

DARIAH Report on researchers' service needs

Jenny Oltersdorf, Markus Matoni, Carsten Thiel

▶ **To cite this version:**

Jenny Oltersdorf, Markus Matoni, Carsten Thiel. DARIAH Report on researchers' service needs. [0] DARIAH. 2016. hal-01351267

HAL Id: hal-01351267
https://hal.science/hal-01351267

Submitted on 5 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



D6.1 Report on researchers' service needs

HaS-DARIAH
INFRADEV-3-2015-Individual implementation and operation of ESFRI projects
Grant Agreement no.: 675570

Date: 22-07-2016
Version: 1.0



Project funded under the Horizon 2020 Programme

Grant Agreement no.:	675570
Programme:	Horizon 2020
Project acronym:	HaS-DARIAH
Project full title:	Humanities at Scale: Evolving the DARIAH ERIC
Partners: DIGITAL RESEARCH INFRASTRUCTURE FOR THE ARTS AND HUMANITIES CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN – KNAW GEORG-AUGUST-UNIVERSITAET GOETTINGEN STIFTUNG OEFFENTLICHEN RECHTS	
Topic:	INFRADEV-3-2015
Project Start Date:	01-09-2015
Project Duration:	24 months
Title of the document:	Report on researchers' service needs
Work Package title:	Basic Services Cluster 2: Supporting the Lifecycle of Digital Humanities – Basic Services Implementation of basic tools and services in the DARIAH infrastructure
Estimated delivery date:	31.07.2016
Lead Beneficiary:	University of Göttingen – State and University Library
Author(s):	Oltersdorf, Jenny [oltersdorf@fh-potsdam.de] Matoni, Markus [matoni@sub.uni-goettingen.de] Thiel, Carsten [thiel@sub.uni-goettingen.de]
Quality Assessor(s):	Francesca Morselli [francesca.morselli@dans.knaw.nl]
Keywords:	Basic Services, Researchers' need, Survey

Revision History

Version	Date	Author	Beneficiary	Description
1.0	June 2016	Oltersdorf, Jenny	University of Applied Sciences, Potsdam	First draft
1.1	June 2016	Matoni, Markus Thiel, Carsten	University of Göttingen	Review and addition of text sections and data
1.2	July 2016	Oltersdorf, Jenny Matoni, Markus Thiel, Carsten	University of Applied Sciences, Potsdam, University of Göttingen	Final adjustments, proof reading
1.3	July 2016	Oltersdorf, Jenny	University of Applied Sciences, Potsdam	Incorporation of comments made by HaS partners after project internal publication
2.0	July 2016	Oltersdorf, Jenny Matoni, Markus Thiel, Carsten	University of Applied Sciences, Potsdam, University of Göttingen	Final version

Table of Content

Executive Summary 5

Introduction 6

Part 1 6

1. Survey design 6
   - *1.1 Aim* 7
   - *1.2 Target groups* 7
   - *1.3 Structure* 7
   - *1.4 Discussion and pre-test* 8
   - *1.5 Dissemination and duration of survey* 8
2. Results 9
   - *2.1 Focus of Professional Activities* 10
   - *2.2 Collaboration Activities* 12
   - *2.3 Administrative and Organizational Activities* 14
   - *2.4 Research Activities* 15
   - *2.5 Software Development* 18
   - *2.6. Conclusions* 19
3. Next Steps 20

Part 2 22

4. Survey data 22
5. Appendix 25
   - *5.1 List of abbreviations* 25
   - *5.2 List of charts* 25

Executive Summary

The report on researchers' needs analyses the results of a survey conducted by the Humanities at Scale project to determine the key basic services needed by researchers in the field of Digital Humanities and those affiliated with DARIAH and its community in particular.

Nature of the deliverable	
R	Document, report
DEM	Demonstrator, pilot, prototype
DEC	Websites, patent fillings, videos, etc.
OTHER	
Dissemination level	
P	Public
CO	Confidential only for members of the consortium (including the Commission Services)
EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)

Disclaimer

The Humanities at Scale is project funded by the European Commission under the Horizon 2020 programme. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Introduction

This report is a result of the work conducted in the Humanities at Scale work package 6 "Basic Services Cluster 2: Supporting the Lifecycle of Digital Humanities Basic Services Implementation of basic tools and services in the DARIAH infrastructure". As part of work package 6 (WP6) task 6.1 investigates demands of Digital Humanities (DH) projects regarding service requirements and tools. Its aim is to evaluate, what tools and services are already in use at the present time, and which ones the target groups want to use in the future. To get this kind of information and to enable a gap analysis a survey was conducted in June 2016. Work in task 6.1 is a joint effort of University of Göttingen – State and University Library[1] (UGOE-SUB) and the University of Applied Sciences Potsdam[2] (FHP).
Part 1 of the report covers the data analysis and interpretative aspects, whereas the original questionnaire can be found in part 2.
The 1st part starts with an overview of the survey design, its aim, target groups and dissemination. It is followed by a chapter on the survey results which is introduced by the quantitative distribution of answers for each group of questions and followed by the interpretation and analysis of the statistics. The section before the last presents the conclusions for further work in WP6. The last chapter of this part discloses thoughts about next steps and open issues.
The survey itself can be found in the 2nd part of the report. The last section covers the Annex with the list of abbreviations and additional visualisations.

Part 1

1. Survey design

The surveys' objective is to reveal needs and requirements of people working in DH projects with regard to tools and services. The survey design is modulated to answer specific questions that are raised in the Humanities at Scale (HaS) WP6. The following section will provide information on the survey design. The objectives of the survey, target groups and structure are introduced as well as information about the discussion and pre-test and dissemination of it.
The survey was created with Limesurvey[3]. Limesurvey is a free and open source application to develop and publish surveys online, collect responses and create statistics.

[1] https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/has/

[2] http://www.fh-potsdam.de/projekt/project-action/show/Project/has-humanities-at-scale/

[3] https://www.limesurvey.org/

1.1 Aim

The survey was designed to answer the following questions:

- How do the target groups evaluate the relevance of providing basic services and tools in a Digital Humanities research infrastructure?
- What services and tools are already in use?
- What do the target groups wish to have or use in the future?

1.2 Target groups

The defined target group for WP6 covers traditional humanities researchers as well as technical oriented researchers and developers in the field of DH. Thus the survey addresses scholars and academics with a focus on *Humanities Research* as well as scholars and academics with a focus on *Technical Development*. Taking this into account we asked people right in the beginning of the survey in which research area they see the focus of their professional activities. In order to avoid excluding people that do not feel comfortable with the choice between *Humanities Research* and *Technical Development*, we decided to implement an "Other" option which we suggested to use in that case. By choosing "Other" respondents were allowed to add their research area manually.

1.3 Structure

The survey is divided into five question blocks. A first group of questions is designed to collect information about the scholarly background. The following four groups of questions are organised around concrete activities that are typical for the lifecycle of DH projects, namely: collaboration, administration and organisation, research activities and software development. Each specific group contains tools and services that are typically used for the activity e.g. "Services to ensure citability of data" in the group of research activities or the "Hosting of source code" in the question block on software development. The list of selectable tools and services in the survey has been derived from WP6 teams' experiences. It was discussed, refined and adjusted in several iterative steps.
The survey participants were asked to rate how important they consider the listed tools and services on a rating scale that ranges from "Not important", to "Of minor importance", to "Important", "Very important" and "I don't know". Each time a survey participant indicates that something is "Very important" to him / her an additional text field appeared to let the person specify which services or tools they already use and / or which ones they would like to use in the future. At the end of each question block an open text field allows to add activities that are not captured but belong to the specific area.
The five question blocks cover a minimum of 7 and a maximum of 35 questions, depending on the replies. As described above, the full 35 questions will be shown only to those people who indicate that something is "very important" for each single question. The average time to complete the survey takes approximately 15 minutes.

1.4 Discussion and pre-test

The survey has been presented and discussed in the context of the joint WP 5/6/7/8 meeting in Göttingen in May 2016. A first pre-test was conducted in advance to reveal unanticipated problems with question wording, instructions to skip questions, etc. Colleagues gave feedback on structure, wording and content of the questionnaire and were asked to think out loud as they were answering the survey questions to find out what the questions mean to them. A second pre-test followed the meeting in Göttingen and gave again valuable feedback which resulted in the rephrasing of some questions and refinements in a small number of answering options.

1.5 Dissemination and duration of survey

In order to address the target groups mentioned in section 2.2 properly, the survey has been disseminated via the following channels:

- Humanities at Scale project consortium[4] email list
- DARIAH-ERIC National Coordinators Committee[5] email list
- DARIAH-ERIC Joint Research Committee[6] email list
- DARIAH-DE consortium[7] email list
- DHd-Blog[8]
- CLARIN-NL[9] newsletter June 2016
- DARIAH-EU Twitter[10]
- HaS-Website[11]

These mailing lists, blogs, newsletters etc. have been selected since they cover the relevant European DH and DARIAH related communities. The survey was available online for 17 days. It was published on 26th May 2016 and closed on 12th June 2016.

[4] http://has.dariah.eu/?page_id=122

[5] http://dariah.eu/about/organisation/national-coordinators-committee.html

[6] http://dariah.eu/about/organisation/joint-research-committee.html

[7] https://de.dariah.eu/ueber-dariah-de

[8] http://dhd-blog.org/?p=6878

[9] http://www.clariah.nl/

[10] https://twitter.com/DARIAHeu

[11] http://has.dariah.eu/?p=408

2. Results

The survey has been developed as means to collect information for the work in HaS WP6. Its focus is determined to fit the concrete objectives of T6.1. As already mentioned, each time a respondent indicates that something is "Very important" to him / her an additional text field appeared to let the person specify which services or tools they already use and / or which ones they would like to use in the future. In the analysis we focussed on exactly this tools and services. The objective of task 6.1 is to collect information about requirements in the field of tools and services for DH projects and to choose two demonstrators out of the results. The focus of the presented report is certainly on the most wanted tools and services. An overview of the distribution of all tools and services per question block as well as the full survey can be found in the Annex for further analysis.
The total number of replies sums up to 121. 62 respondents answered all questions, whereas 59 people decided to quit the survey at a certain point. The analysis is based on the completely filled out surveys only (n=62). In order to avoid a systemic failure we scrutinized a number of correlations in the group of people who have not completed the survey. One of our expectations was, that we would get a lot of replies from people with focus in the field of Humanities research. Our first hypothesis was, that this group would quit the survey when asked for very specific technical evaluations as in the last question block. However, this assumption turned out to not be valid. We located the exit in the middle of the survey after the question block about "Collaboration activities". Contrary to our expectations, there is no relation between the scholarly background and the number of incompletely or completely filled out surveys.

The 59 people who quit the survey did this mainly after the second question block, which means they answered the questions about their scholarly background, main research area and quit within the question block about "Collaboration Activities". Therefore we assume that people who dropped out of the questionnaire found it either too long or not fitting into their research interest or competencies.[12] The following table illustrates that:

Answer	10	16,95%
No answer	7	11,86%
Not completed or Not displayed	42	71,19%

Table 1: You indicated that you think it is very important that collaborative and simultaneous writing and editing of texts is supported by basic infrastructure services and tools. Please let us know which services or tools you already use and / or which ones you would like to use in the future. – limited to the number of incomplete responses

One can see, that around 70% of the respondents did not answer the additional question about tools and services they already use or want to use in the future anymore. This is exactly the number of people who quit the survey at that point. Around two-thirds of the respondents who filled out the survey completely claimed a background in *Humanities*

[12] The surveys first page has been clicked 9 times without any following activity.

Research (see next paragraph). The same distribution applies to the group of people who did not completely fill it out.

2.1 Focus of Professional Activities

When asked for the area where people see the focus of their professional activities, around 68% of the respondents claimed a focus in the field of *Humanities Research*, whereas 18% see their focus of professional activities in the area of *Technical Development*. Around 14% choose the option *Other* and specified the research area on their own[13]. (see Chart 1)

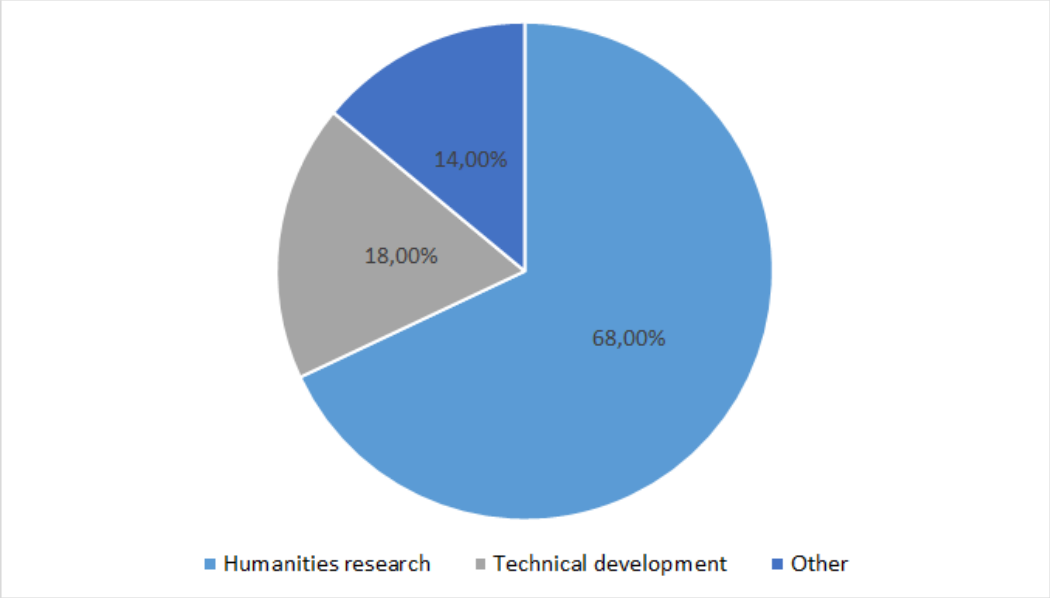


Chart 1: Distribution (in percentage) of professional activities referring to all complete replies (n=62). Question: In which area do you see the focus of your professional activities?

More than two-thirds of the respondents identified themselves as Humanities researchers. Results, in particular of the question group "software development", need to be interpreted against this background.
Interestingly, none of the respondents claimed a research identity as digital humanist. When it comes to the Other option the mentioned areas are the following: "Digital Humanities Research Support", "Library, library- & information science research", "Natural language processing", "Social sciences data management", "Provision of services", "Information science, philosophy of science, and science and technology studies", "Culture, digital culture and cultural policy related research", "I am an information scientist, interested in information use and behaviour.", "Managing infrastructure projects". Each answer was mentioned once.
People were then asked for their primary discipline of research. The given list correlates with the DFG research areas in the Humanities and Social Sciences[14]. The field of Computer

[13] All presented percentage values are rounded.

[14] http://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp

Science is not an original part of that classification but has been added from the task 6.1 team for methodological reasons.[15]
The top 5 of mentioned research areas are: Ancient Cultures (14%), Literary Studies (14%), Computer Science (13%), History (13%) and Linguistics (11%).[16] (see Chart 2)

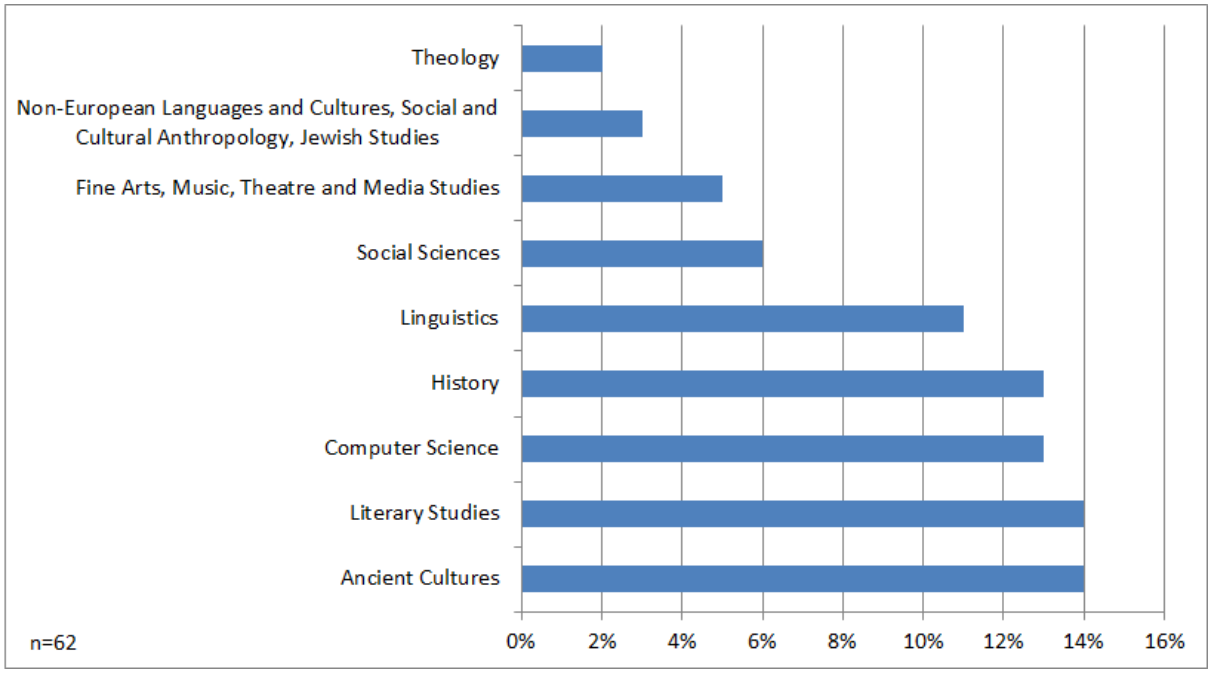


Chart 2: Distribution (in percentage) of primary research areas referring to all complete replies (n=62). Question: What is your primary discipline of research? Please select "Other" and add your discipline in the box on the right side if none applies to you. Some possible options[17] (including "Other") are not displayed in this chart, because no participant chose one of them.

The remaining 18% consider their research focus in a different area and specified that via comments[18]. Respondents come from the broad field of Library and Information Science, including data curation and research infrastructures, Art History, Natural Language Processing, Lexicography, Historical Media Studies, Epigraphy and Computational Linguistics / Artificial Intelligence. One person mentioned a background in the broader field of technical development ("e-science technologies"), one added a background in History manually, without clicking the corresponding checkbox and one indicated an affiliation to "all disciplines of Human and Social Sciences". Considering the manually added research

[15] We added this area since our target group includes people with a focus on technical development.

[16] The presented distribution does not result in 100%, since participants had the opportunity to select "Other" and to add their research area manually.

[17] Possible options were also "Economics", "Education Sciences", "Jurisprudence", "Philosophy", "Psychology" and "Religious Studies".

[18] Even if it was not asked directly and even if it is not the aim of the survey to discuss this issue, it became visible from the data that none of the respondents who added their research area manually indicated an identity as digital humanist. DH degree programs do not exist for a long time, yet. Graduates of these programs have probably not taken part in the survey and the more experienced researchers do obviously not have formed an identity as a digital humanists.

areas, one could subsume them into the applied classification of research areas. The DFG for example includes "Art History" into the class "Fine Arts, Music, Theatre and Media Studies". That disciplines like Art History or Epigraphy are separately listed may reflect national specific self-concepts as well as particular and distinguished self-concepts of individual researchers. For the analysis they were not manually assigned to other disciplines in the classification afterwards.
To sum up, the received replies mainly derive from the field of Humanities Research and are heterogeneous within that group. There is a strong group of people from the area of Ancient Cultures and Literary Studies. Even if 18% claimed a background in the field of technical development only 13% indicated computer science as the primary field of research activity. This leads to the assumption that boundaries are fluid and other research areas qualify as well for successful work in the area of technical development when it comes to DH research projects. However, no other research area generally associated with an IT career, apart from computer science, was mentioned in the survey.

2.2 Collaboration Activities

When asked for the evaluation of the importance of tools and services for collaboration activities, 51% indicated that they consider tools or services for data management in a content management system very important. This answer is followed by "Synchronization of files" and "Collaborative and simultaneous writing and editing of texts". (see Chart 3)

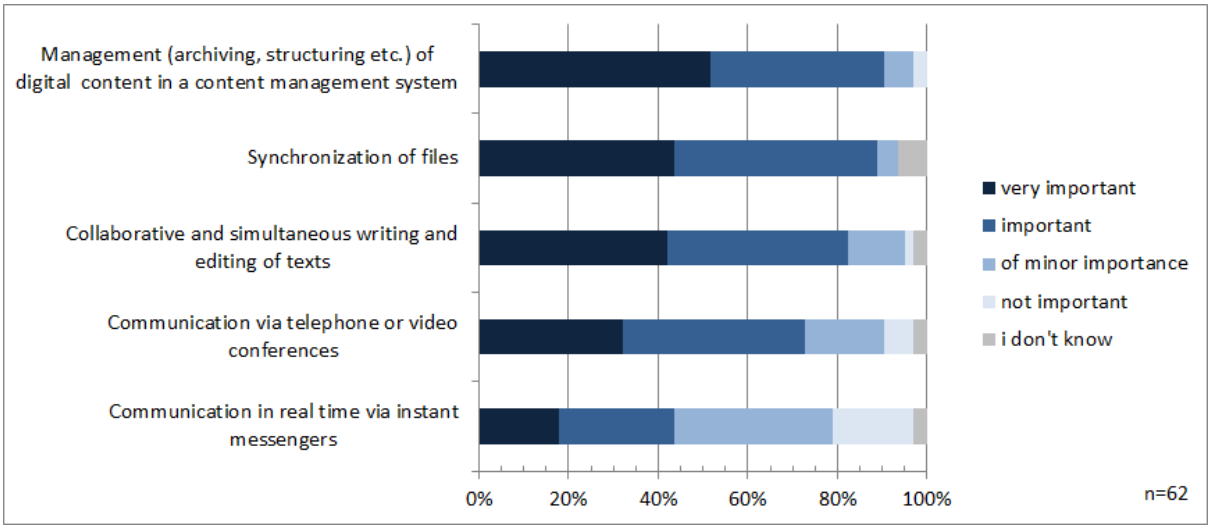


Chart 3: Distribution (in percentage) of collaboration activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services and tools for the following collaboration activities?

People mentioned institutional and national repositories as well as wiki systems and Gitlab/ Github. The research data management system EASY provided by DANS-KNAW was brought up as well as the web-publishing platform for libraries, museums, archives, and scholarly collections and exhibitions - OMEKA. The META-NET network, the research

infrastructure CLARIN, MS Access Database and the website "ArcheoData" were mentioned as well. Most often respondents referred to Dropbox, Google Drive, Typo3 and Zotero as currently used tools and services for data management (in descending order of the frequency of mentions).
The diversity of replies indicated, that the respondents have very different needs and requirements with regard to data management and have very heterogeneous ideas of the performance. The spectrum from Zotero as data management tool to the dedicated data management system EASY and the whole research infrastructure CLARIN shows the variety. Despite the heterogeneity of the answers, we assume that respondents have understood the question in principle, since no one has selected the option "I don't know" which is unique to this question block.
Interestingly, people have little ideas on what tool to use in the future, even if they do not seem to be satisfied with the current situation. The following quotes illustrate that clearly:

"I use Zotero for both bibliography and ancient documents but I would like to use another system for documents without having to create my own database (which is a lot of trouble and not very flexible especially for short projects)."

"I don't know, not using anything like it at the moment (which is not good as I loose overview)."

"We are left to our own devices for archiving (dropbox, etc.) and this is why I find this important, but do not have systematic solutions."

In contrast to the apparent desire for tools and services in the field of data management, participants articulated that they consider "communication in real time via instant messengers" of minor importance. After all, 35% of the answers relate to "Of minor importance" and 17% to "Not important". Which is a very clear result in comparison to the distribution of answers for the rest of that question block.
Additionally, we asked the respondents to comment on issues we might have missed. A total of 26 persons made use of that chance. A number of issues mentioned here referred to future question blocks in the survey - especially very detailed descriptions on software development. Advice was given by 2 respondents to avoid developing new services and/or tools but to concentrate on the proper linkage of existing ones. Interfaces and APIs are mentioned in this context to connect services and to avoid duplication of work.

Key message:
The data show that content management and files' synchronization are the two most mentioned services, even though -as the comment section illustrates - the difference between content management and synchronisation of files is not clear to all respondents. People consider very elementary functionalities like uploading documents to Dropbox as core services of a content management system whereas the structured, sustainable and manageable handling of files are the essential functionalities of a data management

system. Sophisticated functionalities like access and role management, format management, version control, indexing, search, and retrieval characterize a content management system. The synchronization of files is a hosting service, which synchronizes the data between the cloud service and several clients and is part of an advanced content management system. Since it clearly turned out from the data analysis, that by far the most often mentioned service for data management and synchronisation of files is Dropbox, it is necessary and appropriate to find alternatives or to promote existing ones and/or enable low threshold access (from a technical perspective as well as from a administrational perspective) for DH research projects.

2.3 Administrative and Organizational Activities

The third question block asked for information about requirements and needs that are related to administrative and organisational issues in the field of DH projects. The area of project management is the most often indicated topic by the respondents. (see Chart 4)

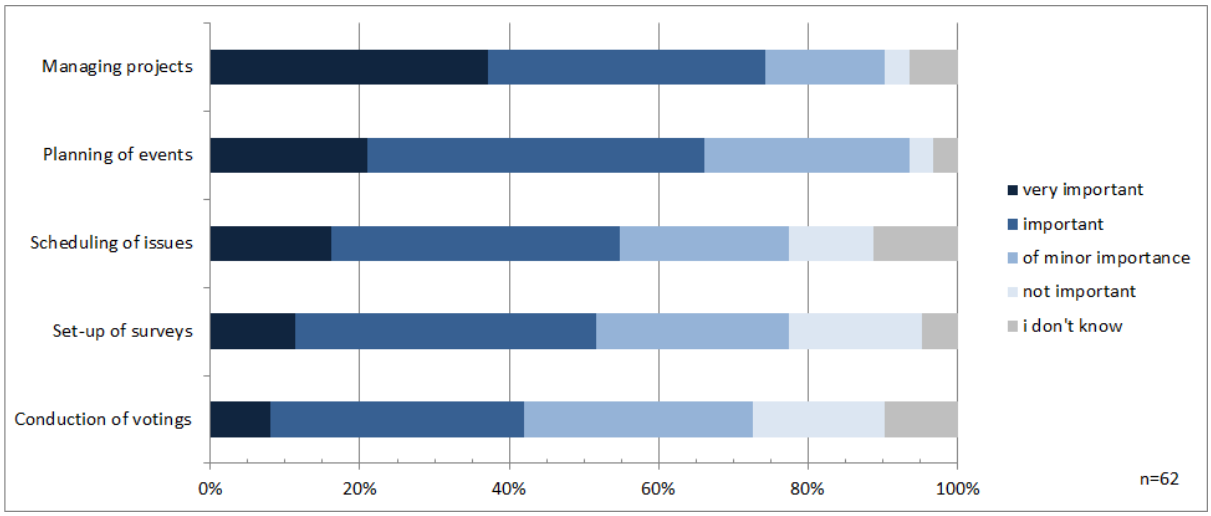


Chart 4: Distribution (in percentage) of administrative and organisational activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services and tools for the following activities?

When asked for tools that are already in use people mostly mentioned wiki-systems like Basecamp and Confluence but also combinations of tools like "Google Docs, Doodle, datumprikker [and] shared calendar on Outlook and Trello". For future use respondents mentioned e.g. project management tools which are tailored to DH projects as well as tools which support collaborative work.

"I am looking for one dedicated to Digital Humanities projects."

"I don't use any specific tools for project management. I would like to have a collaborative project management tool that I can use with my colleagues. However,

project management admits many styles, and I am not sure how this would be resolved by a tool."

It became apparent from the section where people could add additional information on tools and services which we have not listed, that a proper synchronisation between personal agendas, and calendars provided on the web is a requirement for the respondents.

Key message:
Managing projects is very important for the respondents. A suitable free and open source tool is currently not available in the *Digital Humanities*. Thus, the need for a tool that supports organisational and administrative processes during a project is obvious. Many recommendations for planning, reporting, and alignment of ideas, exist in research projects, but not all of them could be handled just by one tool or a combination of management tools.
Standard tools can be used to support particular aspects of project management. So researchers often use a box of tools and a platform that connects those could improve the situation. However, the result of the survey analysis is that a management process for projects cannot only be covered by technical solution.

2.4 Research Activities

By acknowledging all sources that have been used in the preparation of a text, respondents form part of the ongoing exchange of ideas and data that signifies the academic community. The question block about requirements in the field of research activities revealed that the respondents are most interested in that part of research activities. That includes first and foremost services to make publications and data visible, make them easy to reference and guarantee their availability by long term preservation. Depending on discipline, respondents use different kinds of primary data to support their claims. Be it research data in form of pure numbers or text, they claimed an explicit need for services and tools to make data and text citable. This very "fundamental" requirements predominate the specific research activities like *Named entity recognition* or *Optical character recognition*. (see Chart 5)

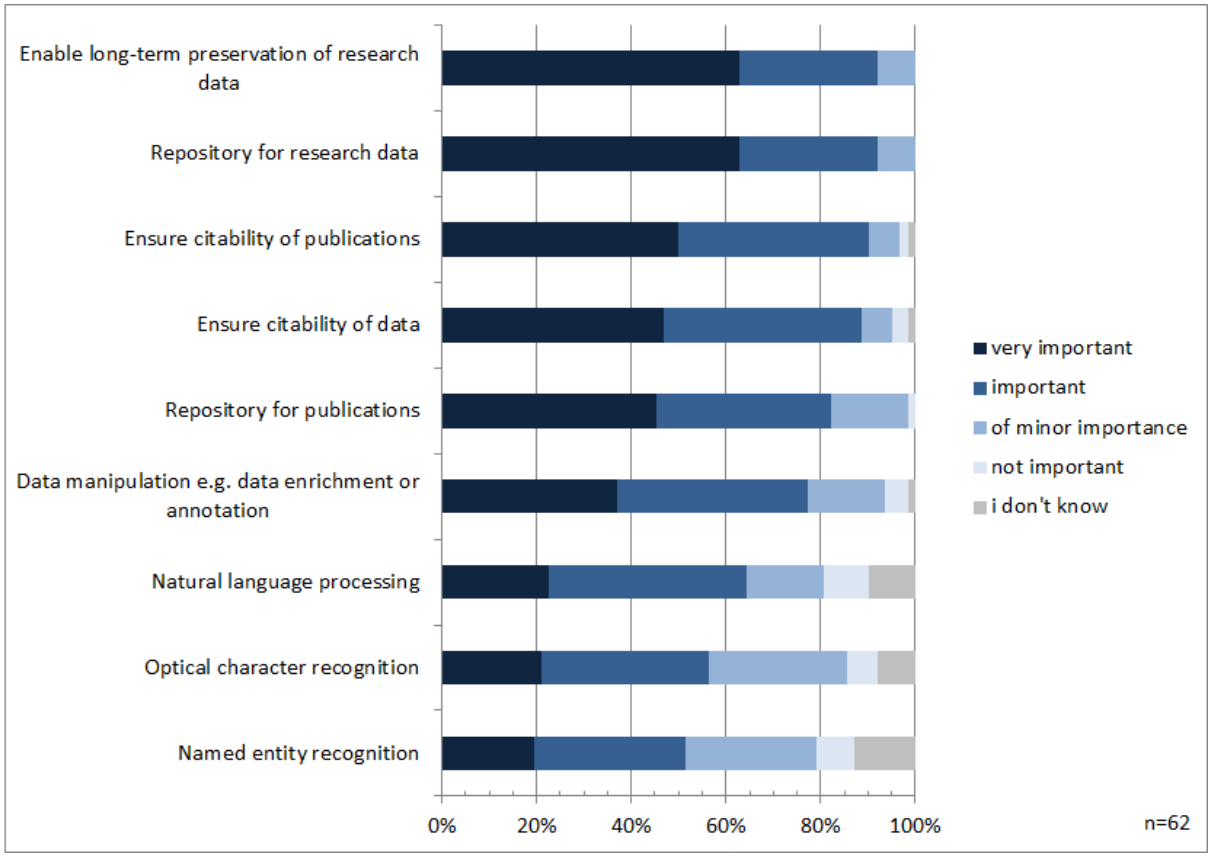


Chart 5: Distribution (in percentage) of research activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services or tools to support your research?

If one looks at the relation of the respondents' scholarly background and their need for this kind of specific tools and services, one notices that there is no relation between research areas and the need for specific tools and services. One could assume that there is a difference between the research areas and the corresponding requirements. However, the data set does not prove that.

The need for services to integrate research data into the research process is obvious from the replies. Research data are on first position still before services to deal with research publications. We assume that services to store, publish and cite articles are well known and easy to use in contrast to services that enable the publication, storage and referencing of data. The comments made by the respondents underpin that. Some people are very well aware of research data repositories gathered in e.g. re3data.org[19]. Others do not have a clue how and where to store e.g. research data. It became clear from the comments that the level of awareness as well as the capability to handle such repositories is quite heterogeneous. Trusted and easy to use repositories that are used by a broad community are what the respondents actually want for their research data.

[19] http://www.re3data.org/

"I store my research Data (texts, digit [sic] copies) in dropbox. I wish a safer alternative that would work well and would be shared by many universities and educational institutions."

"Using [currently]: TextGrid Repository. [Wishes for] future: A nice little Repository for research data that is easy to understand and easy to use. With not-so-many-metadata-input and a great usability."

"We do not have/use an overall repository for research data, but [I] would be pleased if it will be made. Currently we use websites, GoogleDocs and more but they are not general enough."

When it comes to the citability of research data one can conclude from the comments that the current situation is unsatisfying for most respondents. The user friendliness is a very important point raised by the respondents again. Current repositories do not comply with the researcher's requirements as the following statements illustrate.

"User friendly and robust citability of data is not realised yet. Several PID system do exist. The RDA [Research Data Alliance] might be an important network to establish good practices. Trusted Repositories should support the solution by ensuring the cited data is accessible / usable. I use persistent identifiers when I get them from the data archive."

"An important thing missing in the context of citability is consensus among research funders whether 1) researchers are required to always include the grant ID and the persistent IDs of data and publication and 2) in what metadata field exactly they should do this, to allow for automatic retrieval of these IDs. It would make life much easier for everyone involved. Very simple and non-technical, and the UK research councils are a good practice to follow."

Key message:
Repositories already exist and are actively used by researchers. But they need to become more user friendly and intuitive. A focus should be on services for the citability of data. The awareness of existing services needs to be extended and training programs should be offered on how to use the services. The need on research specific services like optical character recognition is relatively low. Regarding data manipulation tools we conclude that they are available in several forms of maturity, quality and sustainability and therefore no special need has been articulated.

2.5 Software Development

The last question block about software development revealed that the *hosting of source code* is by far the most needed service. It is followed by *issue tracking* and services for *authentication and authorisation.* (see Chart 6)

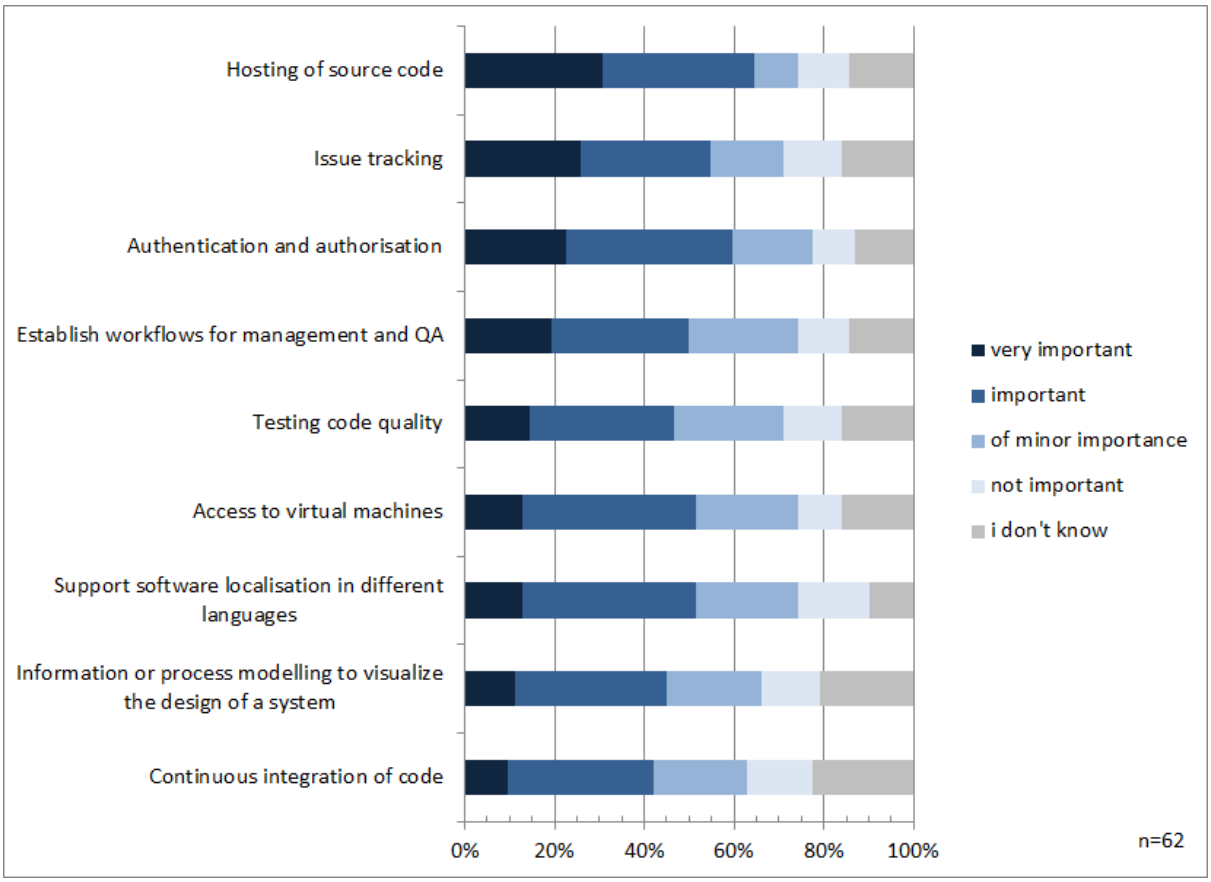


Chart 6: Distribution (in percentage) of software development referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services and tools for the following activities?

The comments in the text field for "very important" prove that "GitHub" is mostly used for public hosting of source code and that there is a clear need of code hosting services where the code is not public, such as GitLab. The following statement from the comments section illustrates that:

> "I see people of my research team using a lot of those features by using github; but github is a private service; there are no repositories for code (as far as I know), so sustainability of github is a concern."

The average rate of "I don't know" replies is much higher here than in the other question blocks.[20] However, this is not surprising since the respondents indicated mostly (68%) that they see the focus of their research activity in the area of Humanities. Thus, the field of software development seems too specific to be adequately answered by a majority of researchers with a focus in Humanities disciplines. The following reply probably illustrates the attitude of the Humanities researchers towards this kind of questions very well:

"[I] don't know (I'm not doing this, but I know it's important)."

By background and their daily research activities Humanities researchers cannot be advanced experts in the technical development of the infrastructure in DH projects. Anyhow, the data shows, that they strongly emphasise the importance of a user friendly and sustainable infrastructure and are willing to give valuable feedback from the user perspective.

Key message:
GitHub is the most frequently used tool for code hosting and issue tracking purposes. However, the two main concerns raised in relation to GitHub relate to its sustainability and to privacy concerns, as the platform is operated by a private company. Thus one important challenge is to provide a solution that offers both the usability as well as the advantages and functionality of using Git with the GitHub ecosystem one the one hand, while also guaranteeing sustainability. Institutionalised code hosting and issue tracking can provide the necessary stability and privacy also for non-public data and code as long as it has a low accessibility barrier and is accepted by the community.
Rated with slightly less importance are authentication and authorisation as well as workflows and QA. Combined with the answers of important they still reach well above 50% of importance as is the case for access to virtual machines and support for software localisation, which is not considered important at all by a large group.

2.6. Conclusions

It becomes clear from the survey that a lot of commercial tools from the private sector are frequently used in DH research projects. This situation does not only apply to the *software development* question but to all question blocks. It is irrespective of the existence of open and freely available tools and services or a real lack of them. People tend to prefer using commercial products like Google Docs, GitHub or Microsoft applications, even if tools like Etherpad exist. Due to its limited functionality in comparison with Google Docs it is not sufficient. The need for improvement and further development of Etherpad for example (with regard to formatting, versioning, general user friendliness) is one result of the survey

[20] Distribution of "I don't know" replies per question group: Software Development: 17%, Research Activities: 4%, Administrative and Organisational Activities: 7%, Cooperation Activities: 3%.

analysis. Based on the empirical survey data one conclusion is, that the preference of commercial tools derives from two reasons:

1. People are not aware of other tools even if free and publicly available alternatives exist.
2. The technical advantages (especially functionality and usability) of commercial tools like e.g. Google Docs preponderate the ones of e.g. Etherpad.

Both aspects have been mentioned several times in the comments section and distribute over all question groups. A clear need for more public relation work and dissemination activities can be derived from survey data. It goes hand in hand with the finding that very few ideas of what to use in the future exist among the respondents. People seem to be aware of the need for improvement but are not experienced enough to clearly name tools or services and/or functionalities they wish to have. This goes along with the need for more detailed and user friendly documentation of the existing tools and services and the development of training opportunities and materials in several forms.
Another important point, which derives from the survey analysis, is that the need for new developments is rated rather insignificant, whereas the proper linkage of existing tools and services should be strengthened. WP6 will consider this when it comes to the next steps.
The survey results provide not only input for the HaS WP 6 but also for WPs 3 and 4. The need for a dissemination and communication strategy regarding existing tools and services as well as concrete training and education services to enable low-threshold access are open issues to deal with in the HaS project.
Regarding the concrete work in WP6, the technical implementation as well as the adjustment and linkage of existing tools and services is in the focus of the next steps. WP6 team will discuss and choose 2 demonstrators to be implemented into the DARIAH infrastructure and discuss how to best link existing tools. The following table lists the most important tools for each question block and will be the basis for that decision.

Question block	Most important tool/service
Collaboration activities	Management of digital content in a content management system
Administrative and organisational activities	Managing of projects
Research activities	Enable long-term preservation of research data / repository for research data
Software development	Hosting of source code

Table 2: Most important tool or service for each question block

3. Next Steps

The next step in T6.1 will be the selection of demonstrators that will be implemented into the DARIAH infrastructure to test the use of the proposed tools.
This process needs to be accompanied by improving upon the existing communication and dialogue between researchers and users on the one hand and the service providers on the other hand. Several of the DARIAH partners already offer some tools and services that can be used by the community, such as the DARIAH-DE developer portal[21] or the DANS data repository EASY[22]. However, the information provided in the comments section of the survey disclose that usability and accessibility as well discoverability in general need to be addressed and improved on all ends. At the same time, it is clear that a common understanding of the importance of providing basic services and tools to the wider community is indeed prevalent.
Additionally Service Level Agreements (SLA) which might be required in different national interdisciplinary contexts and especially in the European Research Area as a whole in order to provide a sustainable environment will be gathered and evaluated.

The survey does not cover the evaluation of SLAs for several reasons:

1. The identification of needed tools and services is strategically more important for the work in T6.2 than the evaluation of SLAs. There is a strong interdependency among T6.1 and T6.2, which depends heavily on the selected tools in task 6.1
2. For internal reasons, work in T6.3 (Develop Service-Level-Agreements and a network of reliable service providers) could not start yet. Without a vision and respective input from T6.3 the incorporation of questions about SLAs was not reasonable.
3. The covered areas for tools and services in the survey are already broad. Since we were totally aware of the chance of a very low respondents rate we tried to avoid that risk be keeping the survey as concise as possible without losing information.

First results from T6.3 will form the basis for further work on SLAs in T6.1.

[21] https://de.dariah.eu/developer-portal

[22] https://easy.dans.knaw.nl/ui/home

Part 2

4. Survey data

This Section contains the full number of responses to all questions of the survey. You find the absolute number of replies in the tables.

In which area do you see the focus of your professional activities?

Humanities research	42
Technical development	11
Other	9

What is your primary discipline of research? Please select "Other" and add your discipline in the box on the right side if none applies to you.

Ancient Cultures	9
Computer Science	8
Economics	0
Education Sciences	0
Fine Arts, Music, Theatre and Media Studies	3
History	8
Jurisprudence	0
Linguistics	7
Literary Studies	9
Non-European Languages and Cultures, Social and Cultural Anthropology, Jewish Studies	2
Philosophy	0
Psychology	0
Religious Studies	0
Social Sciences	4
Theology	1
Other	11

How important do you consider infrastructure services and tools for the following collaboration activities?

	(1)	(2)	(3)	(4)	(X)
Collaborative and simultaneous writing and editing of texts	26	25	8	1	2
Management (archiving, structuring etc.) of digital content in a content management system	32	24	4	2	0
Communication in real time via instant messengers	11	16	22	11	2
Communication via telephone or video conferences	20	25	11	4	2
Synchronisation of files	27	28	3	0	4

(1) Very important, (2) Important, (3) Of minor importance, (4) Not important, (X) I Don't Know

How important do you consider infrastructure services and tools for the following activities.

	(1)	(2)	(3)	(4)	(X)
Set-up of surveys	7	25	16	11	3
Conduction of votings	5	21	19	11	6
Scheduling of issues	10	24	14	7	7
Managing projects	23	23	10	2	4
Planning of events	13	28	17	2	2

(1) Very important, (2) Important, (3) Of minor importance, (4) Not important, (X) I Don't Know

How important do you consider infrastructure services or tools to support your research?

	(1)	(2)	(3)	(4)	(X)
Repository for research data	39	18	5	0	0
Repository for publications	28	23	10	1	0
Ensure citability of data	29	26	4	2	1
Ensure citability of publications	31	25	4	1	1
Enable long-term preservation of research data	39	18	5	0	0
Data manipulation e.g. data enrichment or annotation	23	25	10	3	1
Optical character recognition	13	22	18	4	5
Named entity recognition	12	20	17	5	8
Natural language processing	14	26	10	6	6

(1) Very important, (2) Important, (3) Of minor importance, (4) Not important, (X) I Don't Know

How important do you consider infrastructure services and tools for the following activities?

	(1)	(2)	(3)	(4)	(X)
Hosting of source code	19	21	6	7	9
Issue tracking	16	18	10	8	10
Establish workflows for management and QA	12	19	15	7	9
Testing code quality	9	20	15	8	10
Continuous integration of code	6	20	13	9	14
Information or process modelling to visualize the design of a system	7	21	13	8	13
Support software localisation in different languages	8	24	14	10	6
Authentication and authorisation	14	23	11	6	8
Access to virtual machines	8	24	14	6	10

(1) Very important, (2) Important, (3) Of minor importance, (4) Not important, (X) I Don't Know

5. Appendix

5.1 List of abbreviations

DARIAH	Digital Research Infrastructure for the Arts and Humanities
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
DH	Digital Humanities
HaS	Humanities at Scale
ID	Identifier
PID	Persistent Identifier
RDA	Research Data Alliance
SLA	Service Level Agreements
UK	United Kingdom
WP	Work Package

5.2 List of charts and tables

Chart 1: Distribution (in percentage) of professional activities referring to all complete replies (n=62). Question: In which area do you see the focus of your professional activities? 10

Chart 2: Distribution (in percentage) of primary research areas referring to all complete replies (n=62). Question: What is your primary discipline of research? Please select "Other" and add your discipline in the box on the right side if none applies to you. Some possible options (including "Other") are not displayed in this chart, because no participant chose one of them. 11

Chart 3: Distribution (in percentage) of collaboration activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services and tools for the following collaboration activities? 12

Chart 4: Distribution (in percentage) of administrative and organisational activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services and tools for the following activities? 14

Chart 5: Distribution (in percentage) of research activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services or tools to support your research? 16

Chart 6: Distribution (in percentage) of software development referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services and tools for the following activities? 18

Table 1: You indicated that you think it is very important that collaborative and simultaneous writing and editing of texts is supported by basic infrastructure services and tools. Please let us know which services or tools you already use and / or which ones you would like to use in the future. – limited to the number of incomplete responses 9

Table 2: Most important tool or service for each question block 20