools

It is no surprise, that respondents to this questionnaire are active and interested in using digital methods for their research. Only one out of 40 respondents did not use nor had interest in digital research methods or tools.

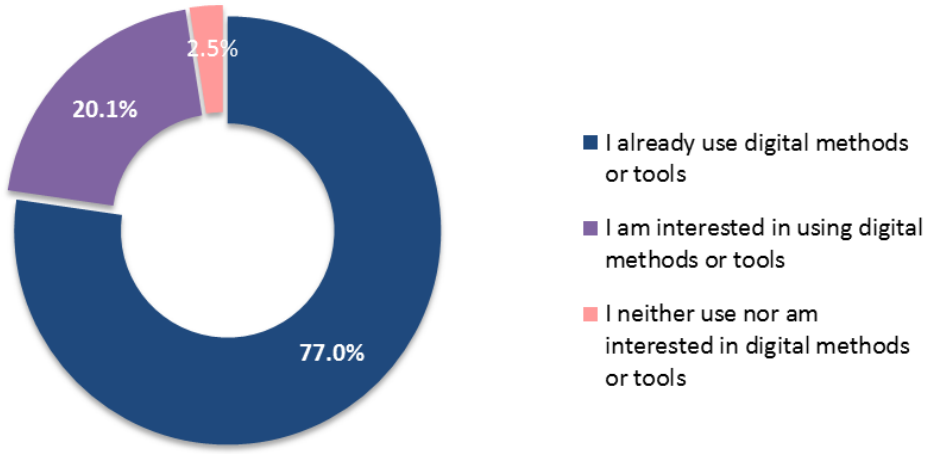


Figure 13Finnish sample, interest and use of digital methods or tools for research, Percent N=239.

As mentioned in the introduction, this questionnaire is structured having in consideration all stages of research and the possibility of making use of digital methods and tools in each of them. When looking at the distribution by purpose, most respondents use or are interested in digital methods for finding and organising research assets, followed by publishing research results and processing data. Less respondents were articulate about annotation and enrichment of research assets.

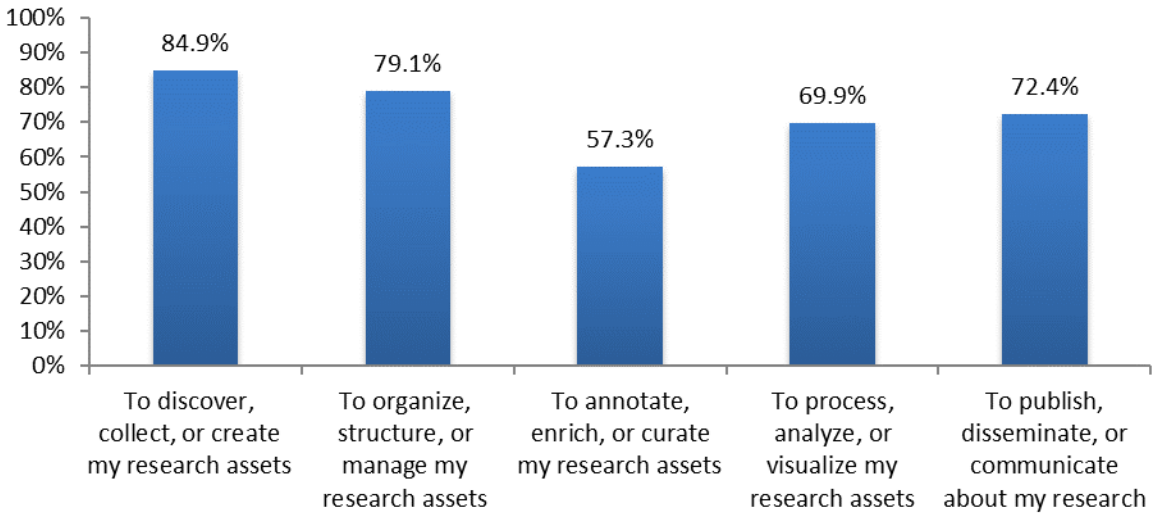


Figure 14 Finnish sample, Purpose of digital methods or tools N=239.

2.3 Specific digital methods or tools used

This questionnaire was not particularly interested in measuring the digital skills of respondents nor was it targeted to advanced digital humanities practice. Therefore, looking at the answers of respondents who described specific digital methods they apply in their research, practices range from search habits in catalogues to creating 3D models. This chapter analyses these digital methods and tools based on their purpose, as displayed in the previous figure. For a more complete overview of digital methods in the arts and humanities an annex has been added at the end of this report.

> *"I use practically anything that is in my computer. From processing of texts to video and database editing, from programming to writing emails"*[4]

Overall most respondents described use of, or interest in digital methods for finding sources and acquiring research data; this followed text editing and processing, digital publication and management of references and bibliography. Many mentioned data analysis methods such as statistical analysis or qualitative data analysis. Finally, less but some practice on digital storage or archiving is found among responses. The illustration below is a representation of what respondents relate with digital research methods.

Figure 15 Finnish sample, Tag cloud with digital research methods (r=66)

[4] *Ihan kaikki mitä tietokoneesta löytyy. Tekstinkäsittelystä videoihin ja tietokantoihin, ohjelmoinnista sähköpostiin.*

2.3.1 To discover, collect or create research assets

Researchers use a combination of digital sources to find publications (books, articles or journals) or other digital research data. Slightly more often they use generic- or scholarly search engines and online journals, than a library OPAC. 7 out of 10 uses often or very often digital archives or data repositories for finding research data. By contrast, only 4 out of 10 respondents seeks or discovers research assets on a social media platform often or very often[5].

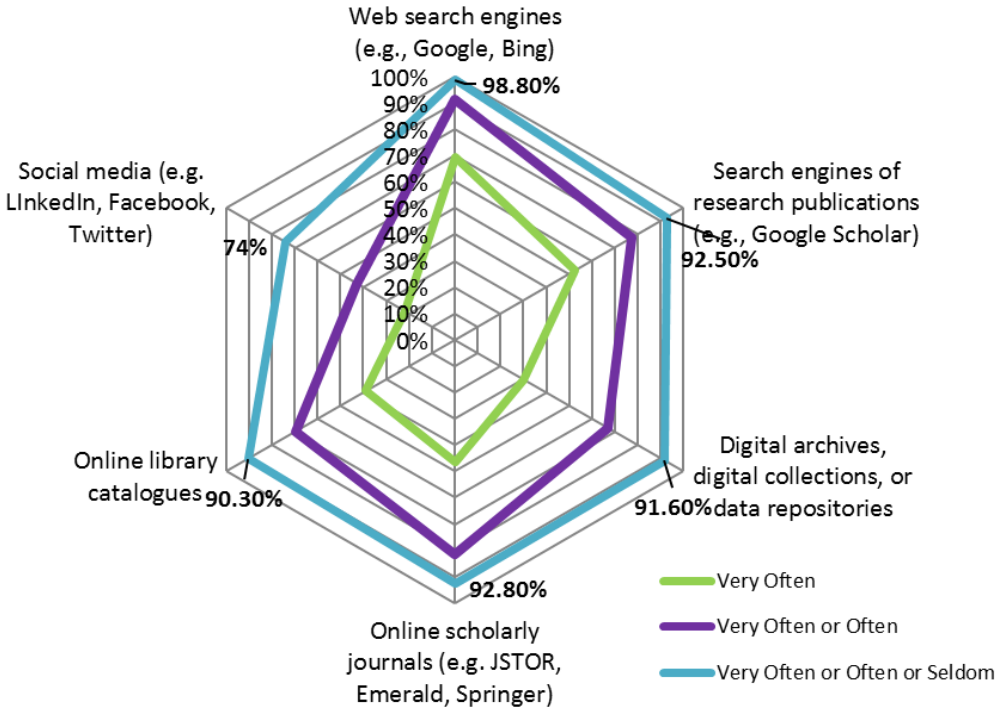


Figure 16 Finnish sample, Frequency of use of services. N=239

To find bibliographic sources (books, articles, journals) some respondents have mentioned OPACs provided by Universities (*Nelli HY, Helios UTA*…), but more often other portals such as *Google Books, Academia.edu, ResearchGate, Frontiers, Ebscohost* or *ACM* are used. Fine arts scholars frequently use *Research Catalogue*. Other than traditional bibliographic sources, other research assets have been collected through specialized digital catalogues or databases: *CSC, Wordsmith* (for corpora or text resources), *National archives, RITVA, JAPA, Ylen äänitearkisto, elonet, IMDB* are some examples of repositories with data used in research.

> *"On internet databases I find better images of ancient coins and inscriptions than I would find in already old and black-and-white paper publications. It is also hundred times easier to find anything"*

[5] Respondents in this questionnaire were given a reference: Very often= more often than once or twice in one month; Often= once a month or less but regularly; Seldom= I have done it before; Never= I have not done this at all.

Data generated by participants or web users represents a significant amount of information for research. This data can be generated purposefully via web questionnaire (*Webropol*, *Google forms*), or exist regardless of the research but be used for such, e.g. social media content, logs or forum data.

Audiovisual data is also recorded by many researchers, most respondents mention audio recordings, but some use other visual recording / capture methods (e.g. *Movavi, Panono 360*).

2.3.2 To organize, structure or manage research assets

> *"I use the online space offered by the university to store my assets and manage my citations. I would like to have as well, applications to archive and organise email (used as research asset)"*[6]

In order to organize, structure and manage research assets, most respondents use some sort of text editor or spreadsheet application. Respondents using digital storage space, prefer using generic storage platforms (*Dropbox, Google drive, Microsoft OneDrive*) than space offered by an institution for keeping their research assets.

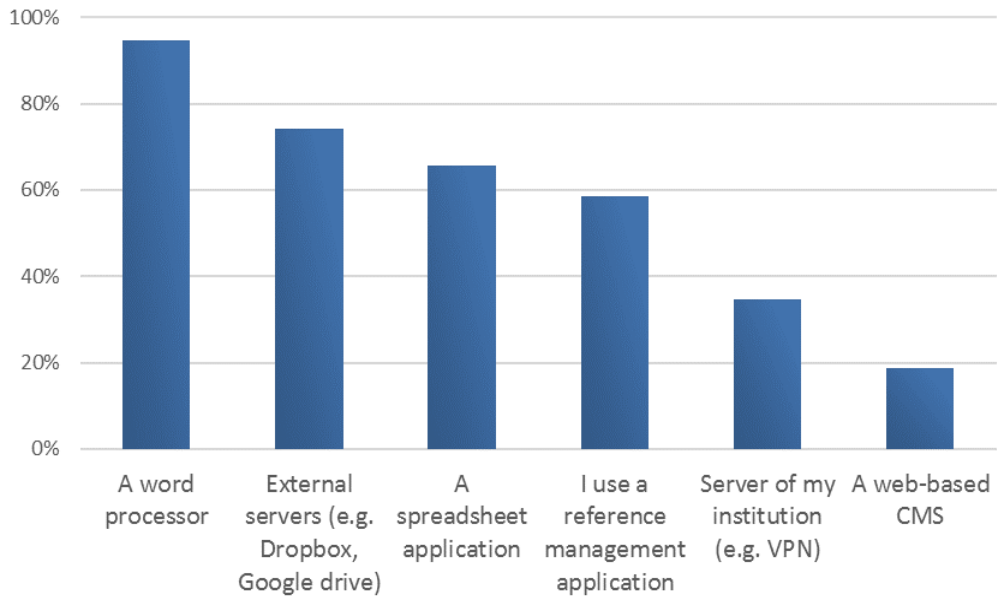


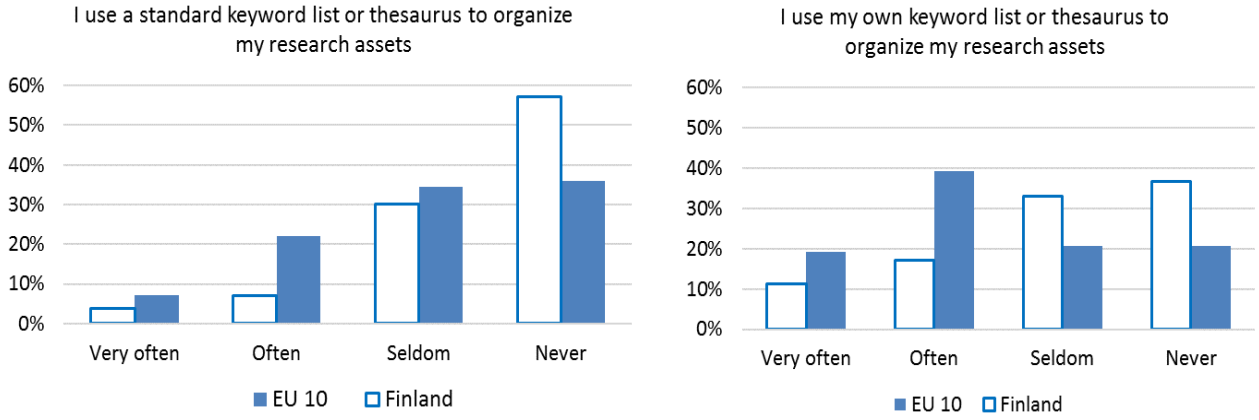
Figure 17 Finnish sample, Applications used to organize and manage research assets (N=239)

When asked about specific management methods or tools, diverse bibliographic management systems were mentioned: *RefWorks*, *Zotero*, *Mendeley*, *Paperlite*, *Bookends*. Very few respondents are active developing bibliographic management systems as part of their work (*Bibframe*, *Bibtex*, *Latex*).

---

[6] *Käytän yliopiston tarjoamaa verkkotilaa aineiston tallentamiseen ja viitteidenhallintaohjelmaa. Haluaisin käyttööni ohjelmistot sekä tutkimuaineistona käytettävien sähköpostien arkintointiin ja jäsentelyyn.*

To organise research assets, very few respondents use vocabularies or thesauri. When thinking of doing research that aims at producing digital results, the use of shared standards and vocabularies increase the chance to make the research re-usable in the future or by others. In this, Finland differs significantly from the cumulative data available from this international survey.



Comparison Finland / EU 10[7], Scholarly activities, percent, EU 10 N=2.122, Finland N=239

At the end of the research cycle, issues arise regarding archiving, long-time preservation, and persistent identification of research data. When asked about archiving or destruction of research data, 3 out of 10 respondents have taken measurements, while 1 out of 10 admits that parts of the research will be destroyed. None of the respondents provide detail on special measures taken or tools used for preserving their research data, other than doing back-ups, in external and in multiple devices, as well as using space offered by their institution.

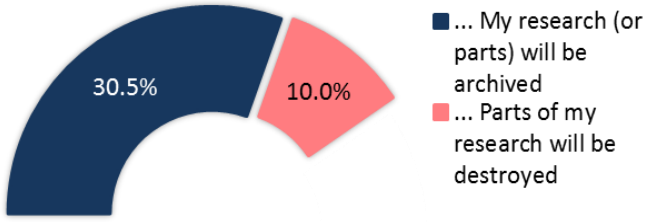


Figure 18, Finnish sample, Regarding the outcomes of your most recent research... (N=239)

2.3.3 To annotate or enrich research assets

In humanities research there is a thin line that separates annotation and enrichment from actual processing and analysing research assets, so this section focuses on methods and tools used for annotating and cleaning-up data.

[7] Data from: France, Germany, Austria, Poland, Switzerland, Ireland, Greece, Slovenia, Serbia and Lithuania.

"Artefacts of participation – physical objects such as post-it notes, big sheets of paper, etc. I like to photograph the big sheets for later reference, and type all of the post-its"

For text annotation, most reference management systems are equipped with digital note taking aids, other tools mentioned by respondents were Evernote, OneNote, Endnote. A digital tool was mentioned to facilitate brainstorming (SimpleMind).

A third of respondents often or very often digitise themselves research assets, but as shown below, much less use a standard thesauri or vocabularies to annotate their digital assets.

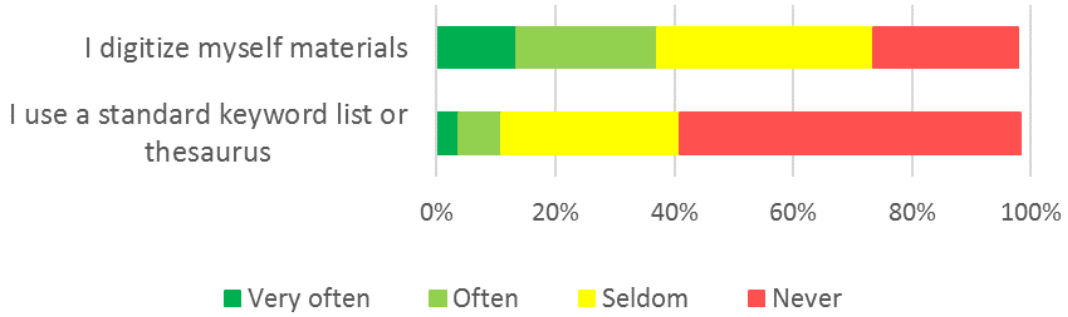


Figure 19 Finnish sample, How often do you perform the following activities (N=239)

2.3.4 To process, analyse, or visualize research assets

In accordance to the methods as classified by DiMPO, this section summarises experiences with digital methods or tools that facilitate content analysis, network analysis, relational analysis, spatial analysis, structural analysis, visualisation, interpretation, contextualisation, modelling and theorising.

"My main work tool is TUSTEP; first I digitise manuscripts from the Middle ages, then I compare different text versions with a machine."[8]

Linguists are in the forefront using digital methods and tools for analysis, topic modelling or quantitative text transformation. For processing larger text corpora different tools are mentioned: *Voyant, NLTK, BNCweb, CQPweb, types2, FiCa, Khepri, Text variation explorer, Wordsmith tools, AntConc.* Other tools mentioned among linguists, without specifying, were translation tools and web-dictionaries.

A large group of respondents works with data collected by themselves either recording it, using web surveys or other collection methods. This data (textual, audio-visual) is rarely used as such for research; it either has to be transcribed or edited. In addition to generic text editors and media players, some mention specialised software that facilitate transcription and media edition

[8] *Pääasiallinen työvälineeni on TUSTEP (Tübinger System der Texverarbeitungsprogramme): 1.Keski-aikaisten käsikirjoitusten digitointi.2. Eri tekstiversioiden koneellinen vertaaminen.*

(*Inqscribe, Elan,* or *AVS Video*). For facilitating qualitative data analysis based on transcriptions respondents have mentioned to use *NVivo, Atlasti, MAXQDA* and *Nudist*, or specialised for audio or phonetics (*Praat*, *Audacity*).

Creative tools by Adobe (*Illustrator, Photoshop*) are used to process images by few.

Respondents interested in larger amounts of data or quantitative analysis of statistics or other metric data make use of software such as *SPSS, SAP, R* and *YT*. Among this group, knowledge of programming languages was common, languages and tools used are *R, Pearl, Python* or *OxygenXML*.

> *"Database software, digital map-making tools, online toponymical databases, text mark-up languages, 3D visualization and modelling"*

A few respondents work in more technical projects: developing immersive educational environments using AR and AV, or doing research in archaeological sites. These mention 3D modelling tools (*AutoCAD, Sketchup*), tools for creating augmented reality environments (*Aurasma, Junaio* and *Wikitude*) or to process and analyse geographical data (*Idrisi, ArcGIS*).

2.4 Collaboration and communication during the research process

"My research has its own entity, which I do alone, but the project provides an opportunity to cooperate with other researchers, e.g. In the form of joint publications and seminars"[9]

All throughout the research process, there are scholarly activities that demand others and need that we undertake activities beyond our computer screens. How digital environments and tools are used to support communication is displayed in the figure below. Here we find some indicator that in Finland collaborative research practice is mainly undertaken in the framework of projects. Also, when communicating, social media channels are places to find and keep in touch with collaborators, research participants or colleagues abroad.

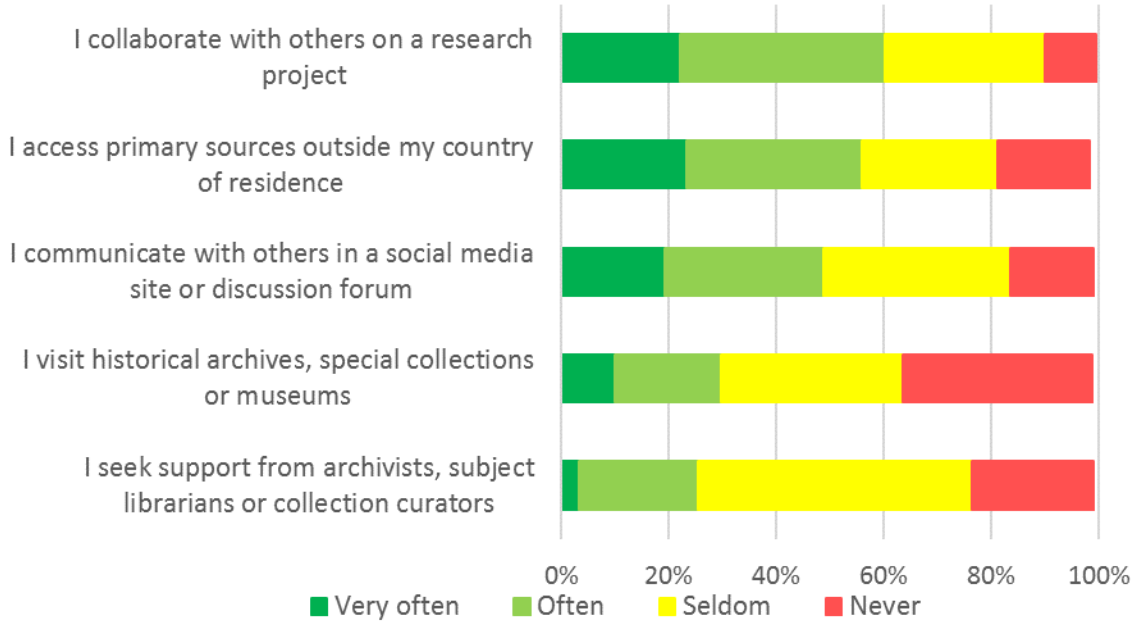


Figure 20 How often do you perform the following activities as you work on your research? (N=239)

When asked to describe more in detail forms of collaboration, some 70 respondents gave some details. Most collaborative projects mentioned are in the framework of larger national projects or facilitated by national funding bodies (e.g. Finland Academy, TEKES, Lapland Dark Heritage, Sibelius Academy, LINE Learning Interactive Environments). Though some people mention being in contact with international researchers, few are active in ongoing international projects (e.g. Erasmus+, COST Action).

Most projects are located as interdepartmental collaboration or between several universities, some mention other sectors (private companies or administration), fewer projects mention involvement with cultural heritage institutions (National archives, National museums, Finnish

[9] *Oma tutkimukseni on oma kokonaisuutensa, jota teen yksin, mutta hanke tarjoaa mahdollisuuden yhteistyöhön muiden tutkijoiden kanssa esim. yhteisjulkaisujen ja seminaarien muodossa.*

Literature Association) and only few researchers active in projects mention collaboration with peers from the same discipline.

Figure 21 If you are involved in a collaborative project, provide some details about it (r=70)

Collaborative spaces also exist in digital form; some respondents share their experiences in the open question about methods used. Email is the most often mentioned form of digital communication, to the point that some are interested in simple ways to include them in their reference management system. Other mention Skype or Skype recordings as well as digital platforms for collaborative writing or code sharing (*Google Drive*, *Overleaf, GitHub*).

Other ways of sharing research steps are possible with some of the applications mentioned in previous sections such as some reference management systems (*Zotero, Bookends, Menderley*) and publishing platforms (*Academia.edu, ResearchGate*). The following section takes a closer look at digital publishing practice.

2.5 Publishing and dissemination of results

When contemplating research outcomes in the digital age, it is relevant to defy traditional publishing formats such as books, articles or thesis. As it would be expected in the humanities, 9 out of 10 respondents have or plan to compile their research in a publication. Nevertheless, half of respondents have produced or plan alternative outcomes (presentations, exhibitions, performances…), and 2 out of 10 consider other digital outcomes for their research.

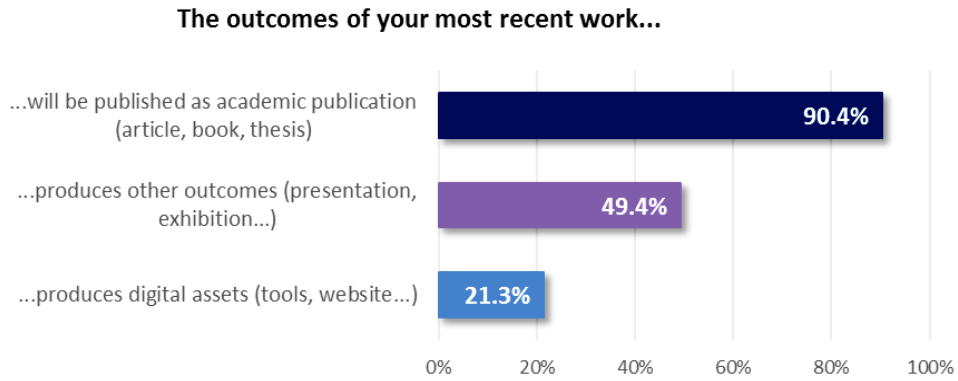


Figure 22 Finnish sample, research outcomes, percent (N=239)

2.5.1 Means of dissemination of scholarly work

At first glance at the figure below, a seemingly moderate activity in dissemination through digital channels is detected among respondents. In order to understand this, it needs be reminded that all questions to which respondents had to estimate frequency, the same time-scale was provided[10]. Being research a time-intensive activity, it is likely that outcomes will occur more seldom than often or very often.

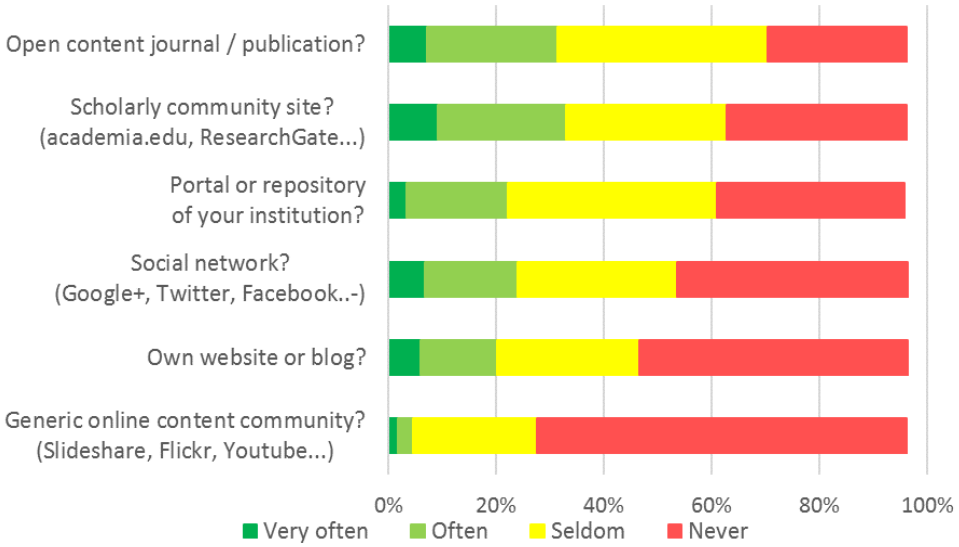


Figure 23 Finnish sample, Means of dissemination of scholarly work, r=237

[10] Respondents in this questionnaire were given a reference: Very often= more often than once or twice in one month; Often= once a month or less but regularly; Seldom= I have done it before; Never= I have not done this at all.

From the figure above, it is no surprise that the sources of information for humanists are as well the platforms chosen to publish their scholarly work: Open scholarly journals, academic community sites and their own institution. Noteworthy is that 25% of respondents have never used a digital channel to disseminate their own work, which might be explained by the number of students responding the questionnaire. Nevertheless, there are quite many (42-68%) who never use social media or generic online content communities for dissemination.

In the open question about use and interest of digital methods only few provide details of digital publication activity. *Power point* or similar tools were often mentioned. From the scholarly communities, *Academia.edu, Research Catalogue* were mentioned, as well as *Open Science Framework* and *GitHub* for code-sharing. *Twitter* and *Wordpress*, were the only generic web services mentioned by respondents.

2.5.2 Language of publication

In contrast to the European core survey, this questionnaire did not distinguish what is the native or first language of respondents. This is a conscious decision because it is not rare in Finland to find researchers with native-level command of two languages and it is a safe estimation that those publishing in Finnish are either Finns or have native-level command of the language.

The figure below evidences that English is the first language used in publications (r=114, 76%) the second being Finnish (r= 98, 41%). Overall, there were more respondents who publish in several languages (r=64, 26.6%), than only in Finnish (r=41, 18%).

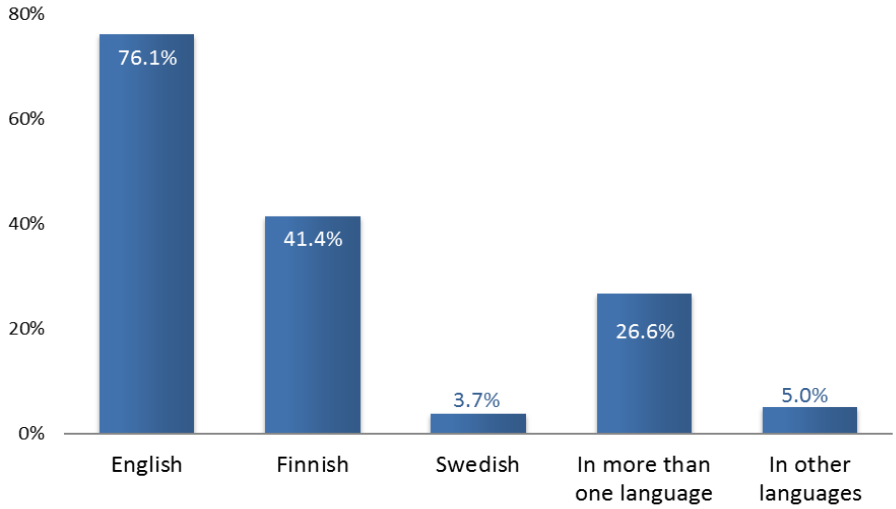


Figure 24 In which language(s) do you publish? (N=239)

Some 10 respondents publish in other languages: German and Spanish (each r=4); Russian and French (each r=2); Estonian, Portuguese, Italian and Turkish. Regional languages mentioned were Sami and Catalán. This diversity can be explained by the amount of international researchers in Finnish universities, where a native-level command of Finnish or Swedish is not always required; and a preference of Finnish researchers to publish in English in order to achieve some attention abroad.

2.6 Use of databases

In Finnish the word “tietokanta” is often used referring to catalogues or databases. This section focuses on the second meaning, thus focusing on almost half of respondents (segment, r=112), who make use of databases as a means of organising and structuring collected research assets. *SPSS, SAS, R, SQL* and *Access* were database systems mentioned by respondents.

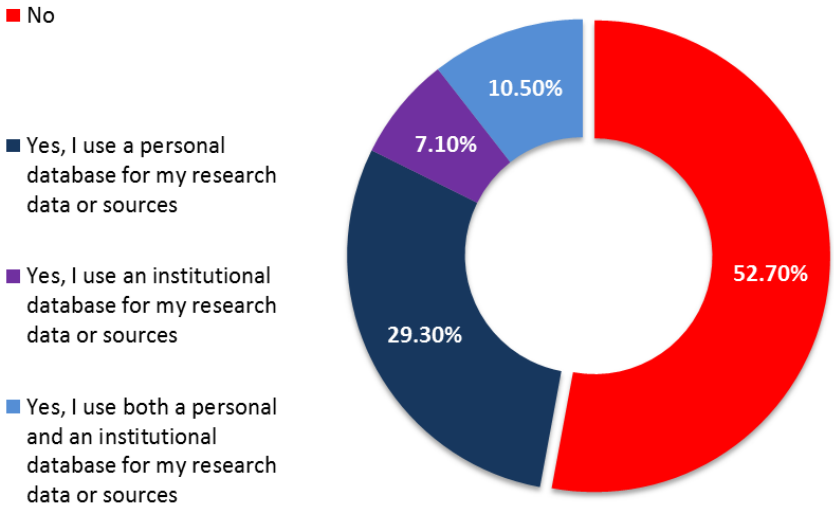


Figure 25 Finnish sample, Use of Databases, percent (N=239)

In the figure below the type of information collected in databases is displayed. From the open responses, those who manage databases specified their content: text variations, multilingual terminologies, survey data from questionnaires, terrain and volumetric data (archaeological or architectural data).

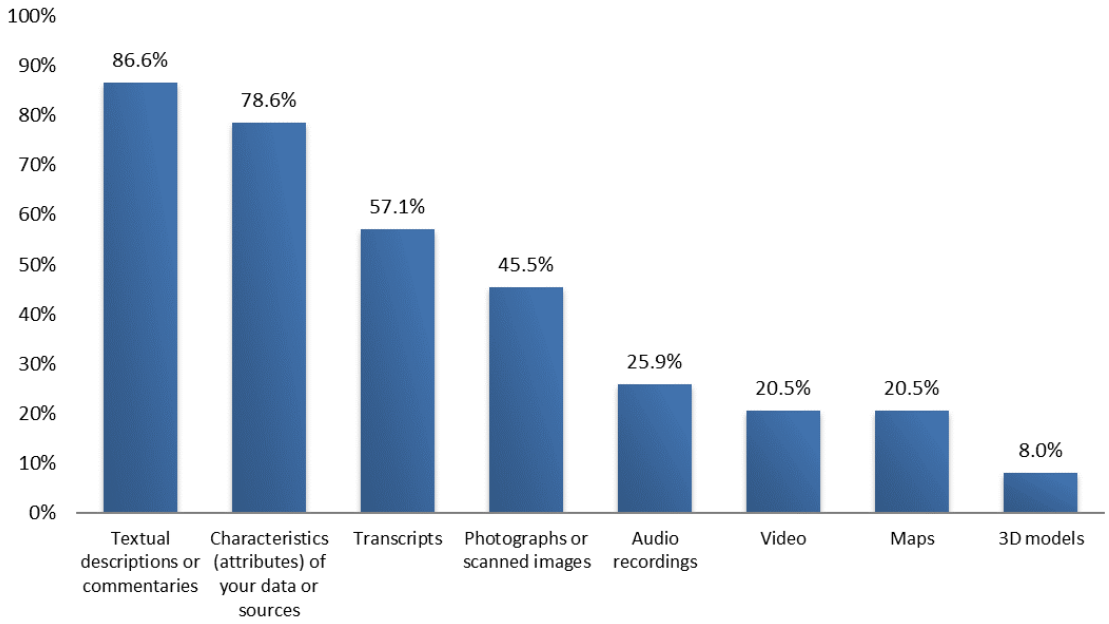


Figure 26 Finnish sample, Database contents (percent, Segment=112)

3. Needs for digital research infrastructures in Finland

In the course of the questionnaire, respondents describe not only their practice using digital methods in their research, but as well their interests and unsatisfied needs when, during research, digital materials or infrastructures come into play. This section provides an analysis of these expressions of need for training or support on digital methods or tools, desire for accessing more research or data in digital form, as well as the human and technical infrastructures important for their research.

Indisputably, the figure below points at the priority to make digital research resources or data more available, either by lowering barriers to access or by digitising research assets. It hints towards a preference among respondents, to network and collaborate with others rather than to learn how digital methods or tools can support in their research. Nevertheless, improved support and access to tools is considered important by 8 of 10 respondents.

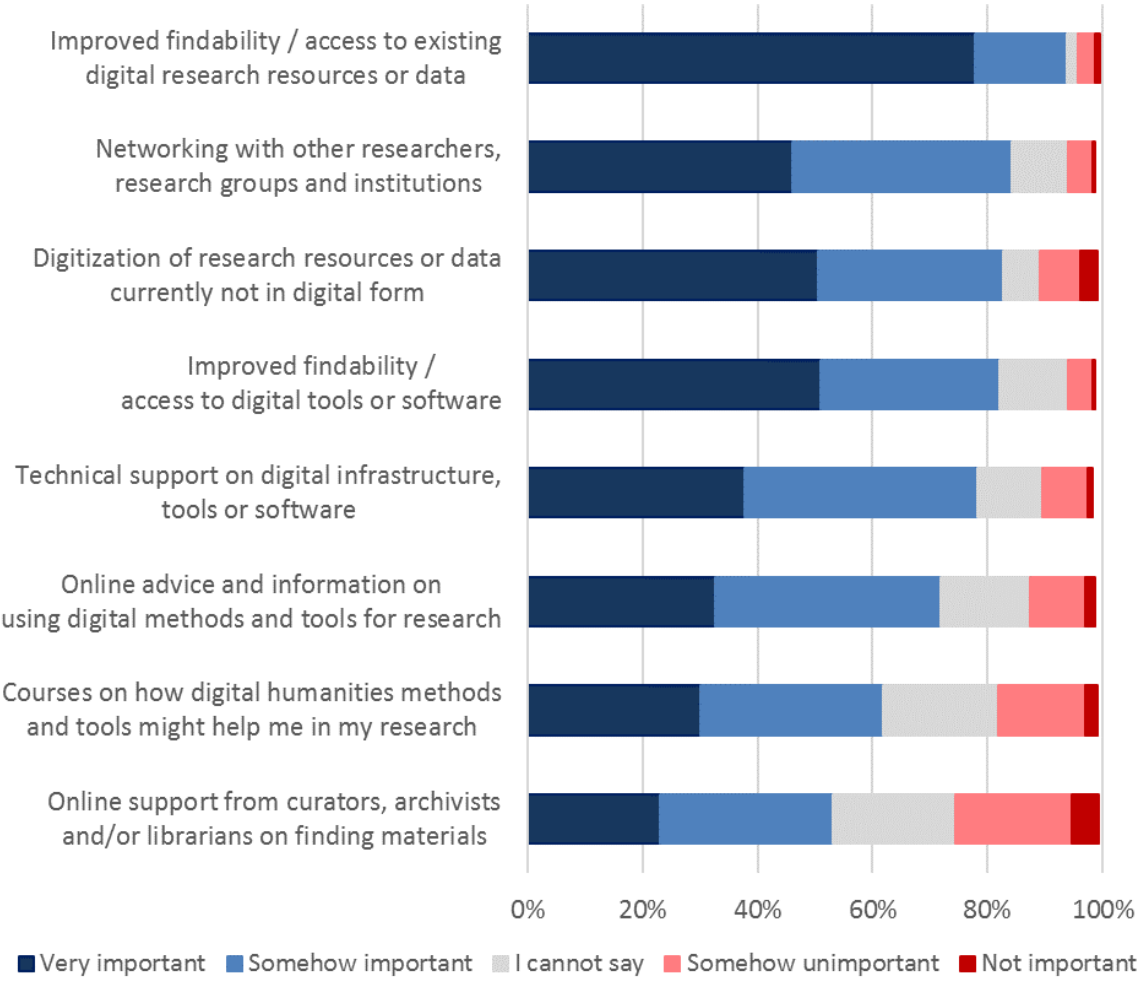


Figure 27 Finnish sample, how important is for your research..., N=237

Respondents to this questionnaire had the chance to add to this scale any digital infrastructural needs not satisfied by their home institution. Though, a majority of respondents did not expand on this topic, two out of ten specified what their priorities would be, if they had all they need.

These responses are combined here with expressions of interest left out of section 2, which focused on actual practice and use of digital methods and tools.

3.1 Improved digital access to research resources or data

Comments of limited access to digitally published research relate directly with pay-walls or insufficient content access to key scientific journals from the home research institution. Repositories like *OpenAthens* and *JSTOR* were mentioned as desired but not available. Often this cannot be resolved by research institutions alone. In Finland the FinELib consortium of Finnish Universities, Polytechnics and research institutions located at the National Library of Finland is the organ that negotiates the licensing of e-magazines, e-books, reference books and reference databases.

Other comments relate to difficulties that could be tackled by each institution, e.g. improving search functionality of library OPACs.

> *"It would be amazing to be able to enter a DOI and instantly have access to the associated article when logged into the library"*

As this questionnaire shows, bibliographic sources are not the only assets for researchers in the arts and humanities. A range of digital textual, audio and visual content, digitised or digital born content is used in research. Respondents expect their institutions to provide clear guidance on how to use or access such materials. Adequate referencing guidelines or explicit licenses are not always attached to digital data, essential for re-use in scientific context.

> *"Not having to re-invent the wheel, and instead, copy reference systems and data management plans from ready-made guidelines provided by the faculty. Reliable guidance for open data is needed"*[11]

3.2 Training on digital methods and use of tools

The statement *"Courses on how digital humanities methods can help in my research"*, was considered important by 6 of 10 respondents (see previous figure). Some wish for "Digital humanities training for doctoral students" to "personalised tutoring".

What characterises the needs of respondents towards training in digital methods, is its adequacy to their research subject. Researchers that fluctuate between disciplines might have it harder to find the right course. E.g. a sociolinguist researching gender expressions and collecting data from respondents using a web questionnaire will find only limited application in a course on survey methodologies, traditionally designed for the field of social statistics.

[11] *Tutkimusten syntyä nopeuttaisi, jos jokaisen ei tarvitse keksiä pyörää uudelleen, vaan saisi kopioida esim. viittauskäytännöt ja datanhallintasuunnitelmat valmiista malleista, jotka tiedekunta valmistaa. Luotettavia ohjeita avoimeen dataan kaivataan.*

"My research topic is multidisciplinary, so I cannot get answers in courses meant for traditional disciplines. Courses on the use of databases are suited for those whose topic fits into some existing discipline"[12]

Need for training in methods for textual analysis (big data analysis, text mining, topic modelling) have been mentioned the most. This is not only because nearly 30% of respondents have research interests directly related to linguistics or literature. Respondents from the fields of social behaviour, history or regional studies are as well interested in these methods. The increasing amount of historical materials being digitised or already accessible in digital form, as well as the amount of social and cultural expressions that today are being directly produced in digital platforms, comes with the interest for learning basic skills to acquire and analyse (at least, breaking down) this data. Some respondents are interested in training that facilitates handling relational data (databases).

"I am interested in using tools for analysing and visualizing networks and connections and to make things like membership lists easily comparable"

Beyond methods, training in the use of tools are as well on demand by respondents: Qualitative analysis tools, tools for publishing and disseminating their research online, or tools for creating infographics and graphical representations of social networks, interactions or geopolitical data.

3.3 Digital technical infrastructures and support

"We are not offered the training or technical support to compare or use software needed for research"[13]

There is a number of respondents that suspect not taking advantages of all the possibilities offered to them, and that they might be missing important information due to increasing workloads. But some of the respondents using or developing digital methods, evaluate the technical infrastructure chosen by their home institution as inadequate: Everything running on Windows instead of Linux, limited digital storage/server space, insufficient budget to buy the software needed.

"Right now, we are given limited options, based upon what people hired at the university are interested in"

Other comments touch upon being able to find somebody who can provide support, which in institutions depends a lot on the human infrastructures for research. Some respondents would appreciate more variety and opportunities for consult when deciding in using a software.

---

[12] *Aiheeni on monialainen, joten en saa vastauksia perinteisille aloille suunnatuista koulutuksista. Esim. tietokantojen käyttöön liittyvät koulutukset sopivat niille, joiden tutkimusaihe sattuu johonkin olemassaolevaan tutkimussuuntauslokeroon.*

[13] *Meillä ei ole tarjolla koulutusta/teknistä tukea tutkimuksessa tarvittavien ohjelmistojen vertailuun tai käyttöön.*

3.4 Collaboration and networking

When course offer does not satisfy or when there is limited time to invest in learning new digital methods, the best option is to find partners and collaborators. Unfortunately, very few comments were done regarding improving chances for collaboration or networking. These few comments mention assistance in carrying out time-consuming tasks (OCR, writing scripts, finding people that can help with technical things); or mention being introduced to those who have the technical skills.

"Connecting with data professionals interested in DH depends on your own initiative, you do not really know where to start"[14]

[14] *Yhteys DH-kiinnostuneisiin tietojenkäsittelijöihin on oman aktiivisuuden varassa - ei oikein tiedä mistä aloittaa.*

4. Conclusions

Before this study, there was very little knowledge about digital research practice specifically in the arts and humanities in Finland (Paju, 2016). This questionnaire aimed at uncovering this practice and take a pulse of the real needs for digital research infrastructures in the humanities.

The responses to this survey indicate that many researchers in the arts and humanities already make use of digital methods and tools. When this practice is analysed in detail, it is more likely that this occurs in initial research steps, for searching, accessing and organising research assets, and less frequently in later research stages, for analysing data, creating or communicating research results.

Digital practice is present in heterogeneous research fields, and, more importantly, the respondents' description of their practice was accompanied by expressions of interest for learning and applying digital methods and using tools more efficiently. This evidences the need for thinking of digital humanities, not as a research field, but as an infrastructure that supports researchers within their own discipline and enables researchers in between fields, to find support and a way to collaborate.

The most frequently mentioned practices and needs are here analysed in connection with the purpose of this survey, which was twofold: guiding the development of HELDIG, the Helsinki centre for digital humanities; and identifying key issues at national level that could benefit from and contribute to DARIAH, the European network for digital research infrastructures in the arts and humanities.

4.1 Access to digital publications and research data

Better access and findability of existing research resources was considered by respondents the most important factor for carrying out research. For the most part of respondents, these resources are scholarly journals and publications. Paywalls of scientific research is a topic currently being discussed internationally and in Finland, recently through the initiative *Tiedonhinta*[15] ('cost of knowledge'). In addition to commercial publishing companies, more and more open science journals are available[16] that provide better chances for scholars to find and publish scholarly work.

Research assets in the arts and humanities are also archival sources, images, videos and audio material, databases, user generated content or metrics. For these other types of digital research data, there are repositories in Finland where data can be found and published (*Tietoarkisto, Termipankki, Avoindata*, *HRI.fi*). These repositories were not mentioned by respondents; therefore, more could be done to inform about their existence.

[15] More information: http://tiedonhinta.fi

[16] Directory of Open Access Journals: https://doaj.org/

Other than publications and presentations of research results, 21% of respondents to this questionnaire are creating new non-traditional digital assets with their research. If this is planned, making and following a data management plan[17] is a best practice that ensures quality, access, re-use and preservation of this data.

4.2 Fit-for-purpose digital humanities methods and tools

Digital humanities methods and tools are suitable when trying to analyse or compare large amounts of textual data, for identifying structures in visual data or changes through time and space. Arts and humanities researchers who today want to make use of digital methods structuring and producing research assets or analysing and visualizing data, often have to look beyond their own research discipline. Some respondents have mentioned traveling abroad for finding the adequate training. If this is so, a first observation would be the need for facilitating remote access to training if needed, or allowing external credits to be computed in the study transcripts.

The good news is that already many projects have worked collecting and developing tools that fit the needs of specific methods in humanities and related fields[18].

Other possibility for collaboration is the European network DARIAH, through their Working Groups[19] contact points are organised in different areas of expertise in the Digital Humanities, here are listed those that could be of interest by the responses to this survey:

- Textual analysis (WG Text and data analytics, WG Lexical resources, WG Analysing and linking biographical data).
- Visualization of research (WG Visual media and interactivity, WG Image science and media art research).
- Digital enrichment and standards (WG Digital annotation, WG Thesaurus maintenance, WG GiST guidelines and standards).
- WG Medievalist sources (MESO)
- Specialised digital humanities courses can be found and added to the DARIAH course registry[20.]

4.3 Importance of human infrastructures

The collaborating scenario suggested by digital humanities (art and humanities researchers with computer scientists) is more likely to be happening in isolated cases and in form of organised projects at the moment, rather than it being a component of the regular activity of a researcher. An indicator of this can be found in this questionnaire by the few respondents in the arts and

[17] Guidelines on Data management plans: Finnish Social Science Data Archive http://www.fsd.uta.fi/aineistonhallinta/en/index.html or Open Science Finland http://openscience.fi/data-management-guide (both accessed 12.12.2016)

[18] Registry of digital research tools for scholarly use: http://dirtdirectory.org/ (accessed 12.12.2016)

[19] More information: http://dariah.eu/activities/working-groups.html (accessed 12.12.2016)

[20] More information: http://dh-registry.de.dariah.eu/ (accessed 12.12.2016)

humanities, involved with researchers outside the humanities (6.7%, r=16). In addition to existing collaboration ecosystems in research (projects, seminars, remote collaboration…), researchers in search for collaborators rely on the human factor (supervisors and people in their departments, or colleagues). So it is important to develop and maintain a network of people who can locally provide information on where to look for support or the adequate technical partner.

5. References

- Petri Paju, Raportti digitaalisen historiantutkimuksen kyselyvastauksista, Project report in the project: Towards a Roadmap for Digital History in Finland: Mapping the Past, Present & Future, 2016. Aalto-yliopisto, Turun yliopisto. https://digihistfinlandroadmapblog.wordpress.com (accessed, 12.12.2016)

- Johanna Puhl, Peter Andorfer, Mareike Höckendorff, Stefan Schmunk, Juliane Stiller, Klaus Thoden: "Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften". DARIAH-DE Working Papers, Nr. 11. Göttingen: DARIAH-DE, 2015 [URN:urn:nbn:de:gbv:7-dariah-2015-4-4].

- Statistics Finland / Suomen virallinen tilasto (SVT): Yliopistokoulutus ISSN=1799-0599. 2015, Liitetaulukko 2. Yliopistojen ulkomaalaiset opiskelijat ja tutkinnon suorittaneet koulutusasteen, -alan (opetushallinnon 1995 luokitus) ja sukupuolen mukaan 2015. Helsinki: Tilastokeskus http://www.stat.fi/til/yop/2015/yop_2015_2016-05-10_tau_002_fi.html (accessed: 10.12.2016).

6. Annex I Glossary of methods and tools

Digital Methods and Practices Observatory (DiMPO) is a DARIAH Working Group[21] gathering information on needs and attitudes of arts and humanities researchers in the evolving European digital scholarly environment.

Digital methods[22] connected with arts and humanities identified by this Working Group appear in all stages of research:

- Capture (capture, conversion, data recognition, gathering, imaging, recording, transcription)
- Creation (creation, designing, web development, programming, writing)
- Enrichment (enrichment, annotating, clean-up, editing)
- Analysis (content analysis, network analysis, relational analysis, spatial analysis, structural analysis, visualisation, interpretation, contextualisation, modelling, theorising)
- Storage (storage, archiving, identifying, preservation)
- Dissemination (dissemination, collaboration, commenting, communicating, crowdsourcing, publishing, sharing)

Digital tools mentioned by respondents to the Finnish questionnaire:

Name	Description
Access	Database system by Microsoft
Adobe CC	Creative tools of Adobe (Image processing, vector graphics, layout, audiovisual processing...)
Arcgis	Geographical data processing
Atlas.ti	Qualitative data analysis system
Audacity	Audio software for multi-track recording and editing
Aurasma	Augmented reality
Autocad	3D design
AVS video	Video editing
Bncweb	Username administration system
Bookends	Bibliographic and reference management system
Cqpweb	Corpus Query Processor
CSC	Data packaging
Elan	Annotations for audio and video
Frontiers	Open science publishing platform
GitHub	Web-based hosting of online services and source code repository
Idrisi	Geographical data processing, analysis and imagining
InQscribe	Transcription aid
Junaio	Augmented reality

[21] More information: http://dariah.eu/activities/working-groups.html (accessed 12.12.2016)

[22] Source: http://www.dariah.eu/about/collaboration/nedimah.html (accessed 12.12.2016)

Name	Description
Khepri	Modular View-Based Tool for Exploring (Historical Sociolinguistic) Data
MapViewer (Oracle)	Development toolkit for incorporating interactive maps and spatial analyses
Matlab	Environment used in computer engineering for machine learning, signal processing, image processing, computer vision, communications, computational finance, control design, robotics...
MAXQDA	Software for Qualitative, Quantitative and Mixed Methods Research (Texts, Surveys, Tables, Images, Media...)
Menderley	Bibliographic management software, publishing platform and research social network
Nltk	Natural Language Toolkit
Nudist	Qualitative data analysis system
Nvivo	Qualitative data analysis system
Overleaf	Collaborative writing platform and publishing tools
Oxygenxml	XML authoring and development
Panono 360	Panoramic image capture
Paperpile	Bibliographic and reference management system
Praat	Phonetics software for analysis, synthesis, visualisation and manipulation of speech
Python	Programming language
R	Free software environment for statistical computing and graphics
Sas/stat	Proprietary statistical analysis tool
Shiny	Programming language
SimpleMind	Mind mapping tool for facilitating brainstorming, idea collection and thought structuring
Sketchup	3D modelling
SPSS	Statistical analysis reference tool by IBM
SQL	(Structured Query Language) Relational database management system
Types2	Corpus tool for comparing the frequencies of words, types, and hapax legomena across sub corpora.
Voyant	Web based text analyser
Webropol	Online survey platform
Wikitude	Augmented reality
Wordsmith	Language bank, community
YT	Python tool for analysing and visualizing volumetric data
Zotero	Bibliographic and reference management system

These and more tools are available on the DiRT Directory is a registry of digital research tools for scholarly use: http://dirtdirectory.org

7. Annex II DARIAH questionnaire Finland

A Your area of research

What is your broad research discipline? *23
Please choose **only one** of the following:

- ○ Arts and Humanities
- ○ Behavioural Sciences
- ○ Information Sciences and media
- ○ Law
- ○ Natural Sciences
- ○ Social Sciences
- ○ Theology
- ○ Other: _____

Please specify your field(s) of research *

_____
(In case you are involved in multidisciplinary research, please write your speciality)

If you want, you can write here the topic of your current research
Please write your answer here:

_____

You are attached to... *
Please choose **only one** of the following:

- ○ ...a university
- ○ ...a research centre
- ○ ...a government department or agency
- ○ ...a private company
- ○ ...none of the above - I am not attached to an institution
- ○ Other: _____

If you do research in a group or collaborating with an institution, please describe shortly this collaboration (e.g. fields involved and what they contribute to your research).
Please write your answer here:

_____

[23] (*) answer required to submit questionnaire.

B Digital practice and needs in research

How do you access following materials?
Check **all** that apply:

	On a desktop or laptop PC	On some mobile device (tablet, smartphone etc.)	Printed, or using an analogue device
Articles in scholarly journals or conference proceedings			
Books			
Archival holdings			
Images			
Maps			
Video			
Audio			

Other materials you consult in digital format contained in your research not mentioned above?
Please write your answer here
_____

Regarding digital materials that you need for your research...
Please choose **one** response for each item**:**

	I totally agree	I partly agree	I cannot say	I partly disagree	I totally disagree
They are openly available on the web					
They are downloadable					
They are machine readable					
I know what licenses they have (what can I do with them and how to cite them)					

Do you use, or are you interested in using, digital methods or tools for your research?
Please choose **only one** of the following:

- ○ I already use digital methods or tools
- ○ I am interested in using digital methods or tools
- ○ I neither use nor am interested in digital methods or tools

If you use, or are interested in using, digital methods or tools for your research:
For what purpose?

Please choose **all** that apply:

- ☐ To discover, collect, or create my research assets
- ☐ To organize, structure, or manage my research assets
- ☐ To annotate, enrich, or curate my research assets
- ☐ To process, analyze, or visualize my research assets

☐ To publish, disseminate, or communicate about my research

Could you specify these digital methods or tools?

Please write your answer here:

_____

How often do you perform the following activities as you work on your research?

Please choose **one response** for each item:

	Very often	Often	Seldom	Never
I visit historical archives, special collections or museums				
I seek information or advice from archivists, subject librarians or collection curators				
I access primary sources outside my country of residence				
I digitize myself materials				
I use a standard keyword list or thesaurus to organize my research assets				
I use my own keyword list or thesaurus to organize my research assets				
I use a bibliographic management application (e.g. Endnote, Zotero etc.) to manage my citations				
I collaborate with others on a research project				
I communicate with others in a social media site or discussion forum				

(**Very often=** more often than once or twice in one month, **Often=** once a month or less but regularly, **Seldom=** I have done it before, **Never=** I have not done this at all)

How often do you use the following services for research purposes?

Please choose **one respons**e for each item:

	Very often	Often	Seldom	Never
Web search engines (e.g., Google, Bing, Yahoo)				
Search engines of research publications (e.g., Google Scholar, Microsoft Academic Search)				
Digital archives, digital collections, or data repositories				
Online scholarly journals (JSTOR, Emerald, Springer, etc.)				
Online library catalogues				
Social media sites (e.g., LinkedIn, Facebook, Twitter)				

(**Very often=** more often than once or twice in one month, **Often=** once a month or less but regularly, **Seldom=** I have done it before, **Never=** I have not done this at all)

Which applications do you use to store and manage your research assets? Select all that apply.

Please choose **all** that apply:

☐ A word processor (e.g., MS Word)

- ☐ A spreadsheet application (e.g. Excel)
- ☐ A database management system (e.g. Oracle, Access, mySQL)
- ☐ A web-based Content Management System (e.g., Drupal,Wordpress)
- ☐ A note-taking application (e.g. Evernote, OneNote, Zotero)
- ☐ Server space offered by my research institution (e.g. VPN)
- ☐ Server space offered by other services (e.g. Dropbox, Google drive)
- ☐ Some non-digital method
- ☐ Other: _____

Do you use a database to manage your own research data or sources?
Please choose **only one** of the following:

- ○ I don't use a database
- ○ Yes, I use a personal database for my research data or sources
- ○ Yes, I use an institutional database for my research data or sources
- ○ Yes, I use both a personal and an institutional database for my research data or sources

If you use a database to manage your research data or sources:
Does this database contain...
Please choose **all** that apply:

- ☐ Descriptors (attributes) of your data or sources?
- ☐ Textual descriptions or commentaries?
- ☐ Photographs or scanned images?
- ☐ Transcripts?
- ☐ Maps?
- ☐ Audio recordings?
- ☐ Video?
- ☐ 3D models?
- ☐ Other: _____

How important do you consider the following for your research?
Please choose **one response** for each item:

	Very important	Somehow important	I don't know	Somehow unimportant	Not at all important
Improved findability / access to existing digital research resources or data					
Digitization of research resources or data currently not in digital form					
Improved findability / access to digital tools or software					
Online advice and information on using digital methods and tools for research					
Online support from archivists, curators and/or librarians on finding materials relevant to my research					

	Very important	Somehow important	I don't know	Somehow unimportant	Not at all important
Courses or workshops on how digital humanities methods and tools might help me in my research					
Networking with other researchers, research groups and institutions relevant to my research					
Technical support on digital infrastructure, tools or software					

If any of the above is very important to your research, you can write here if you have any specific needs that your research institution does not provide.
Please write your answer here:

_____

C Research outcomes

In which language do you primarily publish?
Please select **at most 2** answers

- ☐ Finnish
- ☐ Swedish
- ☐ English
- ☐ Other: _____

Regarding the outcomes of your current or last research work...
Please choose **one response** for each item:

	I totally agree	I partly agree	I cannot say	I partly disagree	I totally disagree
The main aim of my current research is to acquire an academic degree (e.g. MA, PhD)					
I am currently working on an academic publication (e.g. Journal article, Book)					
In addition to written publications, my research has produced / will produce other type of digital asset					
I plan to archive parts of my research in a data repository					
Parts of my research will be destroyed after my research					

How often do you disseminate your work through...
Please choose **one response** for each item:

	Very often	Often	Seldom	Never
...an open content journal or publication?				

	Very often	Often	Seldom	Never
...the portal or repository of your institution?				
...your own website or blog?				
...a scholarly community site (e.g., academia.edu, ResearchGate)?				
...a generic online content community (e. g. Slideshare, Flickr, Youtube)?				
...a social network (e. g. Google+, Twitter, Facebook)?				

(**Very often=** more often than once or twice in one month; **Often=** once a month or less but regularly; **Seldom=** I have done it before; **Never=** I have not done this at all)

D About yourself

Country of residence *

Please choose **one**:

- ○ Finland
- ○ Other: _____

Your age:

Please choose **only one** of the following:

- ○ 19 or younger
- ○ 20-29
- ○ 30-39
- ○ 40-41
- ○ 50-51
- ○ 60-69
- ○ 70 or older
- ○ I'd rather not say

Your gender:

Please choose **only one** of the following:

- ○ Woman
- ○ Man

Your main occupation:

Please choose **only one** of the following:

- ○ Full or associate professor / senior researcher
- ○ Assistant professor / lecturer
- ○ Junior / contract-based researcher
- ○ Post-doctoral researcher
- ○ PhD student
- ○ MA student
- ○ Amateur or independent researcher
- ○ Other: _____

How long have you been working on research?

Please choose **only one** of the following:

- ○ Less than 1 year
- ○ Between 1-3 years
- ○ Between 3-10 years
- ○ More than 10 years

Thank you very much for participating in this survey!