GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

2019

Impact and usability for digital humanities research infrastructures

Malte Vogl, Hanna-Lena Meiners, Klaus Thoden, Michael Haft, Oliver Schmid

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: This working paper is a report on the adaptation and implementation of the DARIAH–DE style guide as well as further requirements for the usability and success in the management of services in productive operation environments. By evaluating questionnaires that were answered by all software projects of the DARIAH–DE infrastructure, we are able to assess the overall usability and impact of DARIAH–DE. We address some areas, where the impact could be higher, by giving recommendations for action. The evaluation is explained in detail by giving step–by–step descriptions of the used software. The used data set and software are published in the DARIAH–DE repository, see Bibliography.

Impact and usability for digital humanities research infrastructures

Malte Vogl[1]   Hanna-Lena Meiners[2]   Klaus Thoden[1]
Michael Haft[3]   Oliver Schmid[4]

[1]Max Planck Institute for the History of Science
[2]Göttingen State and University Library, University of Göttingen
[3]Department Digital Academy, Academy of Sciences and Literature Mainz
[4]Technische Universität Darmstadt

DARIAH-DE
Working Papers

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Abstract

This working paper is a report on the adaptation and implementation of the DARIAH-DE style guide as well as further requirements for the usability and success in the management of services in productive operation environments. By evaluating questionnaires that were answered by all software projects of the DARIAH-DE infrastructure, we are able to assess the overall usability and impact of DARIAH-DE. We address some areas, where the impact could be higher, by giving recommendations for action. The evaluation is explained in detail by giving step-by-step descriptions of the used software. The used data set and software are published in the DARIAH-DE repository, see Bibliography.

Schlagwörter

Digitale Forschungsinfrastrukturen, Digitale Tools, Nutzbarkeit, Einfluss, Erfolgskriterien, Digital Humanities

Keywords

digital research infrastructures, digital tools, usability, impact, success criteria, digital humanities

Contents

1 Overview

This document reports on the adaptation and implementation of the DARIAH-DE style guide as well as the further requirements for usability and success in the management of services in productive operation environments. In addition, in DARIAH-DE's final phase of the project, a checklist has been produced to review the DARIAH-DE tools and services. It was created as part of the Working Group Service Life Cycle. The checklist addresses issues such as usability, interoperability, user-friendliness, and many more questions that should ensure sustainability. The questions were each answered by the project managers. To measure the impact of DARIAH-DE tools and services and to identify factors that could influence the impact, DARIAH-DE developed the Impactomatrix[1]. The Impactomatrix provides an overview of various criteria and factors that contribute to the impact of projects in the Digital Humanities. It serves as a means of identifying measures for the success of projects and virtual research environments. The Impactomatrix was developed within the framework of DARIAH-DE and will be continued by the DARIAH Working Group Impact Factors and Success Criteria. The abstraction of the results of the Service Life Cycle Checklists on the Impactomatrix is the main focus of this report.

1.1 Working Group Service Life Cycle

The working group Service Life Cycle (WG SLC) defined for itself the following tasks and goals:

- Development of internal procedures and methods, e.g. for the integration of services or the integration of new clusters in the DARIAH-DE research infrastructure.

- Support of the integration of services into the DARIAH-DE infrastructure and the DARIAH-DE portal (i.e. editorial team).

- Interdisciplinary orientation of the WG: Technical and scientific support and supervision of all development processes.

- Strategic planning to prepare for future challenges in the whole digital humanities area (including, for example, coordination with the DeISU[2]).

The WG Service Life Cycle thus serves on the one hand to monitor the smooth integration of services and tools into the DARIAH-DE infrastructure, but on the other hand also evaluates them. The WG is represented by one project partner each and by mentors, who accompany projects through the SLC and in particular support the transition into subsequent phases significantly. The mentoring team is fundamentally interdisciplinary, so that both the technical and the academic side of the service to be integrated can be supervised. There should also be contacts between service developers and experts by DARIAH-DE or associated partners.

---

[1]https://de.dariah.eu/begleitforschung
[2]DARIAH-DE eHumanities Infrastructure Service Unit

1.2 Impactomatrix

The accompanying research cluster of DARIAH-DE focused on the collection of success criteria and impact factors that could be used to evaluate digital humanities infrastructure projects and standardize their outcomes. In the DARIAH-DE Working Paper (Gnadt et al. 2017), a total of 101 relevant terms for the topic "impact in the digital humanities" were identified on the basis of own surveys and evaluation of the relevant literature. Central to this is the concept of **impact areas**, a collection of aspects that are particularly important to the functioning of DH projects. Linked to this are the impact factors and success criteria that are important for every specific area. For substantial work, the terms *impact*, *success*, *criterion* and *factor* were defined as follows:

- Impact refers to the form, degree, or diversity of a change in a group's behavior or attitude.

- Success is a positive response to a measure or product that is measurable in magnitude.

- Factors describe properties or means for changing a condition.

- Criteria describe concrete features for distinguishing between states of conditions.

These concepts work together as follows: Impact, that is influence, can be achieved by certain factors. Factors are thus the parameters, whereby impact can be influenced. Complementarily, the criteria are used as a measure with which Impact can be quantified. Other important elements are the requirements of the users of services and tools of a research infrastructure, which can either be explicitly or even implicitly expressed (because convention expects a certain functionality). These requirements vary in weight depending on the target group. In line with the requirements of different stakeholder groups, the success of a tool or research infrastructure can thus be at least partially assessed or measured.

The 101 identified terms were then divided into the three categories factor, criterion and impact area. This resulted in 67 factors, 25 criteria and 21 impact areas. For twelve of these terms, there were multiple assignments, mostly because a clear separation according to factors and criteria was not always possible. As an easy-to-use tool, these catalogs for success criteria and impact factors in the humanities were transformed into an overview that includes an easy to use access to the catalogs: the Impactomatrix. This tool provides a reference that collects factors and criteria for specific impact areas and then suggests pairings on how to increase impact in a particular relevant area. The Impactomatrix can be used as an aid in the evaluation of projects, firstmost to determine in which area a change should be made. On the basis of the factors and criteria collected for the area, own metrics can be set up, which are sensibly collected before and after a measure to increase the targeted impact area. In this way, project-specific key performance indicators can be collected.

In the following, the changes and implementations of the style guide will be discussed and the Impactomatrix will be used to evaluate the impact of the existing DARIAH-DE tools and services.

2 Adaptation and implementation of the style guide

Rules for a style guide were set up previously for DARIAH-DE and are documented in (Romanello, Stiller, and Thoden 2016). In cooperation with the WG Service Life Cycle, guidelines have been set up, that must be met in order to integrate new services and tools into the infrastructure. The focus has been laid on a user-friendly web design, a uniform and recognizable user interface for all components of the infrastructure as well as on detailed guidelines for the quality of the software. During this process, the following seven areas were identified

- documentation

- language and regional adaptation

- licensing

- platform type

- stability

- interoperability

- sample data

For each section, the style guide specifies optional, recommended, and necessary criteria that should be considered for the implementation as a DARIAH service. Finally, this version of the style guide includes a template for the user interface design in the form of mockups[3] as well as the implementation of the designs in code for websites based on the framework bootstrap, so that existing projects could easily customize the design accordingly.

---

[3]http://dariah-de.github.io/StyleGuideTemplate/

	Styleguide
Data Security / Safety	57%
External Impact	55%
Transfer of Expertise	54%
Education	54%
Usage	48%
Dissemination	43%
Innovation	41%
Competitiveness	41%
Relevance	41%
Reputation	40%
Coherence	40%
Collaboration	40%
Transfer of Knowledge	38%
Effectivity	33%
Communication	33%
Transparency	29%
Efficiency	29%
Sustainability	28%
Publications	25%
Funding Perspectives	25%
Integration	24%

Figure 1: Fulfillment of impact areas for the Styleguide

2.1 Adaption

A revised version of the style guide has been created according to templates of an external partner[4]. In addition to general updates such as better accessibility, the support of representations on mobile devices and the integration of logos as SVG files, particular attention was paid to the reduction of external dependencies. This helps ensuring a long-term maintainability of the style guide and reduces the risk of security vulnerabilities. The individual services were changed and adapted to the new styleguide by the respective developers. Any questions or difficulties encountered were recorded in the issue tracker[5] and form a valuable knowledge base for further implementations. For implementations using specific software combinations, there is an option to add reports to the wiki, see e.g. the entry about the implementation of the style guide for the Impactomatrix[6].

2.2 Implementation using the example of the Impactomatrix

The implementation of the new style guide is demonstrated using the example of the Impactomatrix. For the first implementation of the Impactomatrix a so-called template engine was used (Jinja2[7], based on the programming language Python). A template is a basic web page that contains general shape and dependency information. For the respective web pages derived therefrom, certain web page areas, such as the menu or the main content page, are then filled with the desired content according to specifications. Due to this structure of implementation, only the basic website had to be adapted for the new style guide. An example page in the style guide was a good starting point for this. Missing navigation options were just as easy to add as the special style information needed for the layout of the Impactomatrix. Since the Impactomatrix had no logo so far, the logo of the WG Impact & Success was chosen. The website has been presented as a github page[8] previously, so updating the web pages was also easy after a successful conversion of the style guide.

2.3 Classification of the impact of the style guide

For the style guide, as for all DARIAH-DE tools, a questionnaire for the WG SLC was filled out. An evaluation by a mapping to the factors of the Impactomatrix[9] shows that the Styleguide is particularly strong in the areas of Competence Transfer, Education, External Impact, Knowledge Transfer, and Usage (see figure 1). The effect is rather minor in the areas of Publication, Funding Prospects, Integration, and Transparency. This interesting finding shows that the functionality of the style guide is completely fulfilled. While the style guide plays an important role as a recognition factor and in the integration with other tools, this areas are not stronly fullfilled by itself. This is explained, among other things, by the fact that one of the questions of the WG SLC relates to the application of the style guide, see table 1 in the appendix.

---

[4]https://github.com/tschaef/dariah-de-2018-mockups/blob/gh-pages/design-primer.de.md
[5]https://github.com/DARIAH-DE/StyleGuideTemplate/issues
[6]https://github.com/DARIAH-DE/StyleGuideTemplate/wiki/Using-a-templating-engine
[7]http://jinja.pocoo.org/docs/2.10/
[8]https://pages.github.com/
[9]The method is described in depth in Section 3.1.

3 Usability criteria

The success factors for research infrastructures created in the context of the working group Impact Factors and Success Criteria made it possible to focus critically on the transfer of the offered services into productive operation. These include, for example, the availability of manuals and documentation for the various user groups (users / developers / administrators), the use of standard file formats, the availability of an API for easier interaction with other services, or the use of the DARIAH AAI for registration or sign in by users.

3.1 Evaluation method

The fulfillment of the success factors was monitored by the WG SLC and noted in checklists. The evaluation of these checklists then took place via an assignment of the questions for the checklist to the factors of the Impactomatrix. Figure 2 shows how often the respective questions of the WG SLC were positively answered by a service. With the exception of one question, all questions from the SLC checklist could be attributed to factors of the Impactomatrix, see table 1 in the appendix. The remaining question "What best describes the support?" could not be assigned to a specific factor. Each factor affects several impact areas as listed in table 2 (Appendix). Therefore, for each fulfilled factor of a service or tool, the corresponding impact areas were evaluated cumulatively. For example, if a service fulfills the factor "interoperability with other tools", the areas of Effectiveness, Efficiency, Funding Perspectives, Integration, Coherence, Collaboration, Sustainability, Usage, Publication and Relevance are addressed. Each of these areas thus receives a point. The excel spreadsheet containing the mapping between questions and impact factors as well as areas is published in the DARIAH-DE repository, see (Meiners 2019).

The individual regions may be satisfied with different frequencies by the corresponding factors, e.g. the area of Effectiveness can be fulfilled by 26 factors, whereas the area of Knowledge Transfer can only be fulfilled by 8 factors. In order to take this unequal distribution into account, the number of points achieved in the respective area was normalized by the number of maximum possible points. The evaluation took place in a Jupyter notebook, which is published in the DARIAH-DE repository, see (Vogl 2019).

	SLC Question	How often fullfilled
0	Is the current short description in the portal correct and appealing?	25
1	Is the current description in the portal correct and appealing?	25
2	Is there a description of the function of the service? (In the best case a user manual)	31
3	Description of the target group and their size	30
4	Is there an estimate of the cost or effort to run the service?	17
5	Which DARIAH-DE resources or other services are needed to operate the service?	31
6	Are only Open (Libre) Source licenses used to run the service? If not, which licenses are required?	29
7	Is there an API?	25
8	Which standard file formats are supported (for import as well as for export)?	27
9	Has a localization been implemented? If yes, which locales are supported? Which internationalization framework was used?	22
10	Is one instance of the service sufficient for the entire DARIAH-DE community or do projects / institutes / etc. each have their own instances?	32
11	What network or security requirements does the service provide to its environment?	21
12	Does a user documentation exist?	25
13	Is there a developer documentation available?	21
14	Is there a datasheet (fact sheet)	11
15	Are there any criteria as to when the service can be reinstated and a documentation on how this can be done professionally?	16
16	Has a connection to the DARIAH-AAI been made, as far as reasonably possible?	18
17	Has the current version of the DARIAH-DE style guide been integrated as far as reasonably possible?	15
18	Have integrations with other DARIAH-DE services (PID, repository, Bit-Preservation API, ...) been implemented, where reasonably possible?	18
19	If sensibel, is OAI-PMH supported? Has the integration been done with Generic Search?	14
20	Does a documentation of the test / verification suite exist for the service?	12
21	Who is the contact person for queries from users (technical / technical)?	32
22	Does an administrator documentation exist?	20
23	Are there plans for the operation of the service after the current project duration until 2019?	25
24	Are there any tutorials, FAQs or other training material?	23
25	Are there any application examples from research projects?	20

Figure 2: Distribution of fulfilled SLC Questions

3.1.1 Loading and evaluating excel spreadsheets

To load the data into the Jupyter Notebook environment, the Python package Pandas was used. This package allows data manipulation as well as plotting and was extensivly used for producing the figures of this report.

```python
import pandas as pd
df_allSheets = pd.read_excel(\
    'https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-D8CA-E/data',\
    sheet_name=[0,1,2])
```

The variable `df_allSheets` contains all excel spreadsheets as Pandas dataframes. To access, e.g., the spreadsheet on the attributes of the WG SLC in english, one picks the first element (the first element in a list in python is adressed by index 0).

```python
dfAttribute = df_allSheets[0]
```

3.1.2 Plotting and styling dataframes

To get an overview over the number of answered questions for the different projects, one first has to clean the data. Sometimes answers might have been entered as `True` or `False`. These are replaced by 1 and 0 respectively. Empty entries are treated as a No-Answer by replacing the value np.nan by 0.

```python
dfAttribute = dfAttribute.set_index('Attribut SLC Checkliste')\
    .replace([True,False,np.nan],[1,0,0])
```

One can then sum the number of answers for the different questions and create a new dataframe.

```python
dfAttributeDist = pd.DataFrame([x for x in zip(dfAttribute.index,dfAttribute.sum(axis=1))],\
    columns=['SLC Question', 'How often fulfilled'])


dfAttributeDist = dfAttributeDist.append({'SLC Question':'?','How often fulfilled': 0},\
    ignore_index=True)
```

The variable `dfAttributeDist` contains the summed result. For printing this result as an image, one can make use of the `.stlye` method for pandas Dataframes. For this, one first has to define a helper function to center text in cells.

```
align_center = dict(selector="th",
    props=[('text-align', 'center')])
```

The next step is to set an extra style to display the amount of asked questions as lightgrey bars in each cell. The helper function is applied with the `set_table_styles(...)` method.

```
dfAttributeDistStyle = dfAttributeDist.style.bar(axis=0, color='lightgrey', width=80.0)\
    .set_properties(subset=['SLC Question'],**{'width':'35em', 'text-align':'left'})\
    .set_table_styles([align_center])
```

The variable `dfAttributeDistStyle` corresponds to the displayed figure 2 if viewed in a Jupyter Notebook environment.

3.1.3 Fokus on impact areas

One first has to load the impact factors by calling

```
dfFactor = df_allSheets[1]
```

As before one has to unify the different formats of answers.

```
dfFactor = dfFactor.set_index('Impactfactors').replace([True,False,np.nan],[1,0,0])
```

One can then create a dictionary of possible impact ares and a list of all tools. The dictionary is initialized with zeros.

```
dfImpactAreas = df_allSheets[5]
areaList = list(
    set(
        [x.strip() for y in

            [x.split(',') for x in dfImpactAreas['Impact Area'].values]

        for x in y]

    )
)
areaDict = {x:0 for x in areaList}
toolList = dfFactor.columns.drop('SUM')
```

One can now check how each impact area is influenced by the different impact factors. This information is contained in the mapping of impact factors to impact areas, which are loaded into the variable `dictImpactAreas`.

```
dictImpactAreas = dfImpactAreas.set_index('Factor').to_dict()['Impact Area']
normDict = Counter([x.strip() for x in ','.join(dictImpactAreas.values()).split(',') if x])
dfNorm = pd.DataFrame([dict(normDict)]).transpose()
dfNorm = dfNorm.rename(columns={0:'Occurence'})
```

The dataframe `dfNorm` contains for each impact area the number of impact factors, which address it. Using the `.plot()` method for dataframes, one can create a bar graph of this distribution.
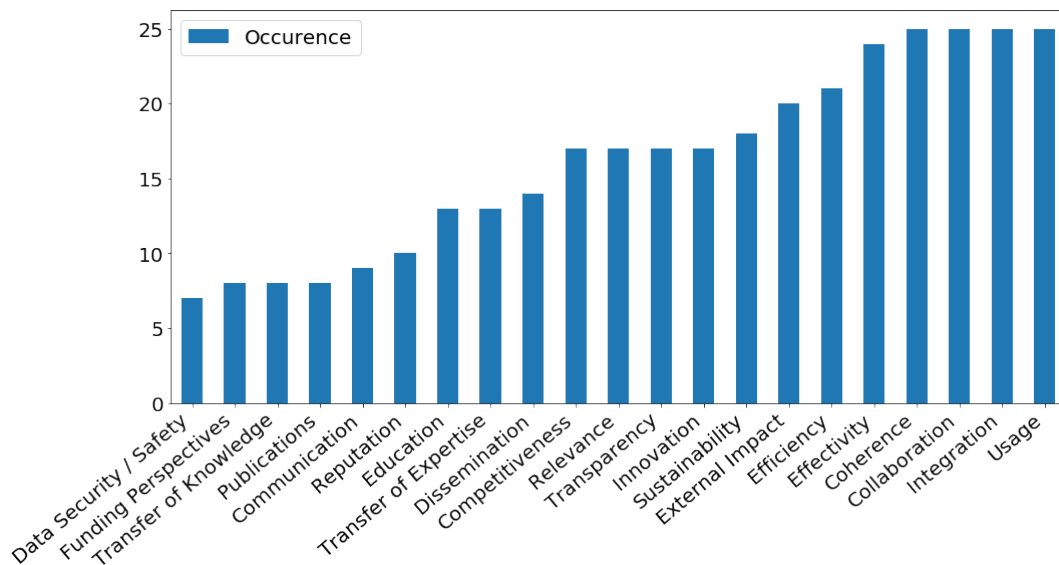

Figure 3: Normalization of impact areas

```
fig, axes = plt.subplots(nrows=1, ncols=1)
params = {'legend.fontsize': 20,
          'legend.handlelength': 2}
plt.rcParams.update(params)
dfNorm.sort_values('Occurence').plot.bar(figsize=(16,8),ax=axes)
axes.tick_params(labelsize=20)
fig.autofmt_xdate(rotation=40)
```

Running this procedure in a Jupyter Notebook setting produces figure 3. For the normalization of later graphics, one can make use of the dictionary `normDict`.

Using the above variables and dictionaries, one can create a matrix for the fulfillment of each impact area by the different tools. For this, a new dictionary `weight` is defined, which in the end will contain the necessary information.

```
weight = {}
```

The following routine creates for each tool a temporary copy of the dictionary containing the possible impact areas `areaDict`, see inline note #1. It then checks for each impact factor (c.f. #2), whether first, the impact factor is fulfilled (c.f. #3), and, if that is the case tries to get a list of all impact areas, which this factor affects (c.f. #4). It then increases the counter for each affected impact area by one (c.f. #5). After all factors are checked, the resulting temporary dictionary is saved as a value to the respective tool key in the `weight` dictionary (c.f. #6).

```python
for tool in toolList:
    tmpDict = areaDict.copy() #1
    for factor in factorList: #2
        if dfFactor[tool].loc[factor] != 0: #3
            try:
                areaList = [x.strip() for x in dictImpactAreas[factor].split(',') if x] #4
                for area in areaList:
                    tmpDict[area] += 1 #5
            except:
                print('Tool: {0}, Faktor {1}'.format(tool,faktor))
                pass
    weight[tool] = tmpDict #6
```

One can then create a dataframe containing the non-normalized impact areas for each tool.

```python
dfImpact = pd.DataFrame(weight)
```

3.1.4 Normalization

As one can see in figure 3, different impact areas have a differernt number of possible factors which address them. To take this information into account, one can use the keys for each impact area in the `normDict` dictionary as a normalization factor.

```python
dfNormed = dfImpact.copy()
for index in dfNormed.index:
    dfNormed.loc[index] = dfNormed.loc[index].apply(lambda row: row/normDict[index])
```

One can furthermore use the above styling attributes to come to a visualization of the fulfillment of impact areas in percentage.

For this aim, one has to define an additional helper function, which sets the text color according to percentage of fulfillment. This increases readability of the matrix.

```
def chg_color(val,param=75):
    color = 'black' if val < param else 'white'
    return 'color: %s' % color
```

Making use of the already defined function for centering text in columns, one can create a pandas `style` object which contains an overview over all tools and all impact areas. To come to a representation of percentage, one first multiplies the normalized numbers by 100 (c.f. #1) and then sets the format for numbers to zero precision, i.e. no numbers after the comma are displayed, and furthermore adds a percentage sign (c.f. #2).

```
dfNormed_styled = dfNormed\
.apply(lambda row: row*100).style.background_gradient(cmap='Blues')\ #1
.set_properties(**{'width':'6em', 'text-align':'center'})\
.set_table_styles([align_center])
dfNormed_styled.applymap(chg_color).format("{:,.0f}%") #2
```

Since this results in a large styled matrix object, in the following specific sub-areas of tools are discussed in details.

3.2 Evaluation

For better clarity, the following section will deal with thematically related services.

3.2.1 Basic services

Basic services form the basis for further data- or research-specific services. Confluence provides the DARIAH-DE Wiki, while Etherpad allows to edit documents collaboratively, for example the DARIAH-DE Working Papers. File Server, Monitoring and Helpdesk are ensuring availability of DARIAH-DE services.

Figure 4 shows the distribution of impact area fulfillment for basic services. The services show an overall inhomogeneous expression in the value distribution for the different impact areas. Striking are rather high values in the area of data security and education as well as relevance. In the areas of Reputation, Dissemination, and Transfer of Knowledge, the values are rather low, as was to be expected, since these services are not self-developed, such that documentation had no high priority.

	Confluence	Etherpad	Fileserver / mass storage	Helpdesk	Monitoring
Coherence	60%	56%	44%	72%	48%
Collaboration	60%	56%	44%	72%	48%
Communication	44%	33%	0%	44%	44%
Competitiveness	65%	53%	41%	71%	65%
Data Security / Safety	71%	57%	86%	86%	43%
Dissemination	36%	36%	7%	36%	36%
Education	77%	62%	46%	77%	77%
Effectivity	54%	46%	29%	67%	54%
Efficiency	52%	57%	38%	67%	48%
External Impact	55%	50%	30%	65%	60%
Funding Perspectives	50%	75%	25%	88%	50%
Innovation	47%	47%	18%	59%	47%
Integration	52%	56%	44%	68%	44%
Publications	25%	62%	25%	62%	25%
Relevance	53%	65%	41%	88%	59%
Reputation	50%	40%	30%	50%	40%
Sustainability	56%	61%	44%	78%	39%
Transfer of Expertise	62%	54%	31%	62%	62%
Transfer of Knowledge	38%	38%	25%	38%	38%
Transparency	59%	53%	24%	71%	47%
Usage	56%	56%	36%	68%	60%

Figure 4: Fulfillment of impact areas for basic services

3.2.2 Data services

The services, which directly concern the backup and provision of data, form a close-knit unit. The AAI service (Authentication and Authorization Infrastructure) forms the basis. With this service all members of DARIAH-DE as well as all users of the EduGAIN network can access DARIAH-DE services with their institutional account. Group rights can be used to regulate access to certain services. The publisher service "Publikator" allows users to publish research results in the DARIAH-DE Repository. In the development of the repository, the experience gained during the development of the TextGrid repository was used. Among other things, the Collection Registry serves to manage the metadata of the collections published in the repository.

	AAI	Repository	TextGrid Repository	Publikator	Collection Registry
Coherence	68%	56%	68%	76%	72%
Collaboration	68%	56%	68%	76%	72%
Communication	44%	22%	33%	44%	44%
Competitiveness	59%	65%	71%	82%	82%
Data Security / Safety	71%	86%	86%	71%	71%
Dissemination	36%	36%	50%	50%	50%
Education	62%	69%	69%	77%	77%
Effectivity	54%	50%	62%	67%	62%
Efficiency	57%	48%	57%	67%	67%
External Impact	60%	50%	70%	75%	75%
Funding Perspectives	75%	50%	88%	88%	88%
Innovation	47%	47%	53%	59%	59%
Integration	64%	56%	64%	64%	64%
Publications	62%	25%	62%	50%	50%
Relevance	82%	59%	76%	82%	76%
Reputation	30%	50%	50%	50%	50%
Sustainability	67%	61%	72%	72%	72%
Transfer of Expertise	54%	62%	54%	62%	62%
Transfer of Knowledge	50%	25%	50%	50%	50%
Transparency	53%	53%	65%	71%	65%
Usage	68%	56%	68%	76%	76%

Figure 5: Fulfillment of impact areas for data services

Figure 5 shows the distribution of fulfilled impact areas for data services. All services show high values in impact areas such as Usage, Relevance, or Education. Additionally, the areas of Competitiveness, Funding Perspectives, and Data Security/Safety are also very strong. In the good midfield, there are areas like Coherence, Effectivity, Integration, and Collaboration. These services are rather weakly positioned in areas of external communication, such as e.g. Reputation, Dissemination, Communication, and Publication.

3.2.3 Research specific services

The research-specific services show an even more inhomogeneous picture. Figure 6 shows the values for five different tools. The GeoBrowser[10] allows the seamless representation of geographic and temporal relationships. Topics[11] and DKPro wrapper[12] are tools for evaluating text collections using Natural Language Processing methods. Cosmotool[13] is used to display biographical information as movement profiles. ConedaKOR[14] is a tool for the administration and representation of data collections for image-based humanities research. In the areas of Data Security/Safety, Funding Perspectives and Relevance, all the above tools show high scores. While all tools achieve low scores in the area of Dissemination, in the area of Communication the field is wider. While ConedaKOR achieves 56% of the possible points and DKPro-Wrapper still receives 33%, the remaining tools are lagging far behind with 0-11%. At the same time, there is the overall lower fulfillment of impact areas by the DKPro wrapper compared to the rest of the tools. For example, the other tools are much better positioned in areas such as Relevance, Competitiveness, Sustainability or Effectiveness. One reason for this could be the favoring of the development of the DARIAH-DE topic modeling framework, for which a stand-alone application, the Topic Explorer, was developed, which can take over some use cases of the DKPro Wrapper. Another reason is certainly the fact that it is an interface software to the main tool DKPro Core, so that less effort was invested in the fulfillment of the quality criteria.

---

[10] https://geobrowser.de.dariah.eu/
[11] https://de.dariah.eu/topics
[12] https://de.dariah.eu/dkpro-wrapper
[13] https://cosmotool.de.dariah.eu/cosmotool/personsearch/
[14] https://conedakor-demo.de.dariah.eu/

	ConedaKOR	Topics	Cosmotool	GeoBrowser	DKPro-Wrapper
Coherence	64%	48%	40%	52%	36%
Collaboration	64%	48%	40%	52%	36%
Communication	44%	11%	11%	0%	22%
Competitiveness	71%	53%	41%	59%	35%
Data Security / Safety	86%	86%	86%	57%	86%
Dissemination	36%	29%	36%	36%	14%
Education	77%	54%	31%	46%	23%
Effectivity	58%	42%	42%	46%	25%
Efficiency	57%	38%	38%	48%	29%
External Impact	65%	50%	40%	55%	35%
Funding Perspectives	75%	62%	75%	88%	38%
Innovation	53%	41%	35%	35%	18%
Integration	60%	52%	44%	52%	40%
Publications	50%	25%	50%	50%	25%
Relevance	76%	59%	59%	71%	41%
Reputation	50%	50%	40%	40%	20%
Sustainability	67%	56%	61%	56%	44%
Transfer of Expertise	62%	54%	23%	38%	15%
Transfer of Knowledge	38%	25%	25%	38%	25%
Transparency	59%	41%	41%	35%	29%
Usage	64%	48%	48%	60%	32%

Figure 6: Fulfillment of impact areas for research tools

3.3 Summary usability criteria

In order to get a comprehensive impression of the fulfillment of the different impact areas in all tools, it is useful to consider the sum of all normalized areas. As shown in figure 6, services are most likely to meet the Data Security / Safety area (24 out of 34 possible). The least satisfied area is Dissemination, which only reaches a score of 11 points out of 34 possible points. Overall, three groups of areas can be identified. In the range of under 15 fulfilled points Dissemination, Reputation, Communication, Publication, Innovation and Knowledge Transfer constitute a group. In the midfield of 15 to 20 points achieved are the areas of Transparency, Transfer of Expertise, Efficiency, External Impact, Effectivity, Integration, Coherence, Collaboration and Sustainability. The last group in which areas reached more than 20 points consists of Usage, Competitiveness, Education, Relevance, Funding Perspectives and Data Security/Safety. The first group consists of areas that can be assigned to the comprehensive topic of external representation. In the second group there are areas of user interaction and the third group is composed of areas related to higher level institutions and their interaction. This rough classification suggests that during the development of DARIAH-DE tools and infrastructures great care has been taken on feedback from institutions and users. The self-presentation to third parties, however, seems to have been less important.
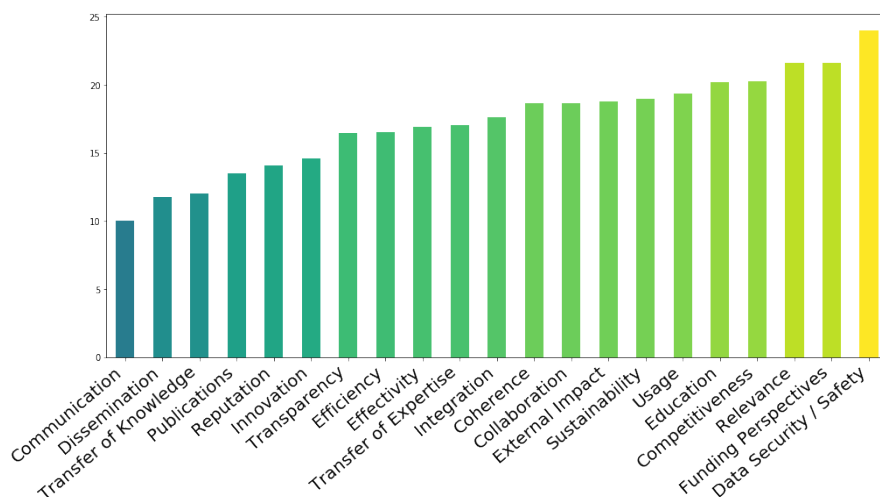
Figure 7: Sum of fulfilled impact areas

4 Summary

As described in "Summary usability criteria" in Section 3.3, the DARIAH-DE tools and services differ in terms of their fulfillment for certain impact areas. This is partly due to the fact, that the development of most of the services and tools initially occurs independently of each other and is always carried out by only one responsible person. The integration into the DARIAH-DE infrastructure usually takes place parallel to the development work or even afterwards, which of course is not necessarily desirable per se, but in some cases can not be avoided. It is also the case, that the different tools and services set

different priorities. Of course, data security and privacy is a key issue in the development of a repository and related services. In general, this impact area has a relatively high status in the development and thus forms one of the basic pillars within DARIAH-DE. Nonetheless, areas that tend to fare poorly after the evaluation are also likely to receive more attention. This applies above all to Dissemination, Reputation, Communication, Publication, Innovation, and Knowledge Transfer. However, it should also be noted that this result is a cross-section of all DARIAH-DE offers, which are in general very diverse. Offers such as the styleguide or the AAI can not be evaluated on criteria such as Data Dissemination, as this is not the purpose of the services. Accordingly unfavorable is the rating of these services in special impact areas. In the following, the weakly-cutting areas will be examined in more detail in order to give possible recommendations for action so that exactly these areas can be positively influenced. For this purpose, it was first examined which factors can be assigned to these areas. For a more precise definition or explanation of the affected areas, readers should consult the Impactomatrix to understand as fully as possible the process. From table 2 it can be seen that there are different possibilities of influencing the allocated area.

A closer analysis reveals that each of the identified areas can be influenced by the factor of knowledge transfer. In addition, four of the five areas can be influenced by the factors of data dissemination and the promotion of exchange of experience. If this result is formulated as a concrete recommendation for action, it should be carefully targeted throughout DARIAH-DE to publish all the steps and results before, during and after development, and to make them available for discussion, so that a dissemination of knowledge does not just remain in small circles. This not only promotes the exchange of experiences and community building, but also opens up the possibility of recognizing, discussing, and remedying existing known or undiscovered problems. Of course, the parallel development and documentation in form of publications is challenging, yet it should be tried to make the large base of knowledge that DARIAH-DE has undoubtedly acquired, more widely available.

Appendix

Mapping of SLC questions to impact factors

The questions of the WG SLC correspond to different impactomatrix factors. Each factor in turn influences different impact areas. The mappings between these components are listed below.

Double entries are included to be able to weight the fulfillment of impact areas.

The excel document containing all datasets is published as (Meiners 2019).

Table 1: Mapping of SLC questions to impact factors

	WG SLC question	Impactomatrix-Factor
0	Is the current short description in the portal correct and appealing?	Providing information and outcomes, Public relations
1	Is the current description in the portal correct and appealing?	Providing information and outcomes, Public relations
2	Is there a description of the function of the service available? (ideally a user manual)	Providing information and outcomes
3	Description of the target group and its size	Community building
4	Does an estimate of the cost or effort to run the service exist?	Measures for long-term use & storage
5	Which DARIAH-DE resources or other services are needed to operate the service?	Re-usability of infrastructure, Embedding of available digital databases/software/tools
6	Are solely Open (Libre) Source licenses used to run the service? If not, which licenses are required?	Open-source (offer), Improved access to resources
7	Is there an API available?	Software-interfaces, Import/Export-functionalities
8	Which standard file formats are supported (for import as well as for export)?	Import/Export-functionalities, Usage & support of standards
9	Has a localization been implemented? If yes, which locales are supported? Which internationalization framework was used?	Multilingualism, Communication facilitation/acceleration
10	Is one instance of the service sufficient for the entire DARIAH-DE community or do projects / institutes / etc. each have their own instances?	Performance, Stability, Availability

Table 1, continued

	WG SLC question	Impactomatrix-Factor
11	What network or security requirements does the service provide to its environment?	Performance, Regular updates (contents & functionalities), Stability
12	Does a user documentation exist?	Documentation of functionalities, Learnability, Operability/Usability
13	Is there a developer documentation available?	Documentation of code, Documentation of functionalities, Operability/Usability, Technical support
14	Is there a datasheet (fact sheet) available	Providing information and outcomes
15	Are there any criteria established as to when the service can be reinstated and a documentation on how this can be done professionally?	Measures for long-term use & storage, Stability, Perfomance
16	Has a connection to the DARIAH-AAI been made, as far as reasonably possible?	Accessibility, Re-usability of infrastructure, Improved access to resources
17	Has the current version of the DARIAH-DE style guide been integrated as far as reasonably possible?	Recognition value, Appealing layout/web surface, Operability/Usability
18	Have integrations with other DARIAH-DE services (PID, repository, Bit-Preservation API, ...) been implemented, where reasonably possible?	Scalability & modularity, Interoperability with other tools, Configurable functionalities, Re-usability of infrastructure, Interoperability with digital resources
19	If reasonable, is OAI-PMH supported? Has the integration been done with Generic Search?	Software-interfaces, Interoperability with other tools, Interoperability with digital resources
20	Does a documentation of the test / verification suite exist for the service?	Stability
21	Who is the contact person for queries from users (professional/technical)?	User support, Technical support
22	Does an administrator documentation exist?	User support,Documentation of code,Operability/Usability,Technical support
23	Are there any plans for the operation of the service after the current project duration until 2019 worked out?	Conservation of data, Availability, Data management
24	Are there any tutorials, FAQs or other training material available?	Learnability, User support, Documentation of functionalities, Operability/Usability
25	Are there any application examples from research projects available?	Learnability,Integration into scientific workflows,User involvement

Mapping of impact factors to impact areas

The corresponding Jupyter Notebook which analyses and visualizes the mappings is published in the DARIAH-DE repository as (Vogl 2019).

Table 2: Mapping between impact factors and impact areas

	Impact factor	Impact areas
0	Open-source (offer)	External Impact, Dissemination, Effectivity, Funding Perspectives, Innovation, Sustainability, Reputation, Transparency, Competitiveness
1	Anonymity (collaboration & communication)	Effectivity, Efficiency, Integration, Communication, Transfer of Expertise, Usage, Transfer of Knowledge
2	Appealing layout/web surface	External Impact, Dissemination, Coherence, Collaboration, Usage, Competitiveness, Transfer of Knowledge
3	User support	Education, Effectivity, Efficiency, Coherence, Collaboration, Transfer of Expertise, Usage, Relevance, Transparency
4	Auswertbare Server-Logs	Data Security / Safety, Efficiency, Integration, Transparency
5	Operability/Usability	External Impact, Education, Effectivity, Efficiency, Coherence, Collaboration, Transfer of Expertise, Usage, Relevance, Competitiveness
6	Providing information and outcomes	External Impact, Education, Dissemination, Effectivity, Integration, Sustainability, Usage, Publications, Relevance, Reputation, Transparency, Transfer of Knowledge
7	Community-Building	External Impact, Dissemination, Effectivity, Innovation, Communication, Transfer of Expertise, Usage, Relevance
8	Dissemination of data	External Impact, Dissemination, Innovation, Sustainability, Usage, Publications, Reputation, Transparency, Transfer of Knowledge
9	Conservation of data	Data Security / Safety, Coherence, Collaboration, Sustainability
10	Data management	Data Security / Safety, Efficiency, Integration, Coherence, Collaboration, Sustainability, Reputation, Transparency, Competitiveness
11	DH functionalities	Education, Effectivity, Efficiency, Innovation, Integration, Coherence, Collaboration, Sustainability, Reputation, Transparency, Competitiveness

Table 2, continued

	Impact factor	Impact areas
12	Documentation of functionalities	External Impact, Education, Effectivity, Efficiency, Innovation, Integration, Coherence, Collaboration, Transfer of Expertise, Usage, Relevance, Reputation, Competitiveness
13	Documentation of code	Education, Innovation, Integration, Coherence, Collaboration, Transfer of Expertise, Sustainability, Transparency
14	Embedding of available digital databases/software/tools	External Impact, Education, Effectivity, Competitiveness, Transfer of Knowledge, Efficiency, Integration, Coherence, Collaboration, Sustainability, Usage, Relevance
15	User involvement	External Impact, Efficiency, Innovation, Coherence, Collaboration, Communication, Transfer of Expertise, Sustainability, Usage, Relevance, Transparency, Competitiveness
16	Einbindung ins Dateisystem	Efficiency, Integration, Coherence, Collaboration, Usage
17	Integration into scientific workflows	Education, Effectivity, Efficiency, Integration, Coherence, Collaboration, Usage, Relevance, Competitiveness
18	Meaningful & significant name	External Impact, Dissemination, Coherence, Collaboration, Competitiveness
19	Dissemination of knowledge (subject-specific and DH-broad)	External Impact, Dissemination, Effectivity, Innovation, Communication, Publications, Relevance, Reputation, Transfer of Knowledge
20	Recognition value	External Impact, Dissemination, Usage, Competitiveness
21	Learnability	Education, Efficiency, Transfer of Expertise, Usage
22	Enabling online-work	Education, Effectivity, Efficiency, Innovation, Integration, Usage, Competitiveness
23	Support of experience exchange	External Impact, Education, Dissemination, Effectivity, Innovation, Communication, Transfer of Expertise, Reputation
24	Import/Export-functionalities	Effectivity, Integration, Coherence, Collaboration, Competitiveness
25	Interoperability with other tools	Effectivity, Efficiency, Funding Perspectives, Integration, Coherence, Collaboration, Sustainability, Usage, Publications, Relevance
26	Interoperability with digital resources	Effectivity, Efficiency, Funding Perspectives, Integration, Coherence, Collaboration, Sustainability, Usage, Publications, Relevance

Table 2, continued

	Impact factor	Impact areas
27	Collaboration functionalities	Effectivity, Efficiency, Integration, Communication, Transfer of Expertise, Usage, Transfer of Knowledge
28	Communication facilitation/acceleration	External Impact, Efficiency, Communication, Transparency
29	Configurable functionalities	Effectivity, Coherence, Collaboration, Relevance, Transparency
30	Workflow management	Efficiency, Integration, Coherence, Collaboration, Communication, Transfer of Expertise, Sustainability, Publications
31	Measures for long-term use & storage	Data Security / Safety, Integration, Sustainability, Publications, Relevance
32	Multilingualism	Education, Effectivity, Coherence, Collaboration, Communication, Publications, Transfer of Knowledge
33	Re-usability of infrastructure	Efficiency, Funding Perspectives, Integration, Sustainability
34	User surveys & tests	Dissemination, Effectivity, Innovation, Coherence, Collaboration, Transparency
35	Public relations	External Impact, Funding Perspectives, Usage, Relevance, Competitiveness
36	Performance	Data Security / Safety, Efficiency, Integration, Coherence, Collaboration, Usage, Competitiveness
37	Regular updates (contents & functionalities)	External Impact, Innovation, Relevance
38	Scalability & modularity	Efficiency, Innovation, Integration, Sustainability, Transparency
39	Stability	Data Security / Safety, Integration, Usage
40	Software Schnittstellen	Effectivity, Integration, Transparency
41	Technical support	Data Security / Safety, Innovation, Coherence, Collaboration, Sustainability
42	Support of successful scientists	External Impact, Dissemination, Funding Perspectives, Innovation, Transfer of Expertise, Usage, Relevance, Reputation
43	Support of open file formats	Dissemination, Effectivity, Integration, Coherence, Collaboration, Sustainability, Transparency
44	Improved access to resources	External Impact, Dissemination, Funding Perspectives, Innovation, Transfer of Expertise, Usage, Relevance, Reputation
45	Availability	External Impact, Effectivity, Integration, Coherence, Collaboration, Usage, Competitiveness

Table 2, continued

	Impact factor	Impact areas
46	Usage & support of standards	Funding Perspectives, Integration, Coherence, Collaboration, Sustainability, Transparency
47	Accessibility	External Impact, Education, Dissemination, Effectivity, Innovation, Usage, Transparency, Competitiveness

Bibliography

Gnadt, Timo, Viola Schmitt, Juliane Stiller, and Klaus Thoden. 2017. "Faktoren Und Kriterien Für Den Impact von DH-Tools Und Infrastrukturen." 21. Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen. http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2017-1-7.

Meiners, Hanna-Lena. 2019. "Evaluation Slc Attributes." DARIAH-DE. https://doi.org/10.20375/0000-000b-d8c9-f.

Romanello, Matteo, Juliane Stiller, and Klaus Thoden. 2016. "Usability Criteria for External Requests of Collaboration (DARIAH2 R 1.2.2/R 7.5)." https://wiki.de.dariah.eu/download/attachments/14651583/R1.2.2-7.5_final.pdf.

Vogl, Malte. 2019. "Analysis of Impact and Usability - Notebook." DARIAH-DE. https://doi.org/10.20375/0000-000B-D8CB-D.