



CESSDA ERIC Agenda 21-24, Tasks 21-22

Cross Pillar Activities Task 1: CESSDA Resource Directory

D6 Review of Zotero/Resource Directory as a Platform for the Tools Directory

Document info

Dissemination Level	PU
Due Date of Deliverable	31/03/22
Actual Submission Date	19/04/22
Type	Report, PU
Approval Status	Approved by Working Group Leader Mari Kleemola
Version	v1.0
Number of Pages	p.1 – p.25
DOI	10.5281/zenodo.6901846

The information in this document reflects only the author's views and CESSDA ERIC is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.





Version history

Version	Date	Comment	Revised by
0.1	08.02.2022	First draft	Marijana Glavica
0.2	29.03.2022	Discussions and second draft	All authors listed below
0.3	30.03.2022	Third draft	Marijana Glavica
0.4	01.04.2022	Peer review	Benjamin Peuch
0.5	11.04.2022	Update after peer-review	All authors listed below
0.6	11.04.2022	Fourth draft	Marijana Glavica
0.7	19.04.2022	Approved fourth draft	Benjamin Peuch
0.8	19.04.2022	Ready for submission	Marijana Glavica
1.0	19.04.2022	v1.0 submitted	Christina Bornatici

Author List

Organisation	Name	Contact information
CROSSDA	Marijana Glavica	mglavica@ffzg.hr
SND	Iris Alfredsson	iris.alfredsson@snd.gu.se
CROSSDA	Irena Kranjec	ikranjec@ffzg.hr
FORS	Christina Bornatici	christina.bornatici@fors.unil.ch
ČSDA	Michaela Kudrnáčová	michaela.kudrnacova@soc.cas.cz

Peer-review

Organisation	Name	Contact information
SODHA	Benjamin Peuch	Benjamin.Peuch@arch.be



Contents

Executive Summary	4
Abbreviations and Acronyms	5
Introduction	6
Need for tools directory	6
Definition of the CESSDA tools directory	7
Structure of the document	8
Zotero as a platform for the CESSDA tools directory	8
Organising information on tools in Zotero	8
Adding, updating and curating information on tools in Zotero	9
Discovering and viewing information on tools in Zotero	10
Interoperability	11
Zotero review summary	12
Review of other services with information on tools	13
COPTR - Community-Owned digital Preservation Tool Registry	13
TERESAH - Tools E-Registry for E-Social science, Arts and Humanities	15
TAPoR - Text Analysis Portal for Research	16
bio.tools	17
Social Sciences & Humanities Open Marketplace	18
Possible future developments of the CESSDA tools directory	19
Keeping the tools directory as a part of CESSDA Resource Directory	20
Need to revise metadata structure for tools	21
Need to develop support for integration with other systems	22
Appendix I - Metadata for describing tools mapped to Zotero fields	23



Executive Summary

This report delivers the results of the Subtask 3 of the CESSDA Agenda 21-24, Tasks 21-22, Cross Pillar Activities Task 1: CESSDA Resource Directory. Its main purpose is to review how fit the Zotero/Resource Directory is as a platform for the tools directory, also taking into account outputs from other projects.

CESSDA tools directory (TD) is a structured collection of information about tools (computer programs, software scripts, services, etc.) for social science research data archives, created and/or used by CESSDA Service Providers (SPs). The tools directory should ease the discoverability of specific tools, thus facilitating their use and encouraging the exchange of knowledge and skills.

The tools directory was created as a part of CESSDA Resource Directory (RD), using Zotero Web Library as a platform. Zotero has many limitations when used for organising information resources that are not bibliographic items. Metadata structures are rigid and search capabilities are limited. Useful features of Zotero are collections and subcollections, and tags.

Five existing resource directory services were reviewed in order to discover similarities and differences between resources covered there and in the TD, to explore features that are available for management and discovery of these resources, and to assess if any of the existing services can be used for the needs of the tools directory. None of the reviewed services offer all the desirable features for the needs of the tools directory. Most of them have a different purpose, scope and target audience. Still, two of the reviewed services should somehow be connected with the TD. These are COPTR and SSH Open Marketplace, since some CESSDA tools are already included in these services and others can be added to make them more visible to broader communities.

TD should be kept as a part of the RD because both have a similar purpose and the same audience, so this will enable the discovery of all relevant resources in one place. Also, in this way, there is no need to develop separate policies for tools; they can rather be included in general RD policies, as is now the case. The new CESSDA Web platform can potentially offer some or all desirable features for the tools directory.

The metadata structure should be revised in the next phase of the tools directory development. Also, interoperability features should be developed for the tools directory to enable an exchange of information with other systems.



Abbreviations and Acronyms

API	Application Programming Interface
CESSDA	Consortium of European Social Science Data Archives
COPTR	Community Owned digital Preservation Tool Registry
CROSSDA	Croatian Social Science Data Archive
ČSDA	Czech Social Science Data Archive
DCC	Digital Curation Centre
EOSC	European Open Science Cloud
FORS	Swiss Centre of Expertise in the Social Sciences
GNU	GNU's Not Unix
HaS	Humanities at Scale
MO	Main Office
NeMO	The NeDiMAH Methods Ontology
OAIS	Open Archival Information System
RD	CESSDA Resource Directory
SND	Swedish National Data Service
SODHA	Social Sciences and Digital Humanities Archive
SP	CESSDA Service Provider
SSH	Social Sciences and Humanities
TaDiRAH	Taxonomy of Digital Research Activities in the Humanities
TAPoR	Text Analysis Portal for Research
TD	CESSDA tools directory
TERESAH	Tools E-Registry for E-Social science, Arts and Humanities



Introduction

This report delivers the results of the Subtask 3 of the CESSDA Agenda 21-24, Tasks 21-22, Cross Pillar Activities Task 1: CESSDA Resource Directory. Its main purpose is to review how fit the Zotero/Resource Directory is as a platform for the tools directory, taking also into account outputs from other projects. The aim of this deliverable is to assess if Zotero can support the needs for the tools directory that would serve the CESSDA community, and to explore other existing resource directory services to inform future development of the CESSDA tools directory (TD). This is the second deliverable for this subtask, while the first deliverable (D5)¹ was about collecting information on data archiving and curation tools, services, products, scripts, etc., that CESSDA Service Providers (SPs) have and could share, and populating this information into CESSDA Resource Directory (RD).

Zotero is currently used as a platform for the RD². Information on tools was collected from the SPs in June 2021. By the middle of October 2021, the tools directory was published in Zotero as a part of the RD, together with information that already existed in the previous version of the RD (reported in D5). Consultation with the chief technical officer from the CESSDA Main Office was held in September 2021, to discuss possible integrations of the RD on the CESSDA website, and general issues on the development and sustainability of RD. In February 2022, a discussion forum was organised with 11 representatives from 9 SPs to discuss current features and future developments of the tools directory.

Need for tools directory

Success in providing research data services is highly dependent on the availability of various tools and services, which can support data curation for long-term preservation and facilitate data reuse. There are no integrated data archiving software solutions on the market that can support all activities of every research data archive. Data archives may differ in their mission, organisational characteristics, disciplinary orientation, and the set of services they offer. Social science data archives in the CESSDA community strive to follow a common framework to organise their workflows and procedures, based on the Reference Model for an Open Archival Information System (OAIS). To support data archiving services, some SPs are developing their own tools, while others are relying solely on already available tools.

¹ Not public.

² CESSDA Resource Directory: <https://www.cessda.eu/Tools-Services/For-Service-Providers/Resource-Directory> (accessed 29.03.2022).



For organisations that are developing data archiving services, using an already available set of tools can be crucial to start the service, no matter the budget at their disposal. Not only financial resources but also enough experience is needed to develop new software. Nevertheless, knowing which software tools are used for data archiving services can considerably help new organisations to get acquainted with specific processes that need to be implemented in their organisation. Cooperation is also possible by giving support and sharing knowledge on how to effectively use existing tools.

Further, for more mature data archives, there is a constant need to improve and develop new services, so having information on tools used by other data archives can encourage ideas for cooperation on the possible joint development of tools.

From a formal point of view, sharing data archiving tools is one of the obligations for SPs, according to Annex II, obligation 6 of the CESSDA Statutes. Having a tools directory with information on which tools are used and/or developed by CESSDA community members will considerably facilitate the realisation of this obligation. Thanks to the TD, CESSDA could also monitor the use of the tools among SPs.

The development of a tools directory, as is the case with the whole RD, will also contribute to enabling Annex II obligation 11, which is about providing member support for countries with immature and fragile national infrastructures to help them build up needed competence.

Definition of the CESSDA tools directory

The CESSDA tools directory is a structured collection of information about tools (computer programs, software scripts, services, etc.) for social science research data archives, created and/or used by CESSDA Service Providers. The tools directory should ease the discoverability of specific tools, thus facilitating their use and encouraging the exchange of knowledge and skills. It is part of the CESSDA Resource Directory and is aimed at the same audiences. The primary audience includes CESSDA SPs and partners, but the directory can also be useful to data professionals from other, similar organisations around the world.

The tools directory includes information on tools for supporting core data archiving processes: pre-ingest (acquisition), ingest (curation), access (dissemination), and preservation, together with tools for all other related activities such as communication and user support, collecting performance indicators and other. These tools can be free (open source) or proprietary (available for a fee or as freeware), developed by SPs or other providers.



Information on tools that were collected for the first version of the tools directory is available in Appendix I. The RD collection development policy, curation process and development strategies are the content of upcoming deliverable D1 for this Agenda 21-22 Task.

Structure of the document

The text is organised in three sections. The first part presents the features and limitations of Zotero. Five existing services are reviewed in the second part. The third part is about possible future developments of the TD.

Zotero as a platform for the CESSDA tools directory

Zotero is a bibliographic manager. Its main purpose is to support the collection, organisation, citing and sharing of research information. Zotero supports the creation of group libraries, which enables collaborative maintenance of information. Resources can be managed by using the Zotero free program which can be installed on a personal computer or by using Zotero Web Library, available as a service hosted on Zotero servers. For the purpose of the RD, Web Library is used, in combination with the locally installed program for some editorial tasks.

Organising information on tools in Zotero

Item types supported by Zotero are generally designed to support citation of bibliographic items. Item types are pre-determined, without a possibility to add custom item types. For tools, item type 'Software' was used to separate tools from other resources in RD, but this is not appropriate for all included tools because some can also be services that are built with various software programs and scripts. Nevertheless, item type 'Software' was more appropriate for tools than any other available item type in Zotero.

For each item type, a selection of **metadata fields** are available to describe a specific resource³. All metadata that describes information on tools (see Appendix I) could not be properly matched to Zotero fields. Available metadata fields for Software item type are

³ Zotero Item Types and Fields: https://www.zotero.org/support/kb/item_types_and_fields (accessed 29.03.2022).



appropriate to describe a single piece of software or a script, but not for services, which can be built using several pieces of software. No appropriate fields are available for recording all information on tools in any Zotero item type, not just for the Software item type. For storing additional information, the 'Extra' field is an option. This field was used to store information on tools that can later be recorded in a dictionary for controlled entry and/or for faceted browsing. Such information includes who is using what (User), and who supported the development of a tool (Funder). All other information items about tools relevant for the TD was recorded in the 'Abstract' field as separate sections.

Items can be organised in **collections and subcollections** and by using **tags**. The same item can be included in more than one collection and several tags can be assigned to an item. These are useful features for the tools directory because they enable the main categorisation of tools' functional areas, and additional information on specific functions. Tools were included in the 'Technical infrastructure' category, and for functional areas of tools, subcategories were created following the CoreTrustSeal requirements and the OAIS reference model, i.e. organisation of subcategories for 'Technical infrastructure' was aligned with the organisation of the main category 'Digital object management', with the addition of 'User support & communication' included as a subcategory. Keywords about tools specific functions were translated into tags.

Adding, updating and curating information on tools in Zotero

A **user account** is needed in order to edit information in Zotero group libraries. User permissions for viewing and editing a group library can be set up. The RD is created as a group library, and it is set to be publicly visible (i.e. by all users), but only authorised users can edit information. User roles and permissions in Zotero do not have enough granularity to separate contributors from editors, which would have been a useful feature for maintaining a curated resource directory.

Adding and updating information on resources in Zotero is fairly easy and intuitive if Zotero is used as a bibliographic manager. For that purpose, detailed documentation is available on the Zotero website⁴. A very useful feature for bibliographic resources is the possibility to download metadata from external sources via DOI, URL, ISBN or other identifiers. However, for any other custom use, which calls for additional metadata for describing resources (and such is the case with the TD), Zotero's logic can be confusing.

The main source of information for the TD is input from SPs. Ideally, SPs should be able to add and update information directly to the system. For that scenario, the system needs to

⁴ Zotero documentation: <https://www.zotero.org/support/> (accessed 29.03.2022).



be easy to use, without imposing unnecessary work on contributors. With Zotero, this was not practical because the information needed for the TD did not fit the Zotero metadata structure. Another reason for not letting contributors enter and/or update information directly into Zotero is how limited user permission management support is in Zotero, without the possibility to differentiate contributors from editors. Also, logging of actions is very limited in Zotero, allowing only the date of the last change and the username of the person who created the change to be stored. Since explanations were necessary for each metadata field, a separate collection tool was needed in which these explanations could be added. A collection tool was created using Google Spreadsheets.

For **editorial work**, using a separate collection form (Google Spreadsheet, or any other external form) is not ideal because all collected resources had to be manually copied later into Zotero by the editorial team. This process is hard to track, consumes time and can lead to errors. Available features that can enable editorial work are: duplicate detection, which is available only in the locally installed Zotero application; and automatic recording of information on the person and the date the resource was added or last updated. However, only the last action is recorded, which is not enough for tracking the editorial work.

Discovering and viewing information on tools in Zotero

Metadata **search** capabilities of the Zotero Web Library are limited. It is possible to search only through the following fields: Title, Creator, and Year⁵. These limited search capabilities are far from enough if the user wants to search for any other information, e.g. information on who is using the tool (recorded in Extra field), or other relevant metadata elements (recorded in Abstract field).

Useful features for discovering information are **browsing** by collections and subcollections, and searching and browsing by tags.

The default **view** of the Web Library is a table view, displaying only Title, Creator and Date as columns, and an icon for item type (which is, for the tools directory, always the same, i.e. Software), as shown in Figure 1. **More details** are visible for each item in a separate pane when the user selects an item. Users can add additional columns to the table view using the Column Selector option. Still, only a limited amount of metadata for tools can be shown as columns, and the existence of the Column Selector option is not obvious, so it is not likely that users will notice this option if not instructed.

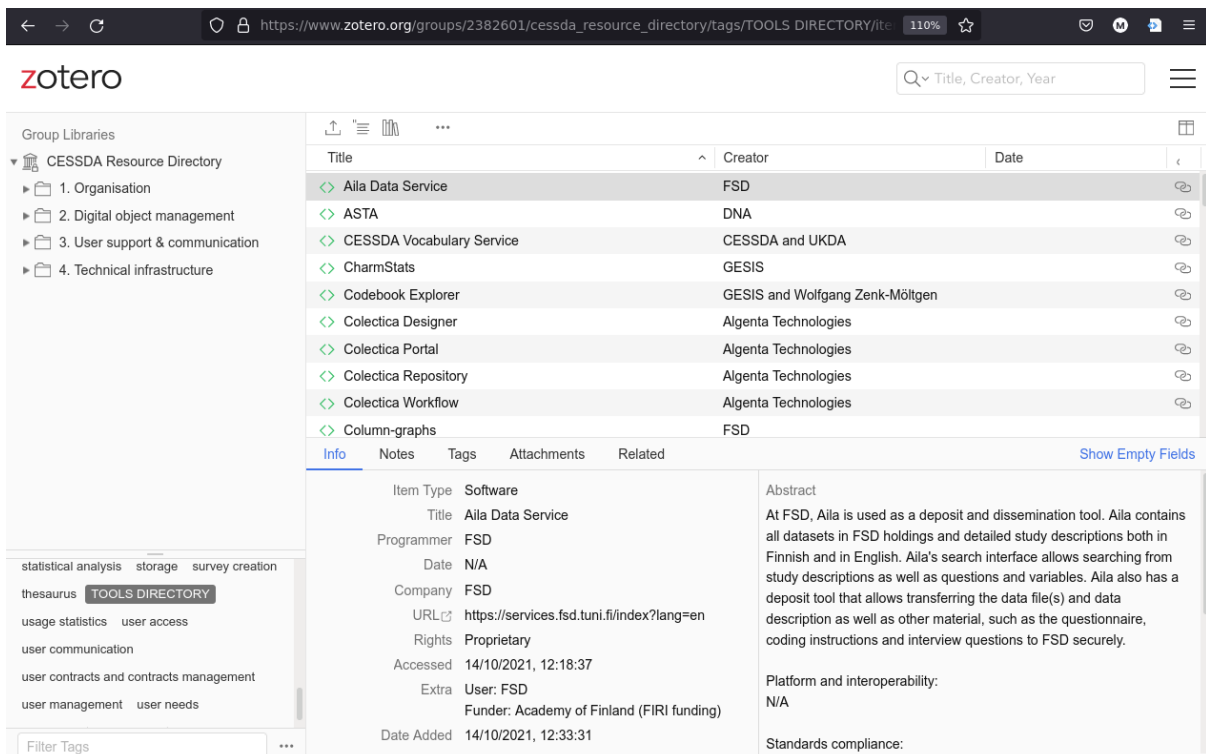
⁵ There is also a full-text search functionality available in Zotero that is looking at the text included in attachments (e.g., pdf of a report). This is not relevant for the RD, since only metadata is collected. Indeed, no files are attached to descriptions of resources which can be indexed for full-text searching.



Sorting by columns is enabled, which is a useful feature if the appropriate columns are shown in the table view. As mentioned before, not all information relevant for tools can be properly mapped to Zotero metadata fields, which are designed for citing literature, so it is not possible to add appropriate fields for tools in columns.

The locally installed Zotero programme has more advanced search capabilities, but it is not likely that the majority of users will use this programme for discovering resources. This could be expected only from users who are already using Zotero for other purposes, and thus are familiar with Zotero features.

Figure 1: Tools directory - part of the CESSDA Resource Directory⁶



Title	Creator	Date
Aila Data Service	FSD	
ASTA	DNA	
CESSDA Vocabulary Service	CESSDA and UKDA	
CharmStats	GESIS	
Codebook Explorer	GESIS and Wolfgang Zenk-Möltgen	
Colectica Designer	Algenta Technologies	
Colectica Portal	Algenta Technologies	
Colectica Repository	Algenta Technologies	
Colectica Workflow	Algenta Technologies	
Column-graphs	FSD	

Field	Value
Item Type	Software
Title	Aila Data Service
Programmer	FSD
Date	N/A
Company	FSD
URL	https://services.fsd.tuni.fi/index?lang=en
Rights	Proprietary
Accessed	14/10/2021, 12:18:37
Extra	User: FSD Funder: Academy of Finland (FIRI funding)
Date Added	14/10/2021, 12:33:31

Abstract
At FSD, Aila is used as a deposit and dissemination tool. Aila contains all datasets in FSD holdings and detailed study descriptions both in Finnish and in English. Aila's search interface allows searching from study descriptions as well as questions and variables. Aila also has a deposit tool that allows transferring the data file(s) and data description as well as other material, such as the questionnaire, coding instructions and interview questions to FSD securely.

Platform and interoperability:
N/A

Standards compliance:

Interoperability

Zotero data can be exported to several formats. Some formats are appropriate to use for citation purposes (e.g. BibTeX, Endnote XML and similar), and some are useful for importing

⁶ Screenshot taken on April 4th, 2022 from https://www.zotero.org/groups/2382601/cessda_resource_directory/tags/TOOLS%20DIRECTORY/items/XZ5Z8PJ/library.



data to other systems (e.g. MODS, CSV). The Zotero programme supports more export formats than the Web Library. Only information on categories can not be exported via the user interface, but it can still be extracted from a locally stored database (SQLite) that is created when the Zotero programme is installed on a personal computer. The Web API enables requests for data from other systems via DOI, URL, ISBN, etc.

Zotero review summary

Zotero is an excellent solution for organising bibliographic references. There are also many useful features of Zotero for building and editing a resource directory, such as: the availability of all common metadata fields for bibliographic references; support for group libraries; the possibility to control who has the right to edit and/or view libraries; possibility to download metadata from external sources via DOI, URL, ISBN, etc.; possibility to organise resources into categories and subcategories; tagging resources; automatically recording date added or last modified, and the user who made a change. All of these features are available on the web interface, i.e. Zotero Web Library. In addition, the locally installed Zotero programme offers even more useful features (e.g. support for duplicate detection and more powerful search capabilities). End users can browse resources through the web interface, using categories, subcategories and tags. Searching for resources is possible by title, creator and date.

Zotero is built on open source technologies and it is free to use. Its implementation didn't require any additional financial resources and it was still an improvement over the Google Sheets used in the first versions of the RD. Furthermore, information entered into Zotero is not locked in, i.e. it can be relatively easily exported and used in other systems. Because of these advantages and lack of resources, Zotero was chosen by CESSDA MO to serve as a platform for the development of RD during the CESSDA Work Plan project in 2019.

However, as information on tools was added to the RD and as RD matured during the year 2021, a need for additional and different functionalities arose. Specific and more flexible metadata are now needed to allow structured descriptions of information that is relevant to answer the RD's (and thus also the TD's) aims, and thus relevant to the CESSDA community. Specific metadata should also be more prominent on a user interface and searchable to improve discoverability of tools and resource items. Finally, since the development of the RD depends on community contributions, a possibility for community members to add information directly to the system would be of great value, together with support for the editorial process enabling editors to review submitted information and track their work for reporting purposes.



Review of other services with information on tools

Several resource directories which contain information on tools (some solely on tools, some together with information on other types of resources) have been developed in the past, as project outputs or community efforts. These are collections of curated resources with a defined scope, audience, policies and metadata structure. Each of these services is available on a dedicated web platform, with a variety of features to support information collection, curation, and discovery of resources.

Five⁷ existing services were reviewed in order to discover similarities and differences between resources covered there and in the TD, to explore features that are available for management and discovery of these resources, and to assess if any of the existing services can be used for the needs of the TD. Each service is described using the following structure:

- **Definition and scope** - short description of the service; what kind of resources/tools are available through the service?
- **Audience** - who are potential users of resources/tools?
- **Policies** - who can contribute resources/tools to the service and how? who approves of the inclusion of a particular resource/tool? who updates and maintains included resources/tools?
- **Metadata** - which metadata is used to describe resources/tools?
- **Features** of a platform (web application) that is used for resource management (contributions, editorial workflows), discovery (searching, browsing) and interoperability (data exchange capabilities).

Based on this information, a judgement was made for each reviewed service about its relevance to the needs of the TD. Interesting features that can be considered for the future development of the TD are pointed out.

COPTR - Community-Owned digital Preservation Tool Registry

<https://coptr.digipres.org>

COPTR helps practitioners find tools needed for long-term digital preservation tasks. It contains information on digital preservation tools and digital preservation workflows. Primary audiences are long-term digital preservation professionals and practitioners.

⁷ Initially, it was planned to include the SSH Training Toolkit in the review, but this was abandoned as this service has a largely different scope (it is about training resources, not about tools) and in the meantime its content was included in the SSH Open Marketplace.



Development of COPTR is not bound to a specific project with limited duration and resources; it is rather a very lively community effort. Anybody can contribute to and/or update a resource in COPTR. As long as a tool has a 'Purpose' (a one-sentence description of its function) and has been categorised by content type and function, the COPTR entry is considered useful. Having a user account is not necessary to contribute, but every edit is logged. COPTR is built as a wiki, so anyone can add information and correct what someone else has entered. A form is available for adding information to COPTR pages in a structured way, together with community guidelines on how to get involved.

Resources can be discovered by searching or browsing. The browse function is created as a grid (COPTR Tools Grid⁸), which allows browsing lists of tools that address specific content types and lifecycle stages (based on the DCC's Curation Lifecycle). Browsing is also possible by function and file format.

COPTR uses MediaWiki software and Semantic MediaWiki extension that helps to search, organise, tag, browse, evaluate, and share the wiki's content. Two main channels for accessing data are available: MediaWiki API and XML dumps of the MediaWiki content.

Relevance to CESSDA tools directory

COPTR has a partly similar target audience and scope. That is why the metadata for the current version of the TD (see Appendix I) was inspired by information on tools available in COPTR. Some tools that are used by SPs are already described in COPTR (e.g. Dataverse, Nesstar, DataCite, DSpace, etc.). More can probably be added, but some CESSDA tools might be out of the scope of COPTR. Information on who is using a particular tool can be entered into the COPTR, but only as an unstructured description, so it is not possible to browse this information.

COPTR could be a good data source to connect with the TD. Possible cooperation between CESSDA and COPTR is worth exploring, although the CESSDA community is more inclined to use OAIS terminology than DCC Lifecycle terminology, on which COPTR is based. Nevertheless, DCC Lifecycle is being referenced by CESSDA archives in the context of research data management, and terminology can be mapped between both models⁹.

Another interesting feature of COPTR to consider for the CESSDA community is the possibility to describe workflows.

⁸ COPTR Tools Grid. https://coptr.digipres.org/index.php/Tools_Grid (accessed 29.03.2022).

⁹ Lindar, M., Rudnik, P., Jones, S. & Horton, L. (2020). "You say potato, I say potato" Mapping Digital Preservation and Research Data Management Concepts towards Collective Curation and Preservation Strategies [Conference Pre-print]. *International Journal of Digital Curation*, 15(1). <https://doi.org/10.2218/ijdc.v15i1.728> (accessed 29.03.2022).



TERESAH - Tools E-Registry for E-Social science, Arts and Humanities

<http://teresah.dariah.eu/>

TERESAH is a cross-community tools knowledge registry. It provides an authoritative listing of the software tools currently in use in SSH. It covers tools, services, methodologies and current standards. The primary audiences are researchers in the social sciences and humanities (SSH), research infrastructure builders and users, and IT personnel.

Initial contributors were DASISH project partners. Development was continued within the Humanities at Scale (HaS) project by DARIAH-ERIC and Göttingen State and University Library. No information is provided on the TERESAH website about adding and updating resources, or how they are currently curated.

The HaS metadata application profile¹⁰, which supports the structured description of digital humanities (DH) tools and services, is used to describe resources. Schema.org was used as the main metadata structure for the application profile, in combination with Dublin Core. Vocabularies are in use for several properties, based on existing taxonomies for digital humanities such as the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH) and The NeDiMAH Methods Ontology (NeMO).

Resources can be discovered by searching, and faceted browsing is enabled by several properties. The interface for searching and browsing is clear and easy to use.

TERESAH has a harvesting capability that allows retrieving information on tools and services from third-party websites and ingesting them into the TERESAH registry. A RESTful application programming interface (API) is available¹¹, in which each type of resource has a URI that one can interact with. Export of all tool metadata is supported via RDF. TERESAH resources are now harvested by the SSH Open Marketplace.

TERESAH source code and documentation are available on GitHub¹².

Relevance to CESSDA tools directory

TERESAH has a different target audience and scope. None of the tools that are currently present in TD are included in TERESAH. Information on who is using a particular tool can not be recorded in TERESAH.

¹⁰ HaS metadata application profile: <http://teresah.dariah.eu/about/application-profile> (accessed 29.03.2022).

¹¹ TERESAH API documentation: <http://teresah.dariah.eu/about/api> (accessed 29.03.2022).

¹² TERESAH source code and documentation: <https://github.com/DARIAH-ERIC/TERESAH/> (accessed 29.03.2022).



The interface for faceted browsing in TERESAH is very clear, especially because there is a short explanation below each facet name, so TD might follow this example. Experiences in developing HaS metadata application profile, described in deliverable D8 of the HaS project¹³, can be useful for future improvements and development of the metadata profiles for the TD, although not the same taxonomies are applicable. The ability to retrieve information from third-party services is one of the interesting features to implement for TD.

TAPoR - Text Analysis Portal for Research

<https://tapor.ca>

TAPoR was originally created as a directory only for tools used for text analysis, but it has evolved to include a far greater range of tools including code snippets that can do the work of tools. In 2018 TAPoR absorbed the Digital Research Tools project (DIRT) and now includes tools that work on non-textual data and tools that provide services used by digital humanists.

Users can contribute to TAPoR by rating tools, adding tags, and commenting on the tools (on how they use the tool, reporting bugs, asking for features or linking to related materials). Users can also create lists of tools that they use together or that complement each other. A webform is available for adding tools, but there is no documentation available that can guide users in that process. Contacts are provided for users who would like to add a tool or write a longer review. No editorial policy is currently published on the TAPoR website, but TAPoR editors are looking for associate editors who would take on the management of an annotated list around a specific discipline, need, audience, theme or technology.

TAPoR is using custom metadata structure to describe tools. TaDiRAH Goals & Methods are used as tags, to better categorise tools.

Searching is possible through all metadata fields, and the browsing function is implemented, which enables filtering tools by information relevant to target audiences.

There is no information on TAPoR website about interoperability features, but TAPoR is included in the SSH Open Marketplace as one of the sources, so it can be assumed that some way of retrieving information exists.

Relevance to CESSDA tools directory

¹³ Claudia Engelhardt, Claudio Leone, Yoann Moranville (2017). *Distributed Metadata Schema and Demonstrator for Open Humanities Methods*, Göttingen State and University Library, DARIAH, <https://hal.archives-ouvertes.fr/hal-01637051> (accessed 29.03.2022).



TAPoR has a very different audience and scope. Some tools (e.g. Dataverse, SPSS) used by the CESSDA community might be present in TAPoR, but the majority of CESSDA tools are out of its scope. The metadata structure is also very different from what is needed for TD. Information on who is using a particular tool cannot be recorded in TAPoR, although lists can be used for that kind of information.

Interesting features of TAPoR are the possibility to add comments on tools and to create lists. Comments are useful for facilitating communication between users and creators of tools, describing their own experiences with a tool, and providing any other additional information. Lists are useful to group related tools together and ease the discoverability of tools by additional information not necessarily included in structured metadata.

bio.tools

<https://bio.tools/>

bio.tools provides essential scientific and technical information about software tools, databases and services for bioinformatics and life sciences. It helps researchers from across the spectrum of biological and biomedical science to find, understand, utilise and cite resources they need in their day-to-day work.

bio.tools development is supported by ELIXIR – the European infrastructure for life science information. The Danish ELIXIR node is now providing long-term funding and support for *bio.tools*, and development is done in cooperation with the ELIXIR community, with a defined governance policy¹⁴. Everyone is welcome to add their own or others' resources to *bio.tools*. Extensive user and technical documentation on *bio.tools* is available¹⁵.

bio.tools is built on standards: the EDAM ontology¹⁶ for scientific aspects, and biotoolsSchema¹⁷ for technical aspects. Both are open, community-defined standards to which anybody can contribute. Creating content domains or "slices" relevant to a specific ELIXIR subcommunity is enabled in *bio.tools* to facilitate description and discovery of resources that are used by different scientific communities.

A Web API is available to provide an easy way to access *bio.tools* data. The source code is freely available to all under a GNU General Public License v3.0.

Relevance to CESSDA tools directory

¹⁴ *bio.tools* governance: <https://biotools.readthedocs.io/en/latest/governance.html> (accessed 29.03.2022).

¹⁵ *bio.tools* documentation: <https://biotools.readthedocs.io/en/latest/index.html> (accessed 29.03.2022).

¹⁶ EDAM ontology: <https://github.com/edamontology/edamontology/> (accessed 29.03.2022).

¹⁷ biotoolsSchema: <http://github.com/bio-tools/biotoolsschema> (accessed 29.03.2022).



bio.tools has a very different audience and scope. All CESSDA tools are out of that scope. Nevertheless, *bio.tools* is an excellent example of a tools directory (or registry) developed by another European research infrastructure consortium. It combines a well-established governance model with community-building mechanisms, supported by extensive documentation and advanced technical features. If CESSDA would consider broadening the scope and audience of the TD, to include not just tools for data archiving and curation practices, but also tools for social science research and data management practices, *bio.tools* could serve as a model for building such a service.

Social Sciences & Humanities Open Marketplace

<https://marketplace.sshopencloud.eu/>

The SSH Open Marketplace is a discovery portal that pools and contextualises resources for social sciences and humanities research communities. It includes tools and services, training materials, datasets, publications and workflows. The target audience is SSH researchers and the support staff of research organisations.

The curation process relies on three components: automatic ingest and update of data sources; continuous curation of the information by the editorial team and contributions from users. Several extensive guidelines for contributors are available: reporting an issue, creating an individual item, enriching an individual item, moderator guidelines, administrator guidelines, and metadata guidelines. Features that allow contributions, feedback and comments are implemented – or will be soon – to ensure that the portal mirrors real research practices.

The data model of the SSH Open Marketplace was designed to be generic and flexible enough to support a variety of sources. During the development phase of the SSH Open Marketplace, data from 15 trusted sources were ingested into the system¹⁸.

The main functionalities of the platform are: searching across resources; refining queries through faceted search; browsing contents by activities or keywords; obtaining a detailed view of every item registered in the website, including a structured set of metadata describing the resource and highlighting related items to facilitate the discovery of relevant resources; contributing to the Marketplace by suggesting new content or enriching existing items; re-using SSH Open Marketplace contents using the application programming interface (API).

¹⁸ SSH Open Marketplace - About the data population: <https://marketplace.sshopencloud.eu/about/data> (accessed 29.03.2022).



Login via EOSC is enabled using existing accounts such as Google, DARIAH, eduTEAMS and multiple academic accounts. The following user roles are available: contributor, moderator and administrator. Each of them has different permissions to work within the system, which enables the creation of editorial workflows.

After the end of the SSHOC project, 3 ERICs and some of their national nodes will ensure the sustainability of the SSH Open Marketplace.

Relevance to CESSDA tools directory

SSH Open Marketplace has a similar scope and audience in part. Although the main focus of the SSH Open Marketplace is on resources for the SSH researchers, support staff for researchers are also targeted as the primary audience. It already includes some tools used and/or developed by the CESSDA community, although some CESSDA tools may not be in the scope. Information on who is using a particular tool can not be entered into the SSH Open Marketplace.

Browsing by activities and keywords is not enough for the needs of the TD. The most relevant features for the TD are: support for adding and enriching content, support for editorial workflow, log in via EOSC, and the possibility to harvest resources from other sources.

Since SSH Open Marketplace already includes several sources that are similar to RD (and TD as a part of RD), the possibility of including RD as a source for the SSH Open Marketplace should be explored.

Possible future developments of the CESSDA tools directory

Based on experiences with Zotero to create the RD and populating it with information on tools, managing collected information, review of other resource directory services, discussions with CESSDA MO technical expert, and input from a discussion forum with representatives of SPs, future developments of the TD are explored in this chapter.

Zotero's main function is to manage bibliographic references. As such, it is not fit to build a resource directory, as it lacks flexibility in metadata structures and has limited features for editing and discovering resources.

None of the existing resource directory or registry services, which were reviewed in this report offer all desirable features for the needs of the TD. This is mainly because they have



a different purpose, scope and target audience. All except the COPTR are targeting researchers, and the scope is disciplinary or topic-specific everywhere except in the SSH Open Marketplace, which covers several SSH fields (although TERESAH might also cover SSH, its content is more relevant to humanities, probably because this is not an active project any more).

Some of these services can be used to record information about CESSDA-relevant tools, but probably none of them can include all relevant tools. TD can contain smaller, very specific pieces of code (e.g. recommended data check syntax files) that might not be relevant to broader communities and thus may be out of the scope of other existing services. Also, there might be a case where some SPs are not ready to share information on tools publicly, but they are ready (or obliged) to share it only with the CESSDA community.

Even if all relevant tools can be described using these relevant platforms, the main purpose of the TD, which is to discover what SPs and partners are developing and using, and to facilitate cooperation between them, cannot be achieved with the current state of existing services. The discovery interfaces of relevant existing services do not allow filtering resources on that basis. This would be possible by implementing additional features to these services, such as the possibility to add comments on tools and to create lists (example from TAPoR), or some other way of exploring related resources (example of communities and domains from *bio.tools*).

Still, two of the reviewed services should somehow be connected with TD. These are COPTR and the SSH Open Marketplace, since some CESSDA tools are already included in these registries and others can be added to make them more visible to broader communities.

On the other hand, the new CESSDA Web platform can potentially offer some or all desirable features for the TD¹⁹. The same features are needed for the whole RD, not just for tools. The discovery of CESSDA training resources is now supported by the new web platform, and the same searching and browsing features are also needed for the RD.

Keeping the tools directory as a part of CESSDA Resource Directory

To sum up, the TD should be kept as a part of RD because both have a similar purpose and the same audience, so this will enable the discovery of all relevant resources in one place. Also, in this way, there is no need to develop separate policies for tools; they can rather be included in general RD policies, as is now the case (RD policy and development strategy is the subject of the upcoming deliverable D1 for the Agenda 21-24, Tasks 21-22). Moreover,

¹⁹ Based on discussion with the chief technical officer of CESSDA MO.



the features of the platform for collecting, curating and discovering tools should be the same. The difference is only in metadata.

Contributions from CESSDA community members are crucial for the success of the whole RD, not just for the TD. Developing support for contributors can be done on two fronts: community and technical. Community work means organising working meetings or so-called "sprints", where a group of people come together (e.g. twice a year) to update information and also discuss other improvements. Dedicated contact points can be appointed in every organisation. Technical aspects can be facilitated by enabling community members to add and update information directly to the platform, thus improving current practices, which consist in collecting information using spreadsheets and then manually transferring them to another system. Since the RD is a curated collection of resources, editors should be able to review and approve or decline the content provided by other contributors. Users should be able to log in directly to the system, ideally by using one of their already existing accounts (e.g. EOSC login, as is the case with the SSH Open Marketplace). Granular user permissions should be available to enable delineation of contributors' and editors' roles.

Searching through all metadata should be possible, with the possibility to refine search results with several facets. Filtering should be possible by the main functional area of the tool and by the organisation that uses a tool.

A list of desired features for the new RD platform is available in the deliverable "D1 CESSDA Resource Directory Policy and Development Strategy" for Agenda 21-24, Tasks 21-22.

Need to revise metadata structure for tools

Metadata for describing tools should be revised to be more relevant for targeted audiences. It would be beneficial and desirable if community members were included in the process of revising metadata structure for the TD, which could be done by organising another discussion forum and similar events that will focus solely on metadata issues for tools. A minimum set of metadata should be defined as mandatory, to ease the contribution effort. Using existing taxonomies should be considered, where possible, to enable interoperability with other services. Differentiation of different types of tools is needed (e.g. computer program for performing a bigger task; a script to perform a specific, smaller task; a service that is built by combining several tools; etc.), and these types should be clearly defined.

Most important information, except for the title and brief description, is the following: who developed a tool and can it be shared/used only within the CESSDA community or also within a broader community, and on which conditions (only for tools and services developed



by SPs and partners); who is using a tool and how (applicable also to tools developed by other providers).

Need to develop support for integration with other systems

Since some tools developed and/or used by the CESSDA community are already included in COPTR and the SSH Open Marketplace, there has to be a way to connect TD with these systems. If a tool is already described somewhere else, users should not be forced to enter the same information again into TD. Only information that cannot be entered in other systems should be sorted in TD.

Different scenarios are possible to achieve this. One is to connect different systems by simple linking, i.e. if a tool is already described somewhere else, only the information that is most relevant for the CESSDA community should be entered into TD, and a link should be provided to the other system. Another scenario is to automatically download available data from other systems into TD at the point of entry.

Another option is to treat RD (which includes TD) as a source of data for the SSH Open Marketplace. This option also depends on the future developments of the SSH Open Marketplace, after the end of the SSHOC project.



Appendix I - Metadata for describing tools mapped to Zotero fields

Information on tools collected from SPs	Explanation	Mapping to Zotero field
<i>For all tools created and/or used by SPs connected to data archiving services</i>		
Title	Title of the tool	Title
Description	Description of the tool. For which purpose it is used?	Abstract
Functional area	One or more keywords that broadly describe the functional area for which the tool is used. <i>e.g. acquisition; ingest; preservation; access; user support; communication</i>	Subcollection under '4. Technical infrastructure': 4.1. Pre-ingest - Acquisition 4.2. Ingest - Curation 4.3. Access - Dissemination 4.4. Preservation 4.5. User support & communication
Function - keyword	Free form keywords that best describe the function for which the tool is used. <i>Some examples: data management; data management planning; transcription; anonymisation; data processing; quality check; redaction; file management; file format identification; file format migration; file recovery; fixity; forensic; metadata extraction; metadata processing; persistent identifiers; storage; metadata server; discovery; visualisation; user management; user communication; requests tracking; etc.</i>	Tags
Creator or provider	Can be an organisation and/or person.	Contributor (for organisations); Programmer (for personal names)
Homepage	URL for more information.	URL
Licence	If the exact name of the licence is not known, open source or proprietary could be entered.	Rights
Note / remark	Any additional information	Abstract



<i>Only for tools that were developed by SPs</i>		
Funder	The name of the organisation or the project who funded the development of the tool (i.e, CESSDA, SERISS, SSHOC, ...).	Extra (with prefix 'Funder: ')
Platform and interoperability	Please enter a short description and specify the programming language, operating systems, databases, plugins, and similar.	Abstract (with prefix 'Platform and interoperability: ') System (for information about the operating system)
Standards compliance	If applicable, please enter which standards are supported and shortly describe how they are implemented. <i>standards examples: DDI (+version), Dublin Core, DataCite, OAIS, PREMIS, METS, RDF, OAI-PMH, OAI-ORE, SWORD, SWORD2, REST, WebDAV, OpenSearch, OpenURL, RSS, ATOM, JSON, XML, etc.</i>	Abstract (with prefix 'Standards compliance: ')
Availability to other organisations	Options to choose from: <ul style="list-style-type: none"> • Already available • Plan to make the tool available in the future • Cannot be shared 	Abstract (with prefix 'Availability to other organisations: ')
Link to source code repository	(if available)	Abstract (with prefix 'Source code: ')
User documentation and/or support availability	Options to choose from: <ul style="list-style-type: none"> • written documentation and support contact available • only written documentation available • only support contact available • not available 	Abstract (URL or e-mail recorded with prefix 'User documentation: ')
Link to user documentation and/or contact e-mail	(if available)	
Technical documentation and/or support availability	Options to choose from: <ul style="list-style-type: none"> • written documentation and support contact available • only written documentation available • only support contact available 	Abstract (URL or e-mail recorded with prefix 'Technical documentation: ')



	<ul style="list-style-type: none">• not available	
Link to technical documentation and/or contact e-mail	(if available)	