

Community Involvement in Research Infrastructures: The User Story Call for Text+

Nanette Reißler-Pipka
Raisa Barthauer
Stefan Buddenbohm
José Calvo Tello
Sonja Friedrichs
Lukas Weimer

Göttingen State and University Library

Measures Leading to Community Involvement

Involving the community in the network of a research infrastructure is a challenging task for all disciplines both on the national and on the European or international level. A successful community involvement guarantees a match of services, tools and user requirements as well as increased acceptance and motivation regarding the cultural change towards open access and sharing of research data. In the context of the NFDI (German National Research Data Infrastructure), community involvement was identified as one of the cross-cutting topics.¹

On the European and international level, examples of community involvement consist predominantly of working groups and strategies aimed at building interest in the respective research infrastructures. These are effective measures to involve the network, but their success highly depends on the continuous engagement of the participants.² Many research infrastructure consortia further approach this problem by integrating individual researchers who are able to represent the interests of parts of the community. A similar measure is the establishment of an ongoing liaison to the relevant professional associations for each

¹ Bierwirth et al. (2020) and Glöckner et al. (2019).

² Examples may be found in the [Research Data Alliance \(RDA\)](#), the [European Open Science Cloud \(EOSC\)](#) or in the field of the humanities in [DARIAH-ERIC](#) or on the national level in [CLARIN-D](#).

research area.³ Comparatively less effort is required to provide a scientific advisory board, an approach that is quite common and mostly undertaken in addition to other measures of community involvement.

In the extensive field of community involvement, user stories are one of many methods to identify users' needs. Other components are, for example, interviews, surveys in the context of community offerings, and the evaluation of usage statistics of websites offering resources and information. These methods involve varying degrees of interaction with the community: from gathering data via anonymously tracking behaviour using an existing service, to collecting numbers according to key performance indicators, to actively participating in community workshops, conducting interviews, or writing user stories. Depending on their design and execution, user stories have the potential not only to capture the pure user needs, but also the context in which the latter arise, thus allowing for the creation of a much more complete picture of research activities.⁴

Context: NFDI (German National Research Data Infrastructure)

In 2013, the [Joint Science Conference](#)⁵ (Gemeinsame Wissenschaftskonferenz, GWK) decided to establish the [German Council for Scientific Information Infrastructures](#) (Rat für Informationsinfrastrukturen, RfII), which has been preparing and supporting the process of forming a National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI) since its constitution.⁶ The German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) coordinates the public discussion and the funding process. The NFDI aims “to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally.”⁷ The formation of the NFDI is a science-led process with the aim of developing and securing data resources in the long term along the FAIR principles (*Findable, Accessible, and Interoperable/Re-usable*)⁸ and across disciplinary and political borders.⁹ It should represent all scientific disciplines in its overall structure, which in turn is to be achieved through the composition and goals of the funded consortia.¹⁰ Organised as an

³ Text+ follows this strategy by creating [Scientific Coordination Committees](#) which consist of elected representatives of the professional associations that cooperate with Text+.

⁴ Warwick (2012), p. 3-6.

⁵ This federal and state organisation decides on joint science funding programmes, <https://www.gwk-bonn.de/>.

⁶ German Council for Scientific Information Infrastructures (2016).

⁷ https://www.dfg.de/en/research_funding/programmes/nfdi/index.html

⁸ Wilkinson et al. (2016).

⁹ NFDI Expert Committee (2020), p.2.

¹⁰ NFDI Expert Committee (2020).

[association](#), the NFDI started officially in October 2020.¹¹ The nodes of the developing national infrastructure network are the – usually discipline-based – autonomously acting consortia. Text+ is one of ten consortia the Joint Science Conference (GWK) decided to fund in the second round of the NFDI call in 2021. Text+ intends to build a research data infrastructure focused on language and textual data, with a special interest in the data domains Collections, Lexical Resources, and Editions.

At the NFDI Conference 2020, the “science-led process with community-driven processes”¹² was named as a framework condition for the development of the NFDI. The RfII has therefore called for an active exchange between infrastructure and community, i.e. providers and users, in order to overcome the given obstacles.¹³ In the field of the humanities, several workshops were organised to sort out common needs and interest in the community at a very early stage of the NFDI building process in 2018.¹⁴ In other disciplines, the interaction with the community is based on workshops and the definition of user types. As such, [NFDI4Ing](#) developed archetypes for typical fields of work, careers, and challenges of its community.¹⁵ Other NFDI initiatives in the humanities, such as [NFDI4Memory](#) and [NFDI4Culture](#), took slightly different measures than Text+ to describe user needs by gathering short problem stories or user stories according to the data life cycle.¹⁶

Text+ decided to launch its call for user stories in spring 2020 to integrate the requirements of its users and to contribute to the further development of the community. The exchange was established by inviting researchers to participate with personal user stories from their daily research experiences. They do not only express a need, a requirement or a wish, but tell a story, give an example, a reason and express valuable ideas for possible solutions and further involvement.

Methodology

Based on recommendations by the RfII regarding previous work, community engagement and connection, Text+ took at least three assumptions into consideration during the application process:

1. **Existing services:** Text+ does not start from scratch but builds upon already established research infrastructure components, which have gained trust and

¹¹ Founded on 12.10.2020 and entered in the register of associations, <https://www.nfdi.de/verein>.

¹² “Aufbau in einem wissenschaftsgeleiteten Verfahren”, “Konsortien entstehen in Community-getriebenen Prozessen”: Schill (2020), p. 3.

¹³ German Council for Scientific Information Infrastructures (2018) and Gehring (2020).

¹⁴ <http://forschungsinfrastrukturen.de/doku.php>.

¹⁵ <https://nfdi4ing.de/>.

¹⁶ <https://4memory.de/problem-stories-overview/> and <https://nfdi4culture.de/downloads.html>.

sustainability due to long-term projects and a wide partner network in CLARIN-D or DARIAH-DE. These existing research infrastructures and research services have to be considered as building blocks.

2. **Community engagement:** Community involvement is essential to provide services that meet researchers' needs. Researchers must therefore be seen as partners and conversations have to be held with the whole community to understand their needs. One way to do this is to involve the community within the research data infrastructure. But their integration can also happen through their involvement in, for example, specialist groups or committees in the infrastructure – or through user stories.
3. **Research proximity:** For research data infrastructures, proximity to researchers is crucial. Researchers have their own individual user perspectives that must be taken into account. Especially in the humanities, research is divided into many different disciplines. For this reason, the research questions that Text+ has to address are diverse and individual. It is up to the development team to abstract from these individual experiences to a concept for a specific research infrastructure.

Taking these three assumptions into account, the Text+ team conceptualised a user stories approach for the collection of individual researchers' requirements and their obstacles in finding, accessing, analysing, and publishing data and also in the context of teaching. Moreover, research projects or research groups were invited to share their experiences and ideas. The collection of user stories was intended to serve as a data set for the analysis, evaluation, and categorisation of user requirements.

Another inspiration for the Text+ user story call came from the Social Sciences and Humanities Open Cloud ([SSHOC](#)) project,¹⁷ which contributes to the SSH domain of the European Open Science Cloud ([EOSC](#)). User stories, use cases, and public consultations,¹⁸ among other means,¹⁹ have been used in this project to engage with the SSH community, to discuss with them, in a joint process, the conceptualisation of services – for example the SSH Open Marketplace – and to achieve a sense of ownership within the community for the resources provided by SSHOC.

The User Story Template

To achieve a consistent structure of the user stories, we provided a template designed to reduce the effort on both sides – the users could use a given structure and we could publish

¹⁷ <https://www.sshopencloud.eu/>

¹⁸ Barbot et al. (2020) and Buddenbohm et al. (2020).

¹⁹ Barbot et al. (2019).

and analyse the results more easily. The template was offered both in [German](#)²⁰ and [English](#)²¹ and consisted of five sections, of which only the first two were mandatory:

- **Motivation:** Each user story had to begin with a paragraph on the motivation of the writer, explaining how the research activity described came about. In addition, to categorise the community which Text+ addresses, the subject area needed to be classified according to the DFG subject classification system ([DFG-Fachsystematik](#)).²² The category “Humanities and Social Sciences” splits up into 13 sub areas and each of these is broken down into two to five more sub-areas. Optionally, specific requirements from the community's point of view could be described in this first section, and possible requirements from the funding organisations' point of view could be considered.
- **Objectives:** In the second section, we asked to describe the aims of the user story, meaning which specific problem or demand should be addressed by Text+. If possible, a reference to a larger Text+ or NFDI strategy could be made here, by ideally specifying the data domain according to the published “[Binding letter of intent](#)” by Text+.²³
- **Solution:** In the third section, suggested solutions to the problems mentioned could be described by the author. These pertained to which resources might be required and what kind of innovations would have to be implemented and how Text+ could contribute to solving the problem, the research question, or reducing the barriers to managing research data. In addition, it was possible to describe which data sources and which services, tools, and components would be needed to achieve this goal. Finally, standards and methods could be described.
- **Challenges:** In a fourth section, possible challenges and risks could be identified which lie outside the control of Text+ or might risk the suggested solution. Alternative solutions could be named here.
- **Review by the community:** The fifth and last section was about a possible evaluation by the community. Here, the authors could indicate their willingness to review the services offered by Text+ during the possible funding period. This feedback loop is intended to determine the success of the proposed solutions.

We asked the authors to stay within 800 words and to deliver the text preferably in English (whereas we also accepted user stories written in German to be translated afterwards).

²⁰ https://www.text-plus.org/wp-content/uploads/2020/09/Text_User-Stories_Template_dt.pdf

²¹ https://www.text-plus.org/wp-content/uploads/2020/09/Text_User-Stories_Template_en.pdf

²² https://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp

²³

https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen_2020/2020_nfdi_Text.pdf

Typically, user stories tend to highlight an individual researcher's own experiences. Therefore, we suggested the first-person perspective. The stories could describe a particular research activity or an entire research process and we also invited researchers to mention any setbacks, errors, or lack of help from research data infrastructures. It was important for further analysis to identify the types of research data and services involved and required, as well as to present the expectable impact on language- and text-based research in the humanities. We also created the template in a way that it could measure the reasons and motivation to use the Text+ services. Its standardised structure facilitated the evaluation of existing needs and desired improvements. With the help and support of the community, Text+ can derive measures for its future development in a user-centered way and constantly check and assure the quality of existing services.

The Call for User Stories

The call for user stories was launched to reach out to the text- and language-based research disciplines relevant to the Text+ initiative. The call was posted on the Text+ website and the [DHd blog](#),²⁴ but it was also disseminated via mailing lists and personal addresses.

User stories serve the purpose of providing an overview of how different target groups and disciplines have used, are currently using, and will probably use the services and tools offered by digital research data infrastructures. This aim was supported by the template structure described above. Ideally, the call should cover all stakeholders, that is individual researchers at various levels of their career, research groups, as well as providers of the infrastructure, who interact with the communities on different levels. In addition, the call aimed to address as many disciplines as possible in order to provide a complete overview. Finally, all kinds of infrastructural services should be covered, with a leaning towards the three [Text+ data domains](#) of Collections, Lexical Resources, and Editions.²⁵ The user stories should refer to existing services and offers, if possible, but should also suggest modifications or extensions that might be of interest to the community or to the individual, and even propose new offers. Such suggestions provide Text+ and the NFDI as a whole with pointers for designing the range of services to meet user needs. Therefore, the call aimed at an integration of user needs with the widest possible range, so that Text+ services could be conceptualised in cooperation with the community.

Adopting the established DFG subject classification system (see also chapter *The User Stories Template*) offers several advantages, but reflects individual researchers only to a limited extent, as demonstrated by experiences with user stories submitted to the Text+ call.

²⁴ Bertino (2020).

²⁵ <https://www.text-plus.org/en/research-data/data-domains/>

The user story [Building Corpora for Comparative Analysis to Identify Declarations of Love in Letters](#)²⁶ is an example of how different subjects can be connected within projects. Other user stories highlight interdisciplinary needs, such as the story [Provision of processable textual data in libraries](#).²⁷ Accordingly, categorising the user stories according to the DFG subject classification comes with certain drawbacks (e.g. a lack of an appropriate consideration of interdisciplinarity). Apart from subject areas or sub-areas, the user stories call was aligned with the three data domains of Text+, respectively:

- Collections
- Lexical resources
- Editions

In addition to the three data domains, we introduced the category Comprehensive, which covers infrastructure-related requirements that do not fit only one of the data domains: these are, for example, user stories on legal aspects or general data management questions ([Textual features derived from copyrighted texts](#)).²⁸

Alongside this call for user stories, Text+ has launched a supplementary call. The [Call for Data](#)²⁹ invites the community to share their research data via Text+ in order to expand the data offering of Text+ and to represent the entire user landscape.

Submissions

The community's general reaction and feedback to the call was overwhelmingly positive. In response to the call, a total of 120 user stories were submitted, both by single and by multiple authors. Most authors closely followed the structure of the template for their stories. However, due to the large number of submissions and the wide range of subjects involved (see chapter *Analysis*), the individual user stories did not only vary in content, but also in their execution. There were major differences, especially regarding the following points:

- **Language:** As mentioned before, we encouraged the use of the English template, but also provided a German one. User stories submitted in the latter language had to be translated by the team for further use, i.e. publication and evaluation. The emphasis was on a translation that reflected the original wording as accurately as possible.
- **Template structure:** Some of the authors had changed the template to adjust the structure to the specific needs of their user stories. In many cases, the order of the sections was changed, several sections were merged, or the given sections were

²⁶ Hastik (2020).

²⁷ Wübbena (2020).

²⁸ Schöch (2020).

²⁹ <https://www.text-plus.org/en/research-data/data-from-the-community/>

ignored at all. During the editorial process and in consultation with the authors, these user stories were adapted after the submission to conform to the template's structure as much as possible.

A further frequent deviation from the template was that not all sections were answered – possibly because some questions could not be answered (see also chapter *Analysis*). However, as described above, not all sections were mandatory and if no essential section was left out, these stories were included in the analysis without any further changes to this point.

- **Other formats:** A very small number of authors submitted lists, collections of ideas or spreadsheets, etc. To make these stories comparable and easier to evaluate, the writers were asked to make adaptations or we had to reject these submissions.
- **Missing information:** Apart from individually modified structures in the user stories, minor changes were sometimes made to the template, such as missing titles, no information on authors or the discipline, or varying formatting. While the formatting could mostly be directly changed, titles for the user stories were created and inserted in agreement with the authors.

For reasons of comparability and clarity, the user stories were edited in a minimal-invasive way. All editorial changes were made in close consultation with the authors and all user stories were published solely with their explicit permission. The authors of four stories did not consent to their publication, in some cases due to ongoing grant applications. All other user stories were published on the Text+ website.

Publication

For the sake of maximum transparency and openness towards the community and the wider public, we decided to publish the user stories as soon as possible.

If not already done by the authors, the user stories were assigned to one of the three Text+ data domains and according to the DFG subject classification system. In many cases, more than one subject area was assigned to the respective user story. When requesting the author's permission to publish their user story on the Text+ website, we also asked for consent to add the name and institution of the author(s). Some user stories were written by project teams, research groups (e.g. [Inscriptions in Germany from the Middle-Ages to Early Modern Times \(Die Deutschen Inschriften\)](#))³⁰ or by institutional networks like Specialised Information Services (Fachinformationsdienst, FID, e.g. [Infrastructural Needs for Romance](#)

³⁰ Arbeitsstelle Inschriften (2020).

[Research Data](#)).³¹ Each user story received an individual URL via permalink. This is an important criterion to ensure that each submission can be referenced.

The user stories are available on the corresponding [subpage](#),³² with the possibility of filtering the submissions according to both Text+ data domains and DFG subject areas to facilitate searching for user stories. If more than one subject area was assigned, the user stories were marked with asterisks.

Analysis

As mentioned before, we received a total of 120 user stories and the authors of 116 user stories gave the permission to publish them. These stories were published on the website of Text+ and their metadata and annotation used for this analysis are also available.³³ The stories were written by more than 163 different authors.

Figure 1 visualizes the percentage distribution of user stories along the Text+ data domains.

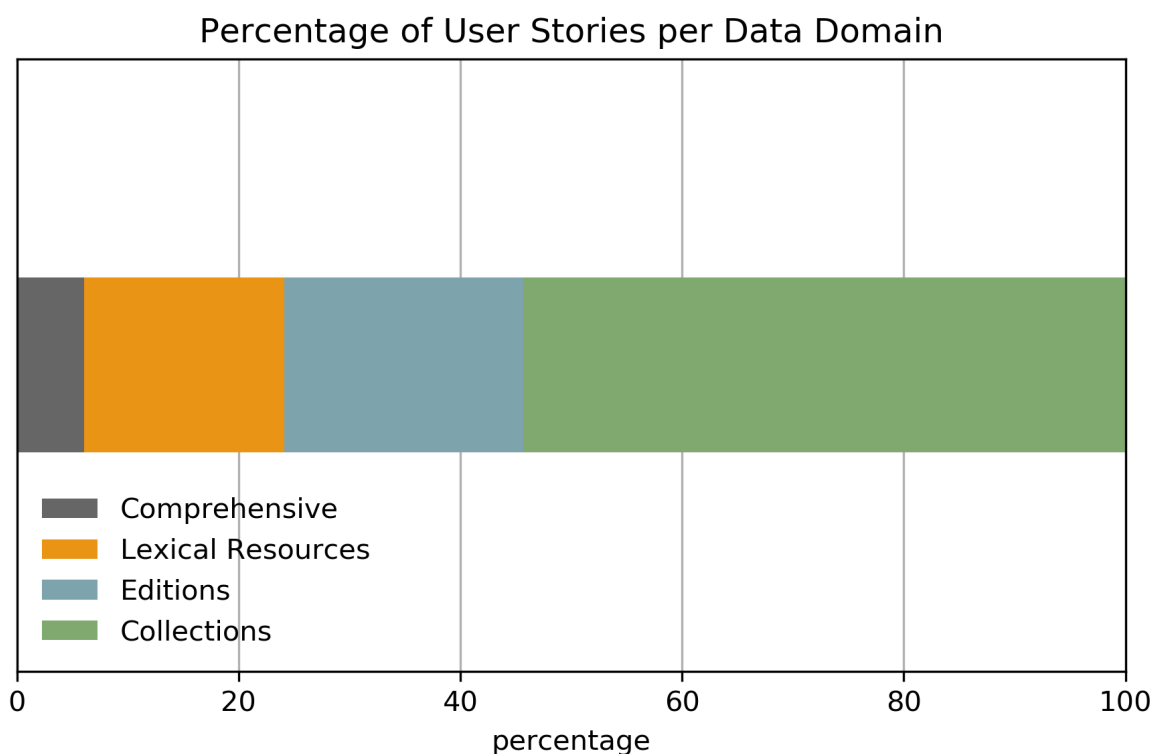


Figure 1. Percentage of Text+ user stories related to data domain

³¹ Vacano/Grüter FID Romanistik (2020).

³² <https://www.text-plus.org/en/research-data/user-stories-en/>

³³ <http://dx.doi.org/10.20375/0000-000E-67ED-4>

Figure 2 visualizes the number of the user stories assigned per subject area, taking into account multiple assignments.

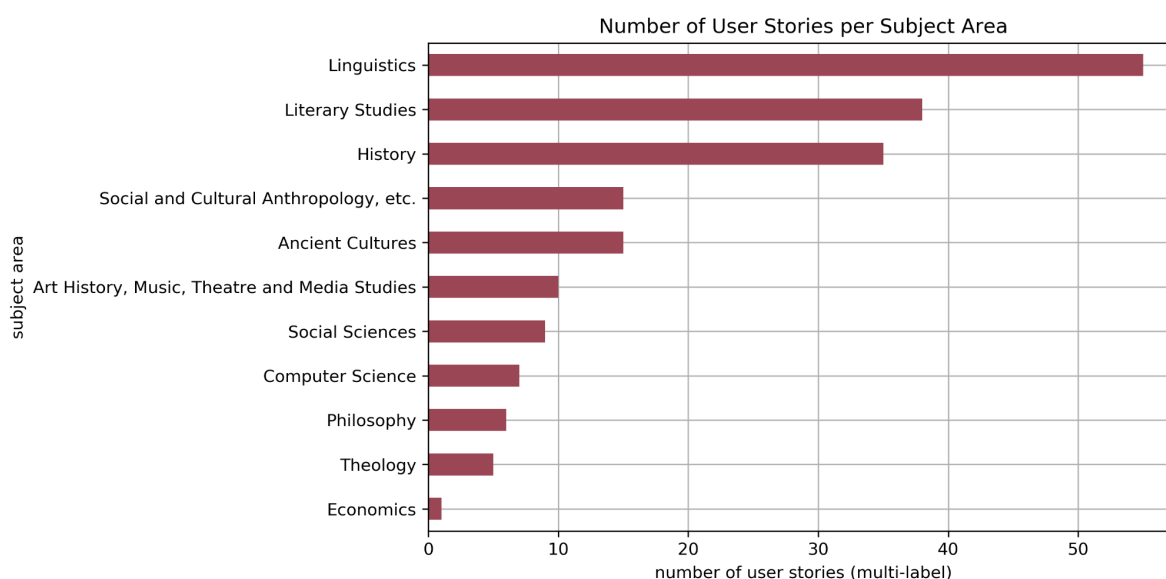


Figure 2. Number of Text+ user stories per subject area

The numbers clearly show that three larger groups of stories are related to Linguistics, Literary Studies, and History. These are followed by Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies; Ancient Cultures and Art History, Music, Theatre and Media Studies. Fewer user stories relate to further areas from the humanities, natural sciences or economics (e.g. [Semantic Uncertainty in DIN norms](#)³⁴ or [Data security, system openness, networking – wishes for the NFDI from view of the MEGA](#)³⁵).

As explained before, the authors could also assign their story to more than one DFG subject area. This is why the sum of the bars' values in Figure 2 is higher than the total number of user stories. For example, 24% of the user stories assigned to Linguistics are exclusively assigned to this subject area. In the case of Literary Studies and History this is the case for 14% and 8%, respectively. This needs to be considered when drawing conclusions about individual subject areas in Text+.

Several members of our team read and edited the 120 user stories and communicated with the authors. Each editor had an in-depth overview of the user stories and their respective context. We decided to annotate each user story with keywords. The keywords were not

³⁴ Stegmeier et al. (2020).

³⁵ Hubmann/Roth (2020).

assigned automatically, but through close reading by the team. For this purpose, we designed a two-layered keywording system.

In the first step, we defined a small set of keywords to be assigned to each user story in order to be able to obtain a general overview about the requirements of each user story from a bird's-eye perspective. This narrow set of keywords contained only five keywords, which refer to the [FAIR principles](#), combining “Interoperable and Reusable” into a single term due to their interdependencies. These were supplemented by two keywords which refer to “Community activities” and to “Software services”.

Each user story could be labelled with up to three of these keywords. This narrow set of keywords was assigned a total of 259 times, with a mean of 2.32 keywords per story and a standard deviation of 0.64.

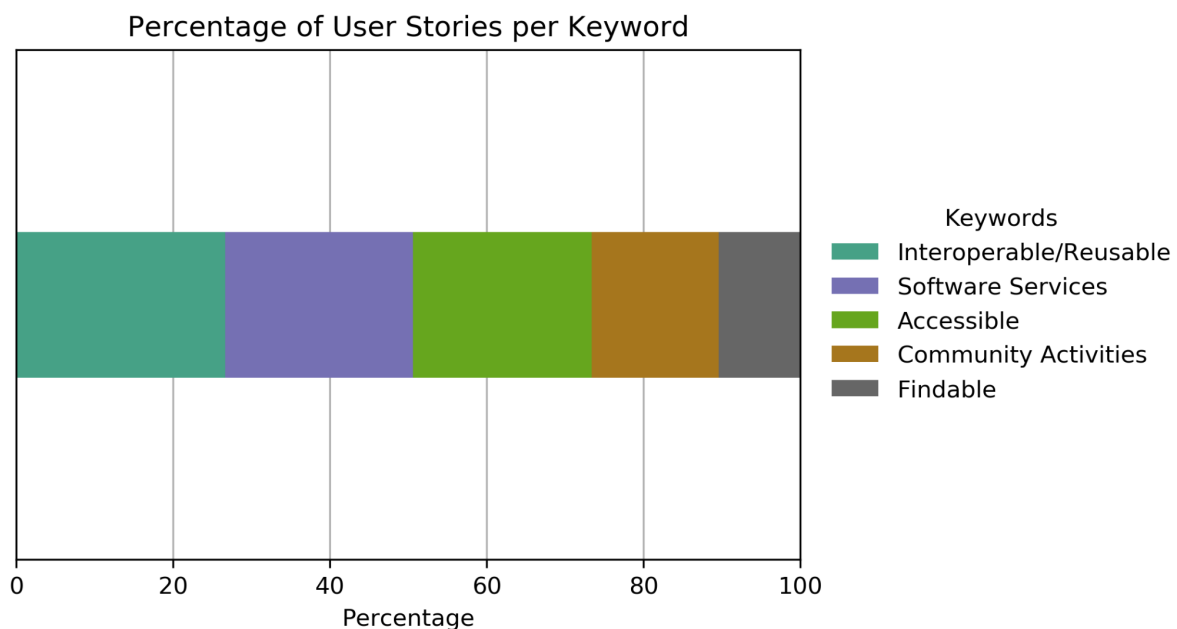


Figure 3. Percentage of Text+ user stories related to five keywords

The percent stacked bar plot in Figure 3 shows the distribution across the different categories, with three keywords having been assigned to a comparably high number of stories: “Interoperable/Reusable” with 69 stories, “Software Services” with 62, and “Accessible” with 59. The category “Community Activities” is close to this group (42 stories). “Findable” is the keyword with the lowest number of stories associated with it (27).

In order to provide a more granular overview, these keywords are put in relation to the Text+ data domains in Figure 4.

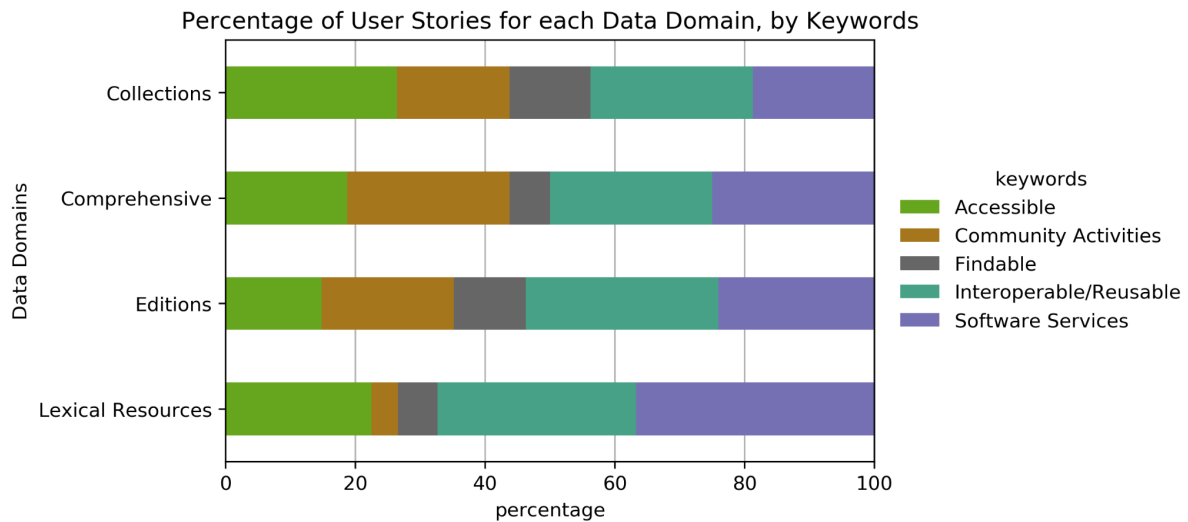


Figure 4. Percentage of Text+ user stories related to five keywords and data domain

This narrow set of five keywords offers a general overview of all user stories and was used as a reference point in the writing process of the Text+ application. As can be seen here, the keywords in the data domains Collections, Comprehensive and Editions show a fairly similar percentage distribution, but the latter differs noticeably in Lexical Resources: here the keyword “Community Activities” is used significantly less and the keyword “software services” more frequently than in the other three data domains. It is also surprising that the keyword “Findable” was not used as often as one would have expected: this aspect of the FAIR principles was addressed by the user stories across the data domains significantly less often than the other aspects (“Accessible” and “Interoperable/Reusable”). Furthermore, it can be seen that “Findable” was addressed almost twice as often in the user stories of the Editions and Collections data domains than in Lexical Resources and Comprehensive.

However, as one can also see from this evaluation, this rather rough set of keywords was not yet informative enough. For this reason, we decided to define and assign a second, wider set of keywords. These emerged from an iterative process in which we re-read and analysed the user stories, considered frequent collocations in the text of the stories, and tried to generalise them. By doing so, we aimed to associate similar, but differently formulated needs and requirements to each other. We ended up with 60 different keywords organised in five groups:

- **Type of situation:** Keywords specifying what the user needs or is interested in (“tool production”, “data production”, “interest in learning”, “tool maintenance”, “interest in data linking”, etc.).
- **Type of tool and services:** Keywords specifying what tool family is required or offered in particular (“OCR”, “lemmatiser”, “sentiment analysis”, etc.).
- **Type of data:** Keywords specifying the type of data being analysed or produced beyond the classification along the Text+ data domains (with values such as “annotation/metadata”, “audio/video”, etc.). In contrast to, e.g., “interest in annotation”, this refers to the produced or analysed type of data.
- **Type of link:** Keywords specifying the interest in linking either resources from two distinct Text+ data domains with one another (e.g. editions with lexical resources) or in linking resources from a single data domain with external resources (e.g. editions with maps). It was specified which types of resources should be linked. The type of link should be understood as a set of keywords that specify the “interest in data linking” observed in the type of situation. We are aware that this higher degree of specificity could cause conflicts when these keywords are compared to the rest of the categories.

As in the previous case, the user story could also be labelled with more than one of these keywords. In total, we assigned 651 keywords, with a mean of 5.66 keywords per user story and a standard deviation of 2.24.

The following percent stacked bar plot (Figure 5) shows the ten most frequent keywords assigned to the user stories in total.

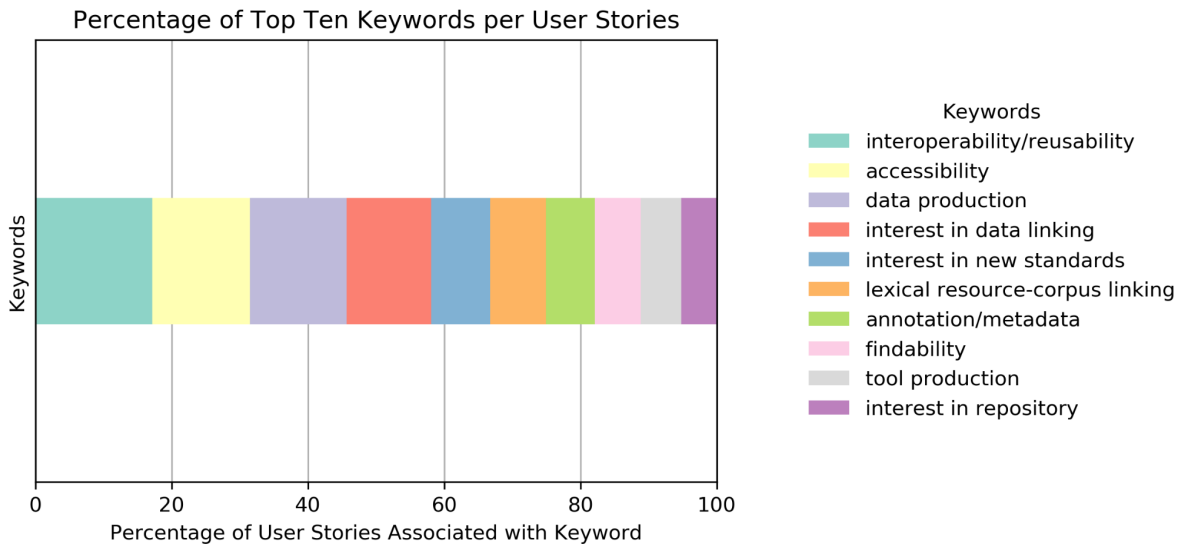


Figure 5. Percentage of Text+ user stories related to keywords

Interestingly, the authors of the stories addressed the role of producing data (57 user stories) more frequently than the role of producing tools (24 stories) or of requiring more data (20 stories). A frequent requirement – expressed in 25 stories – is about linking different resources, in particular with regard to “lexical resource-corpus linking”. Two other frequent keywords express the interest in new standards and in repositories.

Even more granular insights can be gained by analysing the individual data domains, as can be seen in the following figures.

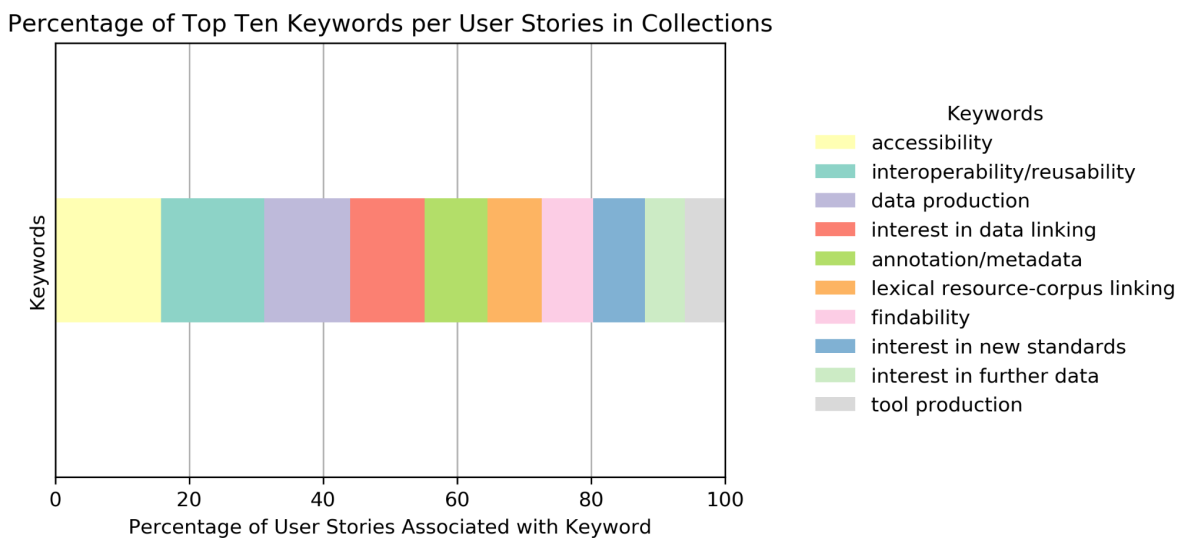


Figure 6. Percentage of Text+ user stories related to keywords in the data domain Collections

Since Collections compose the largest data domain, Figures 5 and 6 are rather similar, with one prominent difference: “interest in repository” was among the top 10 keywords for the

entirety of the user stories (cf. fig. 5), but this is not the case for Collections. Vice versa, “interest in further data” ranks among the top 10 keywords for this data domain, but not for the entirety of the user stories (cf. fig. 6).

Percentage of Top Ten Keywords per User Stories in Editions

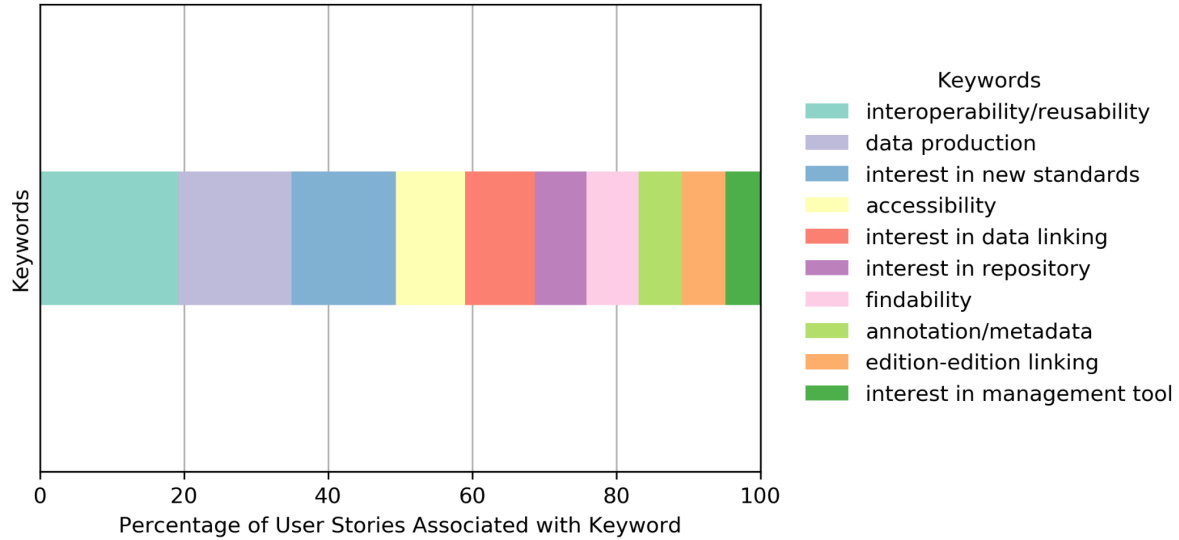


Figure 7. Percentage of Text+ user stories related to keywords in the data domain Editions

Figure 7 shows the top ten keywords applied to the user stories in the data domain Editions. In comparison with the other data domains, “interest in new standards” has a higher frequency. While the “interest in data linking” is similar to other data domains, creating links between several editions (“edition-edition linking”) ranks among the top ten of the keywords. Furthermore, the keyword “researchers network” is on the top-ten keywords for Editions, but not for any other data domain.

Percentage of Top Ten Keywords per User Stories in Lexical Resources

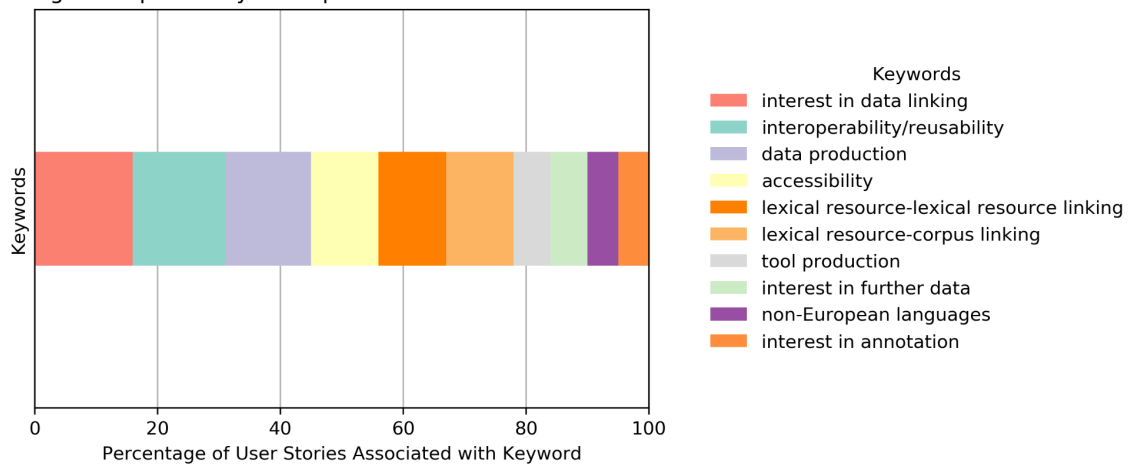


Figure 8. Percentage of Text+ user stories related to keywords in the data domain Lexical Resources

For the data domain Lexical Resources (Figure 8), the most frequent keyword is related to the interest in linking resources (“interest in data linking”). We decided to add more specific keywords describing the resources that should be linked (e.g. “lexical resource-corpus linking”, etc.). These more specific keywords allow for a finer description of the requirements. In the case of Lexical Resources, the authors of the user stories show specific interest in the connection between several lexical resources, but also between these and corpora. Similar to the data domain Collections, an important number of user stories are related to “tool production” and “interest in further data”. As could be expected, the data domain Lexical Resources shows a stronger tendency to consider “multilingual data”, also in “non-European languages”.

Percentage of Top Ten Keywords per User Stories in Comprehensive

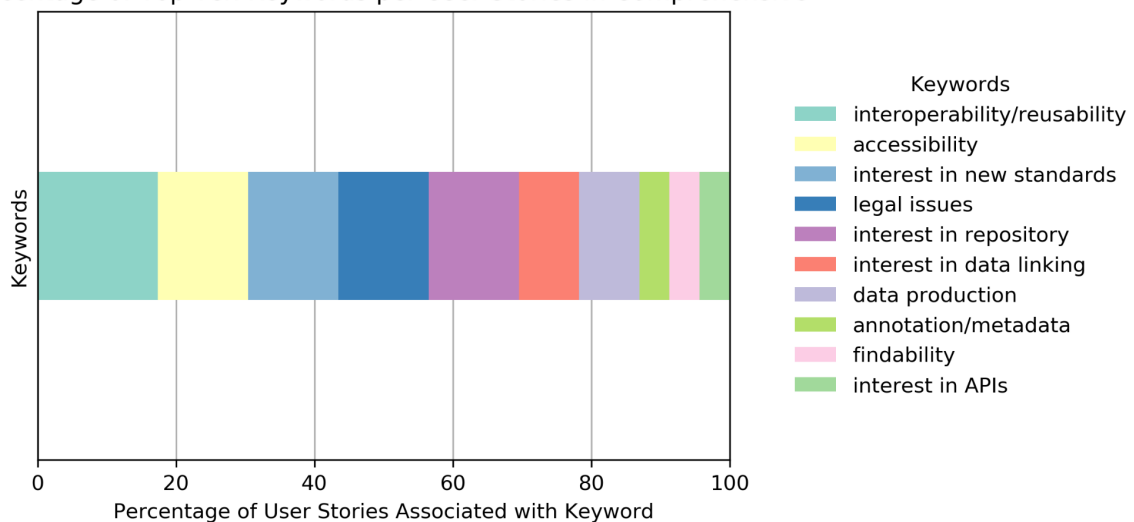


Figure 9. Percentage of Text+ user stories related to keywords in the data domain Comprehensive

Finally, the user stories in the Comprehensive group (Figure 9) show many differences to the rest of data domains because they cover broader subjects or are related to other disciplines. The interest in the FAIR principles is particularly strong here and also expressed in keywords like “interest in repository” (data publishing as a measure to enable reusability) or “interest in federated search” (as well as “findability”). The “interest in new standards” and “legal issues” is also linked to FAIR and relevant for all data domains.

Lessons Learned

Involving the community to elaborate on their requirements and demands regarding research infrastructures is a common procedure. In this section, a number of lessons learned with the help of the Text+ user stories are presented. They might be of use for other projects and consortia, also with regard to potential alternatives and associated drawbacks.

One aspect that caused a considerable extra amount of work was the decision to give contributors the choice of whether to submit their user story in English or in German (see also chapter *The User Story Template*). Due to the fact that the NFDI application process is conducted in English, but also because we wanted to address the international community, all the user stories written in German were translated by the editorial team. The translated user stories were then sent to the authors, who could suggest changes. Giving the opportunity to send user stories in German made it possible for many researchers to express their story not only in their mother tongue, but also in the language that they probably use most often in scholarly communication. However, the amount of extra work for translation was hard to estimate up front because we did not know how many user stories would be submitted and how many of them would need translation.

A further challenge was the fact that user stories are a new type of submission for the community. The kind of inquiries that the team received with regard to particular sections of the user story template made it clear that many authors did not know precisely what was expected of them. Particularly the last section “review by the community” (see also chapter “The User Story Template” in this report), was often misunderstood. In many cases, the editorial team provided individual consultation and sent examples when required. We considered facilitating the communication by publishing some examples beforehand on the website. On the one hand, this would have stated more clearly what was expected in each section. On the other hand, publishing three or four specific examples could have been misleading in that authors with very dissimilar cases to the ones published could have thought that their user story would not fit into the call at all. If such examples are to be included in the call, it is important to make it clear that they are in no way intended to exclude certain topic areas for the user stories.

For this call, we decided to publish a template as a Microsoft Word file, which the authors edited locally and submitted via e-mail. Other options, in particular various web forms, were considered but ultimately discarded. This template brought the following advantages:

1. the authors did not need to create extra accounts for the submission,
2. the format and associated software is widely used by researchers,
3. the format can be edited offline and saved locally.

Choosing other formats, such as LaTeX or TEI-conform workflows, could have led to a bias towards specific communities. However, this template also brought some disadvantages:

1. some authors deleted, merged, or changed sections, and
2. the fact that the keywording of the user stories had to be done by the team post-submission delayed the analysis considerably.

A possible alternative would be to provide contributors with a predefined set of keywords that could be expanded by the authors, if need be. In cases where the authors have not applied these keywords, the team responsible for editing and publishing the user stories could perform this task after submission in close cooperation with the authors. Moreover, the editors could add a new layer of standardised keywords or classes such as the FAIR criteria or [TaDiRAH](#),³⁶ and each project could subsequently evaluate the set of keywords for their stories. The authors of the stories would receive these suggestions and could finally accept, modify, or reject them. This would then facilitate the analysis of the stories while at the same time using keywords that the authors agreed on.

A combination of both an online form and the provision of a textual template could be a promising path forward. In the online form, authors could record formal aspects such as their names, institutions, ORCID identifier, title of story, subject area, data domain, and keywords in a structured manner. This could also be the place to clearly state whether their story should be analysed but not published, or whether metadata and text should be published. The text of the user story would then be written in a text document and be uploaded.

Outlook

The overwhelming response to our call – observable in the large number of user stories submitted over the span of only a few weeks even though this coincided with the summer break, as well as positive feedback expressed personally and via social media – supports our decision to maintain this action as an ongoing call. Beginning with the official start of Text+, we plan to renew the call (modified according to the lessons learned) and keep it as a constant review loop and exchange with the broad community of language- and text-oriented research in the humanities.

Besides the call for user stories, results from neighbouring projects and the broader Text+ network will be integrated: mention should be made of the SSHOC project with its [Open Marketplace Tester Community](#),³⁷ as well as the effort of community involvement in CLARIAH-DE, where different user types and groups are studied via qualitative interviews, community workshops and a usability study. Estimated types of users according to different levels of digital literacy or/and interaction with existing digital infrastructures are tested in a qualitative analysis. The information gathered via interviews, quantitative surveys and the call for user stories all help to get to know the variety of users and the heterogeneous

³⁶ <http://tadirah.dariah.eu/vocab/>

³⁷ <https://sshopencloud.eu/ssh-open-marketplace-testers>

community. Such user studies are valuable and useful across all initiatives providing digital research data infrastructures for the humanities. A collaboration across the NFDI consortia for the humanities started with the “Memorandum of Understanding”³⁸ and will be continued also regarding the community involvement.

References

- Arbeitsstelle Inschriften (2020): Inscriptions in Germany from the Middle Ages to Early Modern Times (Die Deutschen Inschriften). Text+ User Story, <https://www.text-plus.org/en/research-data/user-story-420/>.
- Barbot, Laure; Fischer, Frank; Larrousse, Nicolas; Petitfils, Clara; Illmayer, Klaus; Buddenbohm, Stefan (2020): SSH Open Marketplace. Public Consultation for the DARIAH Community (Version 1.0). Zenodo, [10.5281/zenodo.3935345](https://zenodo.org/record/3935345).
- Barbot, Laure; Moranville, Yoann; Fischer, Frank; Petitfils, Clara; Ďurčo, Matej; Illmayer, Klaus; Karampatakis, Sotiris (2019): SSHOC D7.1 System Specification - SSH Open Marketplace (Version 1.0). Zenodo, [10.5281/zenodo.3547648](https://zenodo.org/record/3547648).
- Bertino, Andrea (2020): Call for User Stories – Text+. DHd-Blog, <https://dhd-blog.org/?p=14043>.
- Bierwirth, Maik; Glöckner, Frank Oliver; Grimm, Christian; Schimmler, Sonja; Boehm, Franziska; Busse, Christian; Degkwitz, Andreas; Koepler, Oliver; Neuroth, Heike (2020): Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung (Version 1.0). Zenodo, [10.5281/zenodo.3895209](https://zenodo.org/record/3895209).
- Brünger-Weilandt, Sabine; Bruhn, Kai-Christian; Busch, Alexandra W.; Hinrichs, Erhard; Maier, Gerald; Paulmann, Johannes; Rapp, Andrea; Rummel, Philipp von; Schlotheuber, Eva; Schmidt, Dörte; Schrader, Torsten; Simon, Holger; Stein, Regine; Teich, Elke (2020): Memorandum of Understanding by NFDI Initiatives from the Humanities and Cultural Studies. (Version 2.0) Zenodo, [10.5281/zenodo.4045000](https://zenodo.org/record/4045000).
- Buddenbohm, Stefan; Barthauer, Raisa; Dung, Joseph (2020). User Requirements for the SSH Open Marketplace - Results from a SSHOC Workshop on January 30th 2020, Göttingen (Version 1.0). Zenodo, [10.5281/zenodo.3759174](https://zenodo.org/record/3759174).
- Calvo Tello, José; Rißler-Pipka, Nanette; Barthauer, Raisa; Buddenbohm, Stefan; Friedrichs, Sonja; Weimer, Lukas (2021): Text+ User Stories Data, DARIAH-DE Repository, <http://dx.doi.org/10.20375/0000-000E-67ED-4>.
- Gehring, Petra (2020): Lernen aus der ersten Auswahlrunde. 2. NFDI-Konferenz 8./9. Juli 2020, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_konferenz_2020/vortrag_gehring.pdf.
- German Council for Scientific Information Infrastructures (RfII) (2016): Enhancing Research Data Management: Performance through Diversity. Recommendations regarding

³⁸ (Brünger-Weiland et al. 2020)

structures, processes, and financing for research data management in Germany. Göttingen, <https://rfii.de/?p=2075>.

German Council for Scientific Information Infrastructures (RfII) (2018): Wide Impact for Research: NFDI Consortia as Stakeholders. Third discussion paper on the development of a national research data infrastructure (NFDI) in Germany. Göttingen, <https://rfii.de/?p=3818>.

Glöckner, Frank Oliver; Diepenbroek, Michael; Felden, Janine; et al. (2019): Berlin Declaration on NFDI Cross-Cutting Topics (Version 1.0). Zenodo, [10.5281/zenodo.3457213](https://doi.org/10.5281/zenodo.3457213).

Hastik, Canan (2020): Building Corpora for Comparative Analysis to Identify Declarations of Love in Letters. Text+ User Story, <https://www.text-plus.org/en/research-data/user-story-303/>.

Hubmann, Gerald; Roth, Regina (2020): Data security, system openness, networking – wishes for the NFDI from view of the MEGA. Text+ User Story, <https://www.text-plus.org/en/research-data/user-story-414/>.

NFDI Expert Committee (2020): The development of the National Research Data Infrastructure (NFDI). Second statement of the NFDI Expert Committee, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/stellungnahme_nfdi_2011_12_en.pdf.

Schill, Kerstin (2020): Der Aufbau der NFDI. 2. NFDI-Konferenz 8./9. Juli 2020, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_konferenz_2020/vortrag_schill.pdf.

Stegmeier, Jörn; Pelz, Peter F.; Bartsch, Sabine; Rapp, Andrea (2020): Semantic Uncertainty in DIN norms. Text+ User Story, <https://www.text-plus.org/en/research-data/user-story-605/>.

Vacano, Johannes von; Grüter, Doris (2020): Infrastructural Needs for Romance Research Data. Text+ User Story, <https://www.text-plus.org/en/research-data/user-story-316/>.

Warwick, Claire (2012): Studying users in digital humanities. In Claire Warwick, Melissa Terras, & Julianne Nyhan (Eds.), *Digital Humanities in Practice* (pp. 1-22). Facet. [doi:10.29085/9781856049054.002](https://doi.org/10.29085/9781856049054.002)

Wilkinson, Marc D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018, <https://doi.org/10.1038/sdata.2016.18>.

Wübbena, Thorsten (2020): Provision of processable textual data in libraries. Text+ User Story, <https://www.text-plus.org/en/research-data/user-story-321/>.

Further Links

CLARIN-D, <https://www.clarin-d.net/de/>.

DARIAH-EU, <https://www.dariah.eu/>.

DFG Research Funding: National Research Data Infrastructure, https://www.dfg.de/en/research_funding/programmes/nfdi/index.html.

DFG Subject classification system,
https://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp.

European Open Science Cloud (EOSC) – Communities,
<https://www.eoscsecretariat.eu/eosc-communities>.

Joint Science Conference (Die Gemeinsame Wissenschaftskonferenz),
<https://www.gwk-bonn.de/>.

NFDI association (NFDI-Verein), <https://www.nfdi.de/verein>.

NFDI4Culture: User Stories, <https://nfdi4culture.de/de/ressourcen/downloads.html>.

NFDI4Ing, <https://nfdi4ing.de/>.

NFDI4Memory: Problem Stories, <https://4memory.de/problem-stories-overview/>.

Research Data Alliance (RDA) – Disciplines, <https://www.rd-alliance.org/rda-disciplines>.

Research Data Infrastructure for the Humanities (Forschungsinfrastrukturen für die Geisteswissenschaften), <http://forschungsinfrastrukturen.de/doku.php>

SSHOC: Social Sciences and Humanities Open Cloud, <https://www.sshopencloud.eu/>.

SSH Open Marketplace Testers, <https://sshopencloud.eu/ssh-open-marketplace-testers>.

TaDiRAH - Taxonomy of Digital Research Activities in the Humanities,
<http://tadirah.dariah.eu/vocab/>.

Text+ Binding Letter of Intent,
https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen_2020/2020_nfdi_Text.pdf.

Text+ Data Call, <https://www.text-plus.org/en/research-data/data-from-the-community/>.

Text+ Data Domains, <https://www.text-plus.org/en/research-data/data-domains/>.

Text+ User Stories, <https://www.text-plus.org/en/research-data/user-stories-en/>.

Text+ User Stories Template, English version,
https://www.text-plus.org/wp-content/uploads/2020/09/Text_User-Stories_Template_en.pdf.

Text+ User Stories Template, German version,
https://www.text-plus.org/wp-content/uploads/2020/09/Text_User-Stories_Template_dt.pdf.