Originally published in: Lindén, Krister/Kontino, Thalassia/Niemi, Jyrki (Eds.): Selected papers from the CLARIN Annual Conference 2023. – Linköping: Linköping University Electronic Press, 2024. 10 Pp.

Standards Information System for CLARIN centres and beyond

Piotr Bański
Department of Digital Linguistics
IDS Mannheim, Germany
banski@ids-mannheim.de

Eliza Margaretha Illig
Department of Digital Linguistics
IDS Mannheim, Germany
margaretha@ids-mannheim.de

Abstract

The present contribution describes features of the CLARIN Standards Information System that have been designed to assist data deposition centres in CLARIN. We also show what is needed and what has been done in order to go beyond the originally designated target, so as to provide service to sibling and descendant research infrastructures, of which DARIAH and Text+ are taken as examples. This paper is aimed primarily at representatives of research infrastructure nodes, responsible for preparing and sharing data deposition information about their centres or repositories. It assumes a degree of technical knowledge or experience in using the XML format and tools, the REST API, and version control systems.

1 Introduction

Many modern research infrastructures (RIs) offer data deposition services for their users. For CLARIN B-centres, the provision of this service is a default characteristic that is subject to certification requirements and that is used as a basis of a measurement needed to calculate one of the CLARIN-ERIC Key Performance Indicators.

The range of data that constitutes language resources or accompanies them is very wide, from the prototypical electronic corpora and dictionaries through, among others, participant lists, tagsets, digital facsimiles, raw audiovisual datasets, language models of various complexity and size, and ending with datasets produced by behavioural or neurolinguistic experiments, as well as documentation of various kinds. Neither the kinds of data nor the formats used to encode it are exclusive to CLARIN. CLARIN's focus has historically overlapped with some areas served by DARIAH and, by a natural extension, with CLARIAH networks that combined DARIAH and CLARIN nodes in some of the European countries, at various points in time. In Germany, the national CLARIN-D merged with DARIAH-DE into CLARIAH-DE in 2019, and, since 2022, many former German DARIAH and all the former CLARIN-D centres (as well as some centres previously not belonging to either of the two) have formed the Text+ consortium, which is part of the German National Research Data Infrastructure, NFDI.[1]

This is illustrated in Figure 1, which does not take historical developments into account, but is rather meant to hint at the resulting inter-RI relationships. The reader should bear in mind that, while CLARIN and DARIAH are multinational networks, Text+ is restricted to Germany.

This paper showcases the Standards Information System (SIS) in the context of an extended network of inter-RI relationships. The main purpose of the SIS since around the year 2021 has been to serve as a platform for sharing and collecting information about data deposition formats supported by CLARIN centres, in lieu of centre-specific recommendations, provided individually in the form of lists or tables, differing in structure and granularity. The information is crucial to end-users who wish to deposit their data for the purpose of archiving or reuse, but it also provides an important insight into the network as a whole: the aggregated recommendations indicate trends in the usage of the particular data formats: a format may be labelled as "recommended", "acceptable", or "discouraged", and – on the safe assumption

[1] References to the home pages of the research infrastructures mentioned in this paper are gathered at the end.

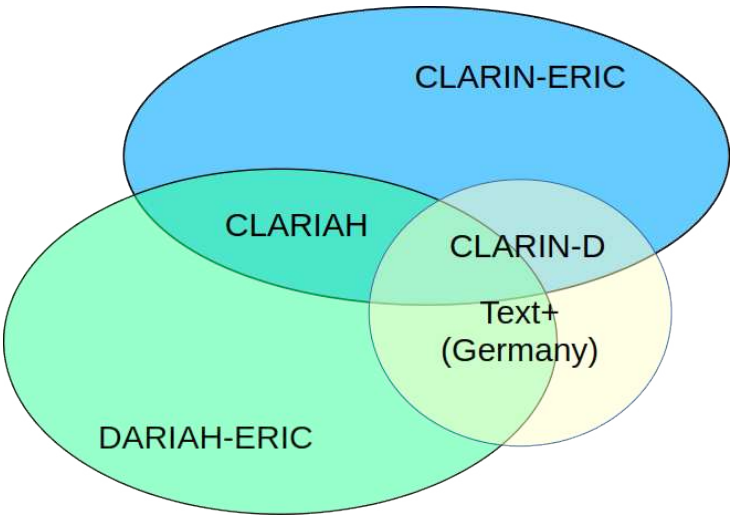

Figure 1: Relationships between selected research infrastructures representing broadly understood Human Language Technology as well as Arts and Humanities.

that CLARIN centres follow technological innovations – the changing recommendations may constitute an important input to standardisation initiatives as well as to the industry at large.

Apart from its informative goal, the SIS attempts to assist the user in thinking about their data: about the ways in which the various parts of the user's deposited data function and also about the formats that data may come in. Sections of the SIS, available from the side menu on its homepage, provide, among others, information on the functions of data in the Human Language Technology (HLT) setting as well as information on the data formats and the relationships both among formats and between formats and the standards that the formats are typically tied to.[2]

The following simplified definitions are assumed in this paper: a **standard** is a document that, following a systematic process of community consultations and revisions carried out by a standards setting organisation, sums up and recommends the best practices for dealing with certain tasks; a **(technical) specification** is similar to a standard, but its origin is less procedural and more community-oriented; a specification often enjoys the status of a *de facto* standard, before it becomes institutionally codified and disseminated. A **data format** is a serialisation of a data model defined by a standard or a specification. Note that this is a very broad statement that denotes, for example, both the XML format as a serialisation of the well-known XML standard defined by W3C, and a very narrowly defined application of XML such as a particular corpus-encoding format compliant with the ISO MAF (Morphosyntactic Annotation Format), heavily restricted by additional data models superimposed on the general XML data model. The end-user rarely has the expertise to distinguish between such cases, and it is part of the task of the SIS to suggest that, among others, "XML" alone is relatively meaningless in the context of data formats, and that it should be further qualified in order to ensure that the user's data is sustainable and interoperable – which are the usual aims of data deposition.

The paper is organized as follows. The history of the development of the SIS is briefly recounted in Section 2. The current features of the SIS and the extended features that support other RIs are elaborated on in Sections 3 and 4, respectively. Related work is presented in Section 5. Section 6 provides a summary of the main points made in the paper and indicates the paths for further development.

[2] See Figure 4 for an illustration of both the side menu and a part of a format information page.

2 Standards Information System: basic information

2.1 From CLARIN Standards Guidance to CLARIN SIS

The current CLARIN Standards Information System[3] extends the former CLARIN Standards Guidance (Stührenberg et al., 2012), contributed to the CLARIN infrastructure by CLARIN-D. Originally, the system provided information about various HLT standards and indicated relationships among them. The practical aim of the Standards Guidance was to assist users in finding standards most appropriate for their purposes. In some cases, names or abbreviations of CLARIN centres claiming to use those standards were provided, so that the user knew which centre to choose for the purpose of depositing data encoded in some specific formats (see Figure 2 for a simplified data model of the original system, indicated by the grey background). A side goal was to provide a taxonomy or even a small knowledge base of standards and technical specifications, served by eXist-DB (Siegel & Retter, 2014).[4]

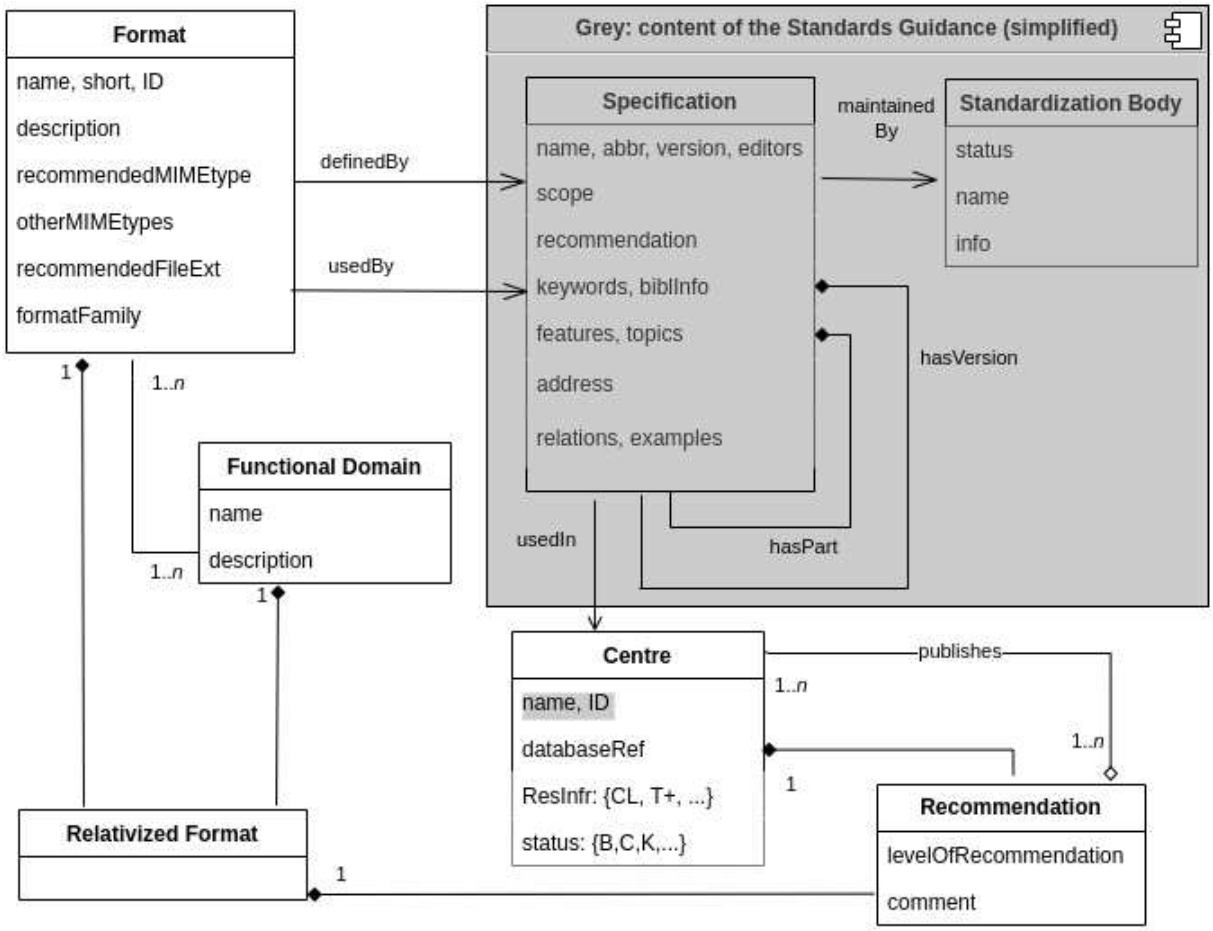

Figure 2: Simplified original data model (on grey background) with additions designed to incorporate format recommendations and research infrastructures other than CLARIN. Filled diamond arrows represent strong aggregation, hollow diamond arrows – weak aggregation, while simple relationships are represented by lines, with simple arrowheads pointing to the objects of the relationships.

Around the year 2019, the CLARIN Standards Committee undertook the task of providing a platform for CLARIN centres to share their recommendations concerning the formats for data that could be de-

[3]The SIS can be accessed at https://standards.clarin.eu/sis/, which is an alias for https://clarin.ids-mannheim.de/standards/. Its GitHub home is at https://github.com/clarin-eric/standards and the documentation is in the project wiki at https://github.com/clarin-eric/standards/wiki. The SIS is listed as a knowledge base at Fairsharing.org: https://fairsharing.org/4705.

[4]The content of the Standards Guidance is still available in the current SIS, in the "Standards and Specifications" menu item on the left. It is not actively maintained.

posited at each of them. That task eventually resulted in adapting and extending the Standards Guidance to become the tool to aggregate the information provided by centres and to visualise it in various ways. Figure 2, improving upon Bański and Hedeland, 2022, shows the extensions to the original data model needed to embrace the new functionality. The Standards Committee faced several challenges, among others concerning the classification of formats, the initial retrieval of lists of data deposition formats for some centres, or devising the least troublesome ways for the particular centres to submit their recommendations to the system. Due to the diversity of ways in which data formats are used in HLT research, it is also a challenge to visualise the recommendations and maintain them with minimal effort. The deliverable has evolved since 2021 from a complex set of Google spreadsheets that put together formats, format categories, and CLARIN centres, to the current XML format integrated in the SIS.[5]

2.2 The SIS data model

In Figure 2, the greyed area of the data model represents, with minor simplifications, the state of the CLARIN Standards Guidance up to 2021. In some cases, information about standards mentioned centre names, hence the grey background in the Centre class. In the SIS, centres are represented consistently, with an indication of the research infrastructure that they belong to (see Section 4) and of their status within that infrastructure. For CLARIN, this means their status as B- or C-centres.[6] CLARIN centres reference the CLARIN centre registry [7] as the authoritative source of information.

Format	Centre	Domain	Recommendation	
PDF/A	EKUT	Image Source Language Data	recommended	
PDF/A	EKUT	Textual Source Language Data	recommended	
PDF/A	FIN-CLARIN	Documentation	recommended	
PDF/A	FIN-CLARIN	Textual Source Language Data	acceptable	
PDF/A	IDS	Documentation	recommended	
PDF/A	IDS	Image Source Language Data	recommended	
PDF/A	DANS	Documentation	recommended	ⓘ
PDF/A	DANS	Other	acceptable	ⓘ
PDF/A	DANS	Textual Source Language Data	recommended	ⓘ
PDF/A	MI	Image Source Language Data	recommended	
PDF/A	MI	Textual Source Language Data	recommended	
PDF/A	ZIM	Image Source Language Data	recommended	
PDF/A	ZIM	Textual Source Language Data	recommended	
PDF/A	LAC	Contextual Data	recommended	

Figure 3: Fragment of format recommendations by CLARIN centres concerning the PDF/A format. Centres may comment on their recommendations (the circled *i* shows the comment in a pop-up).

A crucial element of the SIS is the set of functional data domains that serve to fine-tune the purposes for which the individual data items are collected: for example, data encoded in the PDF/A format are perfect for the purpose of documentation, but definitely not ideal for the purpose of providing annotation for audiovisual sources, or collections of statistical data. This is illustrated in Figure 3, which is a screenshot of a fragment of aggregated format recommendations.

While functional domains are hard-coded in the system, the instantiation of the Format class in the data model of Figure 2 may take two forms. In most cases, there already exists a format description – a

[5] Much of the history behind the task described here is documented at https://www.clarin.eu/content/standards.

[6] K-centres are typically outside the scope of the SIS, unless they are paired with a centre that offers data depositions.

[7] The CLARIN centre registry is available at https://centres.clarin.eu/ and a list of certified B-centres can be found at https://www.clarin.eu/content/certified-b-centres. Note that data deposition services are sometimes offered by centres other than "B", and that B-centres may temporarily become C-centres pending re-certification.

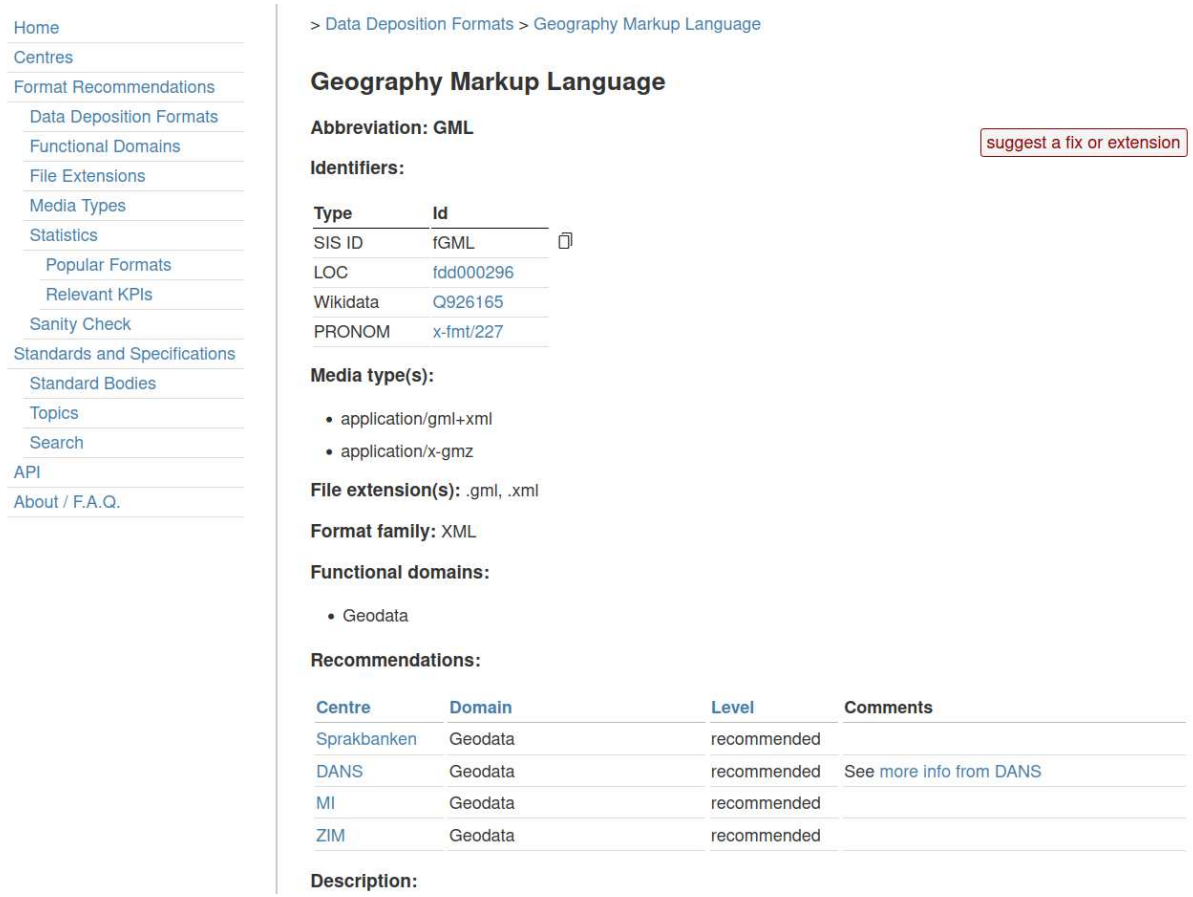

Figure 4: Format description information on GML, with cross-references to the Library of Congress, PRONOM and Wikidata at the top, and other details derived from the system. The description part is suppressed. On the left is the side menu that offers various visualisations of the underlying data.

document that describes the format and is linked from the list of recommendations (see Figure 4 for an example screenshot). In Figure 3, the fragment "PDF/A" is a link, and clicking on it displays the basic information about the PDF/A format, as well as links to related formats. It is also possible that a format does not yet have a corresponding description document in the SIS as is the case of formats mentioned in the recommendations listed in Figure 5. In such cases, the format name is not a link. Instead, it is followed by the $\oplus$ character and clicking on that symbol opens a pre-configured GitHub issue where the basic information on the given format can be provided, so that a physical format information document can be created on that basis. This is a way to ensure that the inventory of formats handled by the SIS can be extended according to the new or modified recommendations formulated by centres, and that the recommendations are not limited to the existing format descriptions.

A recommendation is a qualified pairing of a centre with what the model calls a "relativised format", i.e., a data format viewed from the perspective of the function that the data in that format is expected to fulfil: in the example of the PDF/A format adduced above, the domain for which this format is universally recommended is "Documentation", followed closely by "Textual Source Language Data" – although in the latter case, Figure 3 shows that not all centres are uniform in advocating that format as ideal for "Written unstructured/plain text or originally structured text (e.g. HTML) without linguistic or other mark-up added for research purposes", which is how the SIS defines the "Textual Source Language Data" do-

Figure 5: Fragment of recommendations that do not point to an existing format description document. Clicking on the ⊕ character (note the pop-up) opens a pre-configured GitHub issue.

main[8]. A complete SIS recommendation qualifies a relativised format with a degree of recommendation that the given centre determines, by choosing one of the three recommendation labels: "recommended", "acceptable", and "discouraged". An example of XML encoding of a relativised format is shown in Figure 6, where the submission of data in the domain of "Audiovisual Source Language Data" in the format identified by "fMP3" is discouraged. Additionally, the centre (in the case, IDS Mannheim) provides a comment on the reason for the negative recommendation.

```
<format id="fMP3">
  <domain>Audiovisual Source Language Data</domain>
  <level>discouraged</level>
  <comment>lossy formats should be avoided if possible</comment>
</format>
```

Figure 6: Instantiation of a relativised format with a comment (part of a centre's list of recommendations)

There is a many-to-many association between formats and functional domains: data encoded in a specific format can usually take on many functions, and conversely: a single functional domain is served by many formats. This relationship is never encoded directly – it is derived from recommendations provided by centres. If no centre were to submit a recommendation similar in structure to that in Figure 6, no association between the MP3 format and the "Audiovisual Source Language Data" domain would be derived in the SIS.

3 Standards Information System: data submission and exploitation

The current offer of the SIS towards centres can be summed up in the following three points:

1. increasingly user-oriented way of submitting information,
2. increasingly attractive way to benefit from data aggregation,
3. a way to reuse the data submitted by the centres.

Below, we elaborate on each of these points.

3.1 Data submission

For the purpose of the first data submission, all that is expected from a given centre is a single document that contains a list of formats provided together with a statement that expresses the centre's willingness to accept the particular format in some functional domain. Recall that the possible levels of recommendation are "recommended", "acceptable", and "discouraged", where the last one indicates that the centre may either have insufficient capacity to prepare such data for long-time preservation, or that the process may take a long time. Conversely, the value "recommended" indicates a promise that the deposition process

[8]The list of domains supported by the system is accessible at https://clarin.ids-mannheim.de/standards/views/list-domains.xq and the list of supported formats is at https://clarin.ids-mannheim.de/standards/views/list-formats.xq.

should be relatively painless to both parties. Submitting small-sized initial lists meant to be iteratively extended with further domains or further recommendations is also an option. The SIS offers templates that can be used for that purpose.

Note that, for many centres, the members of the Standards Committee have entered the initial recommendations on the basis of documents published by those centres. That step required a lot of interpretation on the part of the submitter, in order to adjust various kinds of the original recommendations or their varying granularity to the format used by the SIS. Such recommendations are considered "seeds" and should be reviewed, and – probably in many cases – corrected and extended by the given centre. The users are warned in such cases that they are looking at recommendations that have not been curated yet. That warning is eliminated after the centre submits curated information and appoints a contact person.

The preferred way for data submission is by pull requests (PRs) directed at the SIS source code deposited on GitHub. CLARIN developers are familiar with GitHub, so submitting a PR presents no obstacle. For technically less advanced users, the SIS offers an alternative way through editing the recommendation documents, which may be exported from the section of the SIS devoted to the given centre (even if the set of recommendations is empty). These exported files contain placeholders and templates, added in order to make the data input easier. They are additionally constrained by document grammars (W3C XML Schema and ISO Schematron), which signal errors and provide closed lists of options to choose from, where feasible. Finally, many places in the SIS offer an option to switch to editing a templated GitHub feature request, in a single click. This final way is naturally best used only for minor fixes. The wiki system that accompanies the SIS source, linked from the SIS instance, provides additional instructions and illustrations.

3.2 Data aggregation and visualisation

Aggregating structured data from several sources presents an opportunity to visualise the data in various ways and to provide statistics. For this purpose, the SIS offers, among others, word-clouds based on the format keywords, tabular displays of various sorts, extracted lists of file extensions and media types for use in processing pipelines, as well as higher-level statistics concerning, for example, the most "popular" file formats relative to the intended function of the submitted data. For CLARIN, the data aggregated in the SIS make it possible to dynamically compute the Key Performance Indicator (KPI) "collections of standards and mappings", measured by calculating the percentage of centres offering data deposition services and having published their format recommendations (see de Jong et al. (2020) and Bański and Hedeland (2022)) for discussion and further references).

3.3 Information recycling: the SIS API

Finally, the SIS offers a way for the centres to reuse the data that was submitted, via a REST API. This way, the SIS may be used as the sole tool for the maintenance of centre recommendations (and, in the case of CLARIN, to satisfy one of the B-centre certification requirements; see Bański and Hedeland, 2022 for discussion). There is no need to manage two separate instances of data: one for the SIS, and one for the centre itself to display. The API offers a way to receive the information that the centre has provided, to be transformed and styled in the way that the centre wishes.[9]

Additionally, other potentially useful information, e.g. format descriptions, can be obtained via the API and reused. Information obtained from the SIS is available under the CC0 "No Rights Reserved" waiver, with a non-binding request for the SIS to be recognised as the source.[10]

4 Extending the SIS beyond CLARIN

The SIS is in the process of constant development and receives upgrades of functionality on a nearly weekly basis. The most recent work has been influenced by meetings with the Text+ Standardisation

[9] See the example result of an API query for the data of IDS Mannheim at https://clarin.ids-mannheim.de/standards/rest/data/recommendations/IDS-recommendation.xml. The API also supports searching and exporting recommendations with some filtering criteria, such as centre, domain and recommendation level.

[10] See https://creativecommons.org/public-domain/cc0/ for the explanation of how CC0 works.

Group of the Collections cluster, and resulted in partial internationalisation of the underlying functionality: it is now possible to use language tags for centre descriptions and comments on recommendations, and to retrieve that information via the SIS API.

As for the needs of the sibling infrastructure DARIAH, including the cases where the national CLARIN and DARIAH nodes operate as CLARIAH, the SIS offers a functional domain inventory that goes beyond pure language-oriented applications[11]. Depending on the decision by the DARIAH governance (or by the individual repositories) to use the SIS, it remains to be seen whether the repertoire currently offered is going to require further adjustments and fine-tuning given the needs of DARIAH centres.

Figure 7: Switching between research infrastructures. The currently active RI "Text+" is shown on the white background. Hovering over one of the tabs displays a tool-tip to guide the user.

The SIS also provides a functionality that enables users to easily switch between RI environments and to filter the content according to the selected RI. Figure 7 shows how the switch works. "Text+" on the white background indicates that the Text+ environment is currently active. In this case, only Text+ centres and their format recommendations are listed, while the information concerning other centres is hidden. Moreover, language preferences are also taken into account in RI environments. For Text+, which prefers the German language, descriptions and comments are shown in German, as long as centres have provided them. Otherwise, the system falls back to English.

Extending the SIS beyond CLARIN opens new challenges and exposes some limitations of the system. First, some CLARIN centres may appear under different names in research infrastructures other than CLARIN. Currently, the system only allows a single name for a single centre. Whether this is acceptable or whether the centre list needs to be split depending on the RI remains to be seen.

Second, since format recommendations are grouped by the given centre, they are considered to be the same for the same centre across the RIs. When a single centre is a node in multiple RI networks, the SIS assumes that its format recommendations are the same in all these RIs. That means that it would not be possible, for example for the IDS, to recommend the CHAT format in CLARIN but discourage it in Text+. Whether this restriction is going to be problematic remains to be seen when more centres have provided their data.

5 Related work

Similarly to the SIS, re3data.org (Pampel et al., 2013)[12] standing for Registry of Research Data Repositories, provides information about global repositories for deposition of, and access to, research data across various academic fields, and assists researchers in finding a repository suitable to their data and its requirements. The content types of the research data in re3data.org, e.g. audiovisual data and raw data, are more general than those offered by the SIS functional domains, but nevertheless comparable. The SIS is more specific than re3data.org and, naturally, more oriented towards broadly defined HLT research centres – for example, it offers comprehensive details regarding the acceptability of data formats by the listed repositories.

PRONOM (The National Archives, 2002), the Digital Formats website (Library of Congress, 2023) and Wikidata (Wikimedia Foundation and contributors, 2023) present detailed information about file formats including relations to other formats and tools to support the long-term accessibility and preservation

[11] See https://clarin.ids-mannheim.de/standards/views/list-domains.xq
[12] https://www.re3data.org/

of digital materials. Information provided by these initiatives complements the basic information on the particular formats provided by the SIS, and most of the format information documents in the SIS provide cross-references to these three sources (see Figure 4).

In order to promote collaborative knowledge gathering, re3data.org, PRONOM, and Wikidata allow users to submit information through online forms. The SIS targets a much more restricted audience and uses the means made available by the GitHub environment, from pull requests to unstructured issue reports (see Section 3.1), depending on the user's choice and level of technological proficiency.

Similarly to re3data.org and Wikidata, the SIS offers a REST API, as mentioned in Section 3.3. The SIS API is geared more towards the retrieval of entire sets of recommendations, for reuse by centres.

6 Summary and outlook

The Standards Information System is a dynamic platform that adjusts to the expanding demands of data deposition centres. It used to be a relatively static information booth, which around the year 2020 began to evolve into a partially interactive system. The year 2023 is another road marker on its path, as the system opens towards research infrastructures other than CLARIN-ERIC, in the hope to become a platform for the aggregation, visualisation, and measurement of data deposited in research initiatives oriented towards the Humanities. In March 2024, the CLARIN Technical Centres Committee decided to encourage centres to submit recommendations to the SIS and actively monitor the overall progress.

At the time of writing, there are 36 CLARIN depositing centres recorded in the SIS. For 22 of them, at least rudimentary format recommendations have been recorded, and one is in the process of adding the data, after which the dynamically calculated KPI "percentage of centres offering repository services that have published an overview of formats that can be processed in their repository" should be at 63.8%.[13]

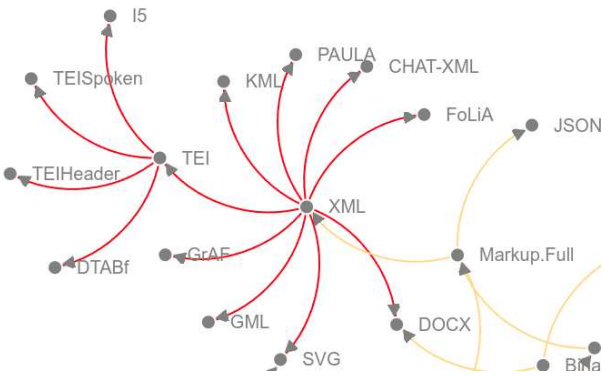

Figure 8: Formal relationships between formats encoded in "format families", a pilot project in the SIS. The figure is a snapshot of a much wider graph. All the leaves and intermediate nodes labelled with format names are clickable and open format description pages similar to that in Figure 4.

Much of the functionality that is needed to serve more than one research infrastructure is already in place. The list of functional domains is already able to accommodate demands that go beyond strictly language-resource-oriented use cases, and can be adjusted to other data functions (with the category "Other" serving as interim storage space). The list of formats is open-ended by design and can be extended both via pull requests and GitHub issues. A preliminary study of formal inter-format relatedness is at the beta stage (see Figure 8) and provides an alternative way to navigate across formats. The system is ready to be used both for CLARIN centres and beyond CLARIN.

Acknowledgments

The SIS has been developed in the context of the work done by the CLARIN Standards and Interoperability Committee (formerly, the CLARIN Standards Committee) and owes much to its former and present

[13] Note that B-centres can sometimes temporarily become C-centres during re-certification. That does not change their classification as data deposition centres, and that is what the KPI calculates.

members, as is only partially evidenced in the CLARIN Bazaar presentations offered in the previous years – a lot of ideas have been discussed, criticised and advanced during the (mostly virtual) committee meetings. We would like to acknowledge the three anonymous CLARIN conference reviewers and thank them for kind words and critical remarks. We are also grateful to the reviewers for the proceedings volume – thanks to their helpful criticism, the text has become much more readable.

Consortia and infrastructures mentioned in the paper

- CLARIAH-DE: https://www.clariah.de/en/
- CLARIN: https://www.clarin.eu/
- CLARIN-D: https://clarin-d.net/en/
- DARIAH: https://www.dariah.eu/
- DARIAH-DE: https://de.dariah.eu/
- NFDI: https://www.nfdi.de/
- Text+: https://text-plus.org/en/

References

Bański, P., & Hedeland, H. (2022). Standards in CLARIN. In D. Fišer & A. Witt (Eds.), *CLARIN: The Infrastructure for Language Resources* (pp. 307–340). De Gruyter. https://doi.org/doi:10.1515/9783110767377-012

de Jong, F., Maegaard, B., Fišer, D., van Uytvanck, D., & Witt, A. (2020, May). Interoperability in an infrastructure enabling multidisciplinary research: The case of CLARIN. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the twelfth language resources and evaluation conference* (pp. 3406–3413). European Language Resources Association. https://aclanthology.org/2020.lrec-1.417

Library of Congress. (2023). Sustainability of Digital Formats. Retrieved March 28, 2024, from https://www.loc.gov/preservation/digital/formats/index.html

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.-J., Gundlach, J., Schirmbacher, P., & Dierolf, U. (2013). Making Research Data Repositories Visible: The re3data.org Registry. *PLOS ONE*, *8*(11), 1–10. https://doi.org/10.1371/journal.pone.0078080

Siegel, E., & Retter, A. (2014). *eXist*. O'Reilly Media, Inc.

Stührenberg, M., Werthmann, A., & Witt, A. (2012). Guidance through the standards jungle for linguistic resources. In *Proceedings of the LREC 2012 workshop on collaborative resource development and delivery* (pp. 9–13).

The National Archives. (2002). *The technical registry PRONOM*. Retrieved March 28, 2024, from https://www.nationalarchives.gov.uk/PRONOM/Default.aspx

Wikimedia Foundation and contributors. (2023). Wikidata. Retrieved March 28, 2024, from https://www.wikidata.org/wiki/Wikidata:Main_Page