Check where centres point for format information #14

Edit   New issue

⊙ Open   **bansp** opened this issue on Mar 30 · 11 comments

bansp commented on Mar 30 • edited ▾   Member

0. Introduction / about this ticket

The lead note of this ticket is part of the release packages of CLARIN recommendations for data-deposition formats, and as such it gets updated at least once per release. The minimal amount of information expected to change cyclically is an update of the date string, after having confirmed that the information is current. Ideally, we are hoping for the centres listed below to gravitate towards section 5.

This page is located at `https://github.com/clarin-eric/standards/issues/14`. Please post remarks, updates and/or corrections in the comments section at the bottom.

CLARIN centres may offer deposition services.
B-centres that offer such services are obligated (this is an (re-)assessment precondition, formulated as part of the CoreTrustSeal requirements) to publish explicit information about data formats that they recommend for depositions. For non-B-centres, this is not a requirement, but it definitely is good practice. That practice has been encoded in one of CLARIN's Key Performance Indicators, using the following measurement: "percentage of centres offering repository services that have published an overview of formats that can be processed in their repository". (Thus the KPI measurement encompasses centres with deposition services, whereas the CTS requirement pertains to B-centres with deposition services, i.e., a subset of the KPI target group.)

The requirement / good practice of publishing explicit information on recommended formats has been so far addressed in the following ways:

1. publishing the information somewhere at the centre (or consortium) homepage;
2. not publishing that information and instead directing users to by now obsolete sets of recommendations that are far too general to represent the given centre's research profile;
3. using a mixture of the above approaches.

The fourth group consists of centres with deposition services that don't publish such information at all, not even as a link. At this point, two centres can arguably be included in that category.

This very ticket is devoted mainly to collecting information on centres with depositing services that point to external sources of information on format recommendations. In other words, we are looking at groups 2 and 3.

The reason for listing this information is:

- to have general info on the "trends" of handling this assessment requirement,
- to identify the specific targets (the targeted resource will often be obsolete; it is good to know whether some centre is pointing there before we e.g. decide to remove it),
- to see if more than one source page is involved, for the given centre,
- to help these centres in identifying and amending this issue (i.e., to define centre-specific data-deposition format recommendations).

Eventually, the appropriate information from each centre will be expected to be collected in the Standards Information System. It is possible for centres to store that information at SIS already at this point, and to present it to users with a dedicated link, such as:

https://standards.clarin.eu/sis/views/recommended-formats-with-search.xq?centre=IMS

(as of early November 2021, the link above leads to no centre-specific information; the IMS neither provides that information on their pages nor points anywhere for it; change the value in the first dropbox to see what info the SIS has about your centre and feel welcome to update it or add it)

1. Centres that point solely to external information

This section lists centres that do not provide information specific to their research profiles but rather point to general and coarse-grained information provided by CLARIN quite a while ago, in most cases in 2009 (the "LRT Standards" document).

1.1. B-centres

Recall that B-centres are obligated (by CTS Requirement 8) to provide explicit information on what formats they are willing to process in the deposition process. The centres below instead point to general and at least partly obsolete guidelines. Amending this situation is at this point easy: deposit the information directly in the SIS.

Centre	LastChecked	LinkTarget	SourcePage
CELR-EKK	13-11-2021	https://www.clarin.eu/content/standards-and-formats	https://metashare.ut.ee/site_media/ms-provider-en.pdf
CLARIN-PL1	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://clarin-pl.eu/dspace/page/faq#what-submissions-do-you-accept
CLARINO B	13-11-2021	https://www.clarin.eu/content/standards-and-formats	https://repo.clarino.uib.no/xmlui/page/faq#what-submissions-do-you-accept
ILC4CLARIN	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/page/faq#what-submissions-do-you-accept
IVDNT	13-11-2021	https://clarin.ids-mannheim.de/standards/ (malformed link)	https://portal.clarin.inl.nl/doc/information_about_deposition_INT.pdf
LINDAT	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://lindat.mff.cuni.cz/repository/xmlui/page/faq?locale-attribute=en#what-submissions-do-you-accept
Språkbanken	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://repo.spraakbanken.gu.se/xmlui/page/faq#what-submissions-do-you-accept (bad link, 13.11.21)

One conclusion that should be drawn from the emerging picture is that the CSC should target the stock FAQ page that many centres use, in order to get the address of the "CLARIN recommendations" changed right there, at the source. The next step will then be to effect an update of that stock FAQ and its instances used by the relevant centres.

1.2. Centres aiming for the "B" status

These centres are not obligated to explicitly publish information about what formats they recommend for deposition. However, some of them already accept depositions, and all are listed as aiming for the "B" status (click on "Type status" to sort them at the top), so at some point they will need to undergo assessment.

Centre	LastChecked	LinkTarget	SourcePage
CEDIFOR	13-11-2021	https://www.clarin.eu/recommendations	https://www.cedifor.de/en/clarin-2/clarin-d/

Centre	LastChecked	LinkTarget	SourcePage
CEDIFOR	13-11-2021	https://www.clarin.eu/recommendations	https://www.cedifor.de/en/clarin-2/clarin-d/cedifor-resource-depositor-guidelines/
CLARIN-LT	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://clarin.vdu.lt/xmlui/page/faq#what-submissions-do-you-accept
ERCC	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://clarin.eurac.edu/repository/xmlui/page/faq#what-submissions-do-you-accept
OTA	13-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://ota.bodleian.ox.ac.uk/repository/xmlui/page/faq#what-submissions-do-you-accept

2. Centres that point to external information in addition to publishing their own explicit information

There is nothing wrong in pointing to an external source apart from the centre's own recommendations, so the centres listed here are not to be considered negatively "flagged" in any sense. When the SIS reaches an appropriate level of maturity, these centres might want to change their external links to point at the SIS.

2.1. B-centres

Centre	LastChecked	LinkTarget	SourcePage
ASV	14-11-2021	https://www.clarin.eu/recommendations	https://repo.clarin.informatik.uni-leipzig.de/public/files/resource_depositor_guidelines_v5.pdf
ASV	14-11-2021	https://www.clarin.eu/content/standards-and-formats	https://repo.clarin.informatik.uni-leipzig.de/public/files/resource_depositor_guidelines_v5.pdf
CLARIN.SI	14-11-2021	http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf	https://www.clarin.si/repository/xmlui/page/data
CLARIN.SI	14-11-2021	https://www.clarin.eu/content/standards-and-formats	https://www.clarin.si/repository/xmlui/page/data

3. Centres that neither point nor publish their own explicit information

This set of centres should be empty, *unless the centre does not offer deposition services*. Please note that rather than amending the existing lack of information on recommended formats, the best course of action for these centres may be to deposit the information directly in the SIS.

3.1. B-centres

"Absence of evidence is not evidence of absence", and it might be that the centres here do publish their own recommendations, in a non-obvious corner of their homepages. Please feel very welcome to post a comment below if you are able to share info on that.

Centre	LastChecked	LinkTarget	SourcePage
IMS	14-11-2021	(none)	https://wiki.ims.uni-stuttgart.de/extern/CLARIN-D
IMS	14-11-2021	"please contact us"	http://clarin04.ims.uni-stuttgart.de/repo/
PORTULAN	14-11-2021	"contact helpdesk"	https://portulanclarin.net/usage/#how

3.2. Centres aiming for the "B" status

These centres are listed as "aiming for B". If they decide to provide deposition services, they might want to do so via the SIS directly.

Centre	LastChecked	LinkTarget	SourcePage
CLARIN-IS	14-11-2021	(no depositions yet)	https://clarin.is/en/services/
NB.NO	14-11-2021	(no depositions)	https://www.nb.no/sprakbanken/en/sprakbanken/ (no info)

4. Other cases

SADiLaR is a centre that aims for the B-status, hence it is not required by the CTS to publish explicit information. It is however within the scope of the KPI measurement, because it offers depositing services.

Centre	LastChecked	LinkTarget	SourcePage
SADiLaR	14-11-2021	stress on original format	https://www.sadilar.org/images/PolicySADiLaRDataAquisitionDistribution1_5LvdB.pdf
SADiLaR	14-11-2021	combination of metadata fields	https://www.sadilar.org/index.php/en/guidelines-standards

5. Centres that link to the SIS, having submitted their recommendations there

This is where all (or most of) the centres listed above should ideally end up -- what is needed for them is to maintain the information served by the SIS and explicitly link to it.

While section 5 is not precisely in the scope of this ticket, it is included (a) to indicate the expected direction of changes and (b) to make sure that when a centre disappears from the listings above, that disappearance is not an editing error.

Centre	LastChecked	SourcePage