From Plans to Impact: Data Management Plans for Humanities Scholars

Jurgita Vaičenonienė
CLARIN-LT national coordinator, SITTI, VMU
jurgita.vaicenoniene@vdu.lt
http://clarin-lt.lt/

CLARIN

Outline

1. The FAIR principles
2. EU and Lithuanian policies in relation to DMPs
3. Data management plans (DMP)
4. DMP tools
5. Humanities-oriented data saving repositories

CLARIN

The FAIR principles

<…> science funders, publishers, and governmental agencies <…> require **data management and stewardship plans** for data generated in publicly funded experiments.

Beyond proper **collection, annotation**, and **archival, data stewardship** includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are **high quality publications** that facilitate and simplify the ongoing process of discovery, evaluation and reuse in downstream studies".

(Wilkinson et al., 2016: 1)

CLARIN

The FAIR principles

To be **Findable**:

- F1. (meta)data are assigned a globally **unique persistent identifier**
- F2. data are described with rich **metadata** (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are **registered** or indexed in a **searchable** resource

(Wilkinson et al., 2016)

CLARIN

The FAIR principles

To be **Accessible**:

- A1. (meta)data are **retrievable** by their identifier using a standardized communication protocol
- A1.1. the protocol is **open**, free, and universally implementable
- A1.2. the protocol allows for an **authentication** and **authorization** procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

(Wilkinson et al., 2016)

CLARIN

Metadata: example

(Source: https://unite-widening.eu/wp-content/uploads/2024/07/DL1-Data-Management-Plan.pdf)

- Document title
- Project Title
- Project Acronym
- Grant Agreement Number
- Project Call
- Funder
- Project starting date
- Duration
- Work Package
- Deliverable
- Deliverable leader
- Deliverable type
- Dissemination level
- Deliverable due date
- Deliverable submission date

CLARIN

The FAIR principles

To be **Interoperable**:

- I1. (meta)data use a formal, accessible, shared and broadly applicable language for knowledge presentation
- I2. (meta)data use **vocabularies** that follow FAIR principles
- I3. (met)data include qualified references to other (meta)data

(Wilkinson et al., 2016)

CLARIN

The FAIR principles

To be **Reusable**:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and **accessible** data usage **license**
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community **standards**

(Wilkinson et al., 2016)

CLARIN

DMP as a means to achieve the FAIRness principles

Data management plans, which "specify an entire life cycle for research data, from creation to interpretation, data documentation and description, short-term storage as well as long-term archiving and data reuse" (*Data Management Plan*, CLARIN D) help researchers improve transparency, reproducibility and compliance with funder expectations and, in the long run, save time, potential of duplicated work, and resources.

CLARIN

EU and Lithuanian policies in relation to DMPs

Consequently, research funders both on the international and on the national level increasingly include requirements for data management (including openness). One of the trend-setters in this regard was the EC, which developed open research data and research data management requirements in its multiannual framework programme for research and innovation, Horizon 2020 (2014 to 2020). <…> A key component is the obligation to create a Data Management Plan (DMP). In recent years, the objective to make data not only open but FAIR (findable, accessible, interoperable and reusable), has been gaining prominence as an important principle for data management and DMPs. This trend will continue in the new Framework Programme Horizon Europe (2021–2027).
(Spichtinger, 2022: 1)

EU and Lithuanian policies in relation to DMPs

New policies and frameworks are emerging at national, European, and international levels to promote the mainstreaming of Open Research. <...> At the European level, the FAIR principles will form a cornerstone of the European Open Science Cloud (EOSC) implementation. <...> the EOSC aims to 'help deliver Europe's contribution to enabling the realisation of scientists', and science's, potential in the digital age' (EOSC, 2021: 11).

CLARIN

Atvirosios prieigos prie mokslo publikacijų ir duomenų gairės. 2016 m. vasario 29 d. nutarimas Nr. VIII-2. (access link)

V SKYRIUS

ATVIROJI PRIEIGA PRIE DUOMENŲ

20. Projekto vadovas turi užtikrinti projekto įgyvendinimo metu gautų duomenų išsaugojimą skaitmeniniu formatu, o pasibaigus projektui – perdavimą saugoti institucijai ir (ar) pateikimą talpyklai. Duomenys turi būti išlaikyti ne trumpiau kaip 5 metus pasibaigus projektui.

21. Projekto, kurio metu bus kaupiami duomenys, vykdytojai paraiškoje pateikia duomenų valdymo planą. Duomenų valdymo planas projekto eigoje gali būti tikslinamas.

22. Duomenų valdymo planui įgyvendinti projekto laikotarpiu patiriamos išlaidos gali būti dengiamos Tarybos finansuojamų projektų lėšomis ir numatomos projekto išlaidų sąmatoje.

23. Duomenys, kurių pagrindu parengtos Gairių IV skyriuje nurodytos mokslo publikacijos, turi būti atvirai prieinami tuo pačiu metu, kaip ir publikacijos. Šie duomenys turi būti prieinami talpyklose ar publikacijų leidėjų nurodytais kitais būdais, ir susieti su atitinkamomis publikacijomis.

24. Projektų vykdytojams, pateikus motyvuotą paaiškinimą, Gairių nuostatos dėl atvirosios prieigos prie duomenų (ar jų dalies) gali būti netaikomos, jeigu:

24.1. duomenys nebuvo gauti projekto įgyvendinimo metu arba mokslo publikacija nėra paremta originaliais duomenimis, t. y. jie nebuvo surinkti ir (ar) sukurti projekto metu;

24.2. reikia apsaugoti rezultatus, ketinant duomenis panaudoti komercinimo ar pramonės vystymo tikslais;

24.3. duomenų atvėrimas nesuderinamas su konfidencialumo reikalavimais;

24.4. duomenų atvėrimas prieštarautų asmens duomenų apsaugos reikalavimams;

24.5. duomenų atvėrimas trukdytų pasiekti projekto tikslus;

24.6. yra kitų teisėtų priežasčių neatverti duomenų.

CLARIN

2024 m. rugpjūčio 30 d. Lietuvos mokslo tarybos nutarimu Nr. V-385 patvirtintas Atvirosios prieigos prie mokslinių tyrimų ir eksperimentinės plėtros rezultatų tvarkos aprašas.
Atvirosios prieigos prie mokslinių tyrimų ir eksperimentinės plėtros rezultatų tvarkos aprašas. 2024. (access link)

18. Mokslo ir studijų institucijos ir MTEP finansuojančios institucijos savo pasirinkta apimtimi koordinuoja mokslinių tyrimų duomenų valdymą ir nustato reikalavimus duomenų valdymo planams. Rekomenduojama į duomenų valdymo planus įtraukti šią informaciją:

18.1. duomenų aprašymą (tipas, formatas, apimtis): duomenų gavimo šaltinis, būdas ir pakartotinio naudojimo galimybės;

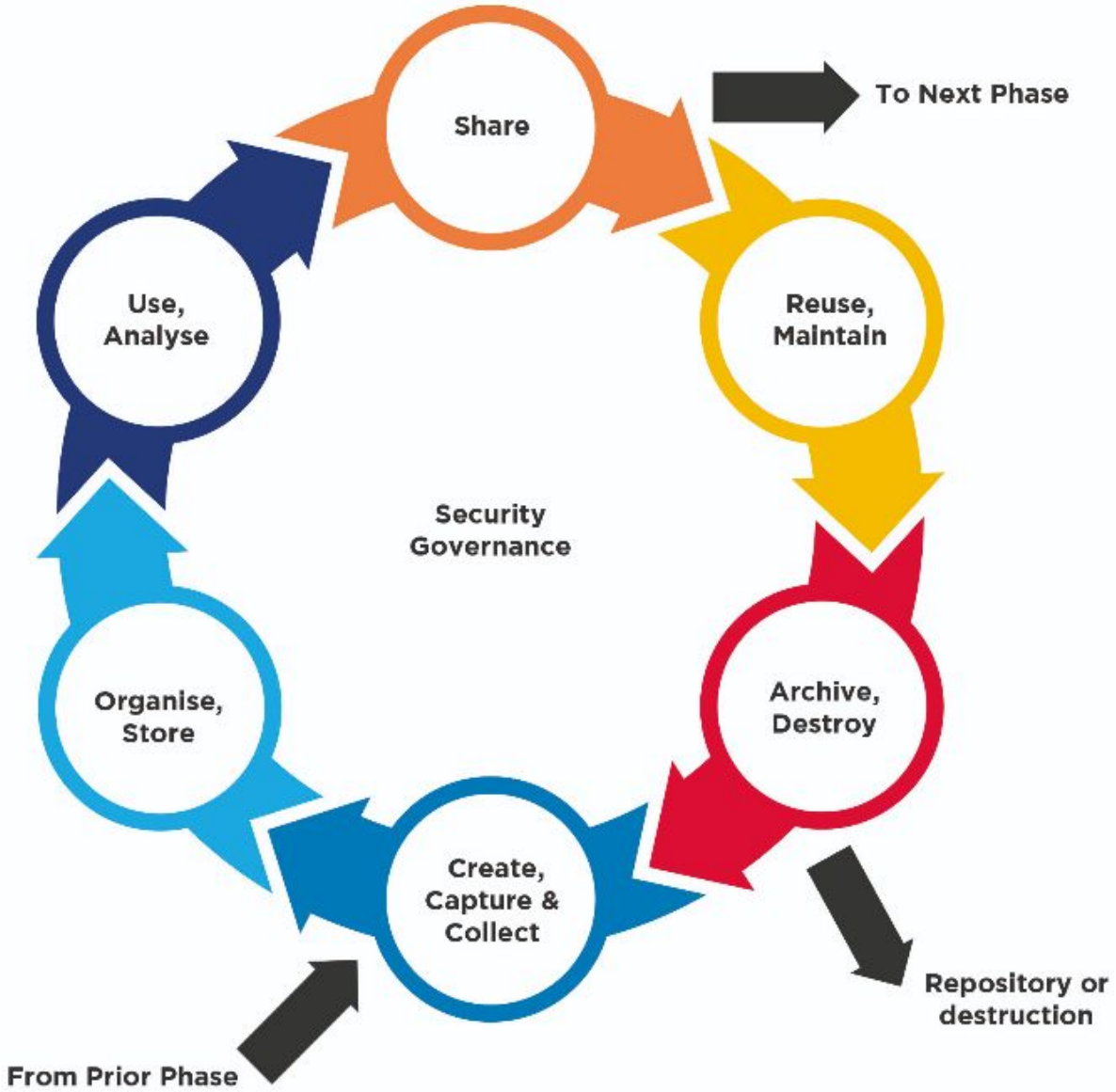
18.2. metaduomenis ir (ar) dokumentaciją (jų gavimo metodai ir struktūra), kuri bus pateikiama su duomenimis ir duomenų kokybės užtikrinimo garantijas;

18.3. duomenų, metaduomenų ir jų kopijų, ypač neskelbtinų duomenų atveju, saugumą užtikrinantį kaupimo būdą;

18.4. duomenims taikytinus teisinius reikalavimus (intelektinė nuosavybė, asmens duomenys), jei tokių yra, ir akademinės etikos ir sąžiningumo užtikrinimo būdus;

18.5. dalinimosi duomenimis sąlygas (kada ir kaip bus dalinamasi, ar yra apribojimų, neleidžiančių dalintis, arba priežasčių embargo laikotarpiui), pakartotinio duomenų naudojimo sąlygas (galimybė parsisiųsti ir modifikuoti duomenis), taip pat ilgalaikio duomenų saugojimo sąlygas ir vietą (kuri duomenų dalis bus ilgalaikio saugojimo).

Data lifecycle



Source: https://data.nsw.gov.au/IDMF/data-management-and-practice/data-management-life-cycle

Benefits of DMPs

Jones (2011: 2) maintains that there are many benefits to managing and sharing your data:

- You can find and understand your data when you need to use it;
- There is continuity if project staff leave or new researchers join;
- You can avoid unnecessary duplication e.g., re-collecting or re-working data;
- The data underlying publications are maintained, allowing for validation of results;
- Data sharing leads to more collaboration and advances research;
- You research is more visible and has greater impact;
- Other researchers can cite your data so you can gain credit.

Benefits of DMPs

- It is an official formal document submitted for funders.
- It shows what will you do with your data during and after your research.
- It ensures that your data is ethical and safely stored.
- Prevents reinventing the bicycle – duplicating sb‘s work.
- Makes your work more visible.
- These procedures increase your planning, save time, and enhance efficiency of your research.

CLARIN

DMP components (Trippel and Zinn, 2015: 2-3)

1. Administrative Data: nature of research project, research questions, purpose of data collection, existing data policies of funder or research institutions;
2. Data Collection: type, format and volume of data; impact on data sharing and long-term access; existing data for re-use; standards and methodologies; quality assurance; data versioning;
3. Documentation and Metadata: information needed for the data to be read and interpreted in the future, details documenting data acquisition; use of metadata standards;
4. Ethics and Legal Compliance: consent of data preservation and sharing; protection of personal data; handling of sensitive data; data ownership, data license;
5. Storage and Backup: redundancy of storage and backup; responsibilities; use of third party facilities; access control; safe data transfer;
6. Selection and Preservation: criteria for data selection, time and effort for data preparation; foreseeable research uses for the data; preservation timeframe; repository location and costs;
7. Data Sharing: identification of potential users, timeframe for making data accessible; use of persistent identifiers; data sharing via repositories and other mechanisms;
8. Responsibilities and resources: for DMP implementation, review and revision at plan and item level, potentially shared across research partners; use of external expertise, costs.

DMP components (Jones, 2011: 30)

1. Data types, formats, standards and capture methods
2. Ethics and intellectual property
3. Access, data sharing and reuse
4. Short-term storage and data management
5. Depositing and long-term preservation
6. Resourcing

DMP components (Engelhardt et al., 2022)

1. Data description and collection or reuse of existing data: Existing data from institutional repositories or digital data collections at the library can be made available for reuse.
2. Documentation/ Metadata and data quality.
3. Storage and backup (during the research project) and data sharing and long-term preservation (at the end of a research project).
4. Legal and ethical requirements.

How to plan your DMPs

- See if the funder offers a DMP template or you are allowed to use any template or tools
- Consult and collaborate with the library, local IT support, legal advisors, data repositories, etc.
- Look for best practice examples (e.g., Zenodo).

CLARIN

Example:

DMP template by Lithuanian Research Council

Skyriai	Klausimai	Paaiškinimai
A. Duomenų rinkimas	1. Kokius duomenis planuojate rinkti ar sukurti?	1.1. Ar egzistuoja duomenys, kuriuos Jūs galėtumėte pakartotinai panaudoti? 1.2. Kokie numatomi duomenų tipai, formatas ir apimtys?
B. Duomenų ir jų atsarginių kopijų kaupimas	2. Kaip duomenys ir jų atsarginės kopijos bus kaupiami projekto metu?	2.1. Kur kaupsite duomenis? 2.2. Kokiu būdu planuojate atkurti duomenis, jei jie būtų pažeisti? Ar kursite duomenų atsargines kopijas?
	3. Kaip užtikrinsite sukauptų duomenų saugumą?	3.1. Kokios gali būti rizikos duomenų saugumui ir kaip šios rizikos bus valdomos? 3.2. Kaip užtikrinsite, kad Jūsų partneris (jei jis bus projekte) galėtų saugiai pasiekti duomenis?
C. Duomenų atranka ir saugojimas	4. Kurie duomenys yra ilgalaikės vertės ir turi būti saugojami?	4.1. Kurie duomenys turi būti saugojami / sunaikinami dėl tam tikrų sutartinių nuostatų, teisinių ar kitų reikalavimų? 4.2. Kokį laikotarpį duomenys bus saugojami?
D. Prieiga prie duomenų	5. Kaip užtikrinsite duomenų žinomumą ir galimybę jais naudotis?	5.1. Kada duomenys taps prieinami? 5.2. Kaip potencialūs naudotojai sužinos apie duomenis, sukauptus jūsų projekto įgyvendinimo metu? 5.3. Kam Jūs suteiksite galimybę naudotis duomenimis ir kokiomis sąlygomis?
E. Atsakomybė ir ištekliai	6. Kam bus priskirta atsakomybė už duomenų tvarkymą / valdymą?	6.1. Kas atsakingas už DVP įgyvendinimą bei šio dokumento periodišką peržiūrą ir koregavimą? 6.2. Ar nuostatos dėl duomenų nuosavybės ir atsakomybės už mokslinių tyrimų duomenų tvarkymą / valdymą bus aptartos su partneriu (jei jis bus projekte)
	7. Kokie žmogiškieji ir kiti ištekliai bus reikalingi rengiant / įgyvendinant DVP?	7.1. Ar reikės įdarbinti specialias kvalifikacijas turintį darbuotoją? 7.2. Ar bus reikalinga speciali, papildoma įranga, įskaitant programinę? 7.3 Ar įvertinote, kad duomenų bankai / saugyklos gali taikyti mokesčius už duomenų kaupimą / saugojimą/atvėrimą?

CLARIN

DMP tools

(Sources: https://dmeg.cessda.eu/Data-Management-Expert-Guide/1.-Plan/European-diversity https://rdm.mpdl.mpg.de/before-research/data-management-plans/)

Argos (OpenAire)

DMPonline (British Digital Curation Centre) (DCC)

DMPTool (University of California Curation Center) (UC3)

Data Stewardship Wizard (Elixir) is one of the recommended tools for DMPs by the European Commission

RDMO for MPG (Max Planck Digital Library)

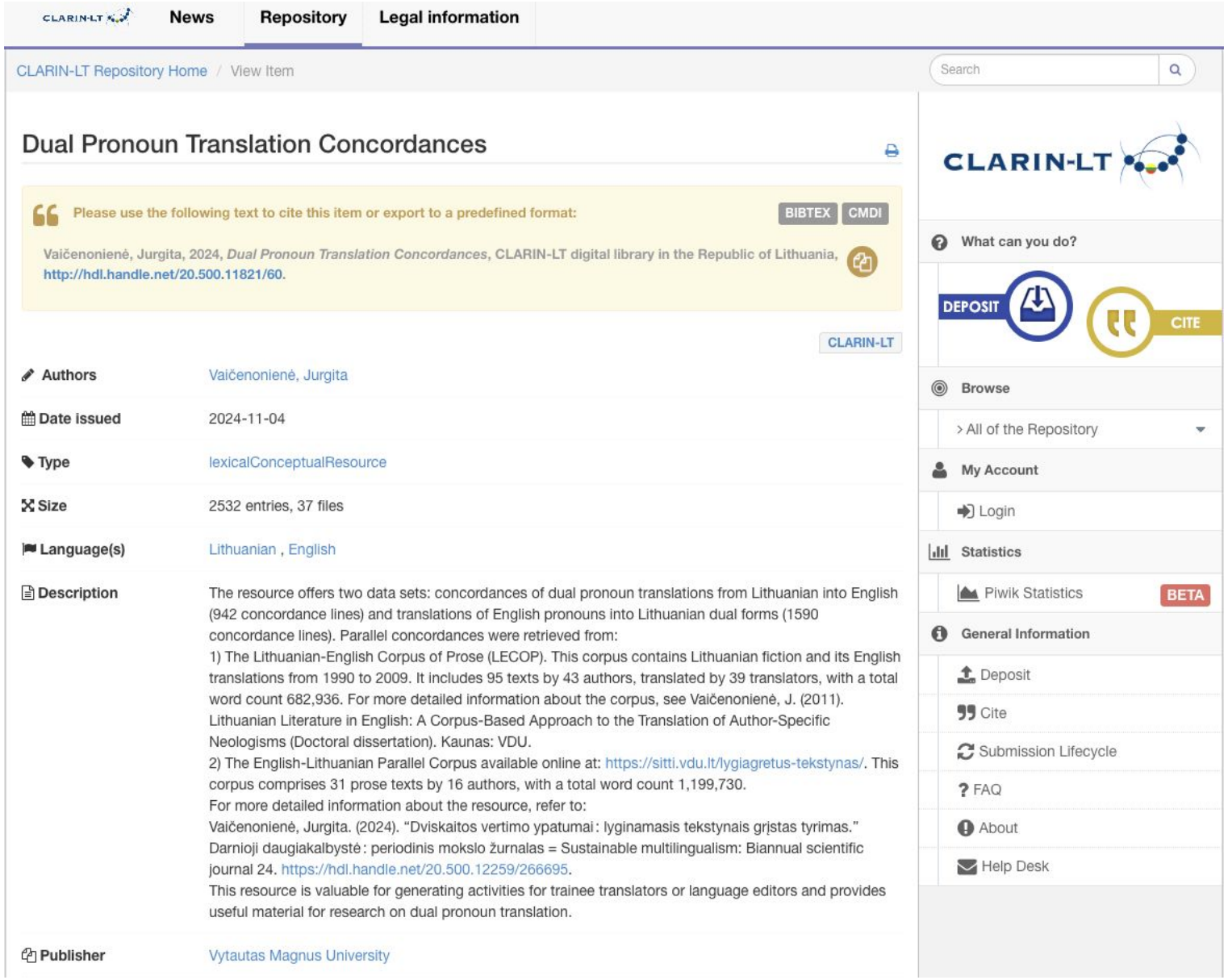
CLARIN

Humanities oriented repositories

- Data repositories can be general (e.g., Zenodo), where researchers of all disciplines can post the data or DMPs, but it is suggested to use, in our case, specifically humanities-oriented research infrastructures which offer domain specific repositories and storage of your data. For example, **CLARIN-LT**, which will be presented in our next presentation.

CLARIN

CLARIN-LT repository – depositing your data according to the FAIR principles



Vaičenonienė, Jurgita, 2024, *Dual Pronoun Translation Concordances*, CLARIN-LT digital library in the Republic of Lithuania, http://hdl.handle.net/20.500.11821/60

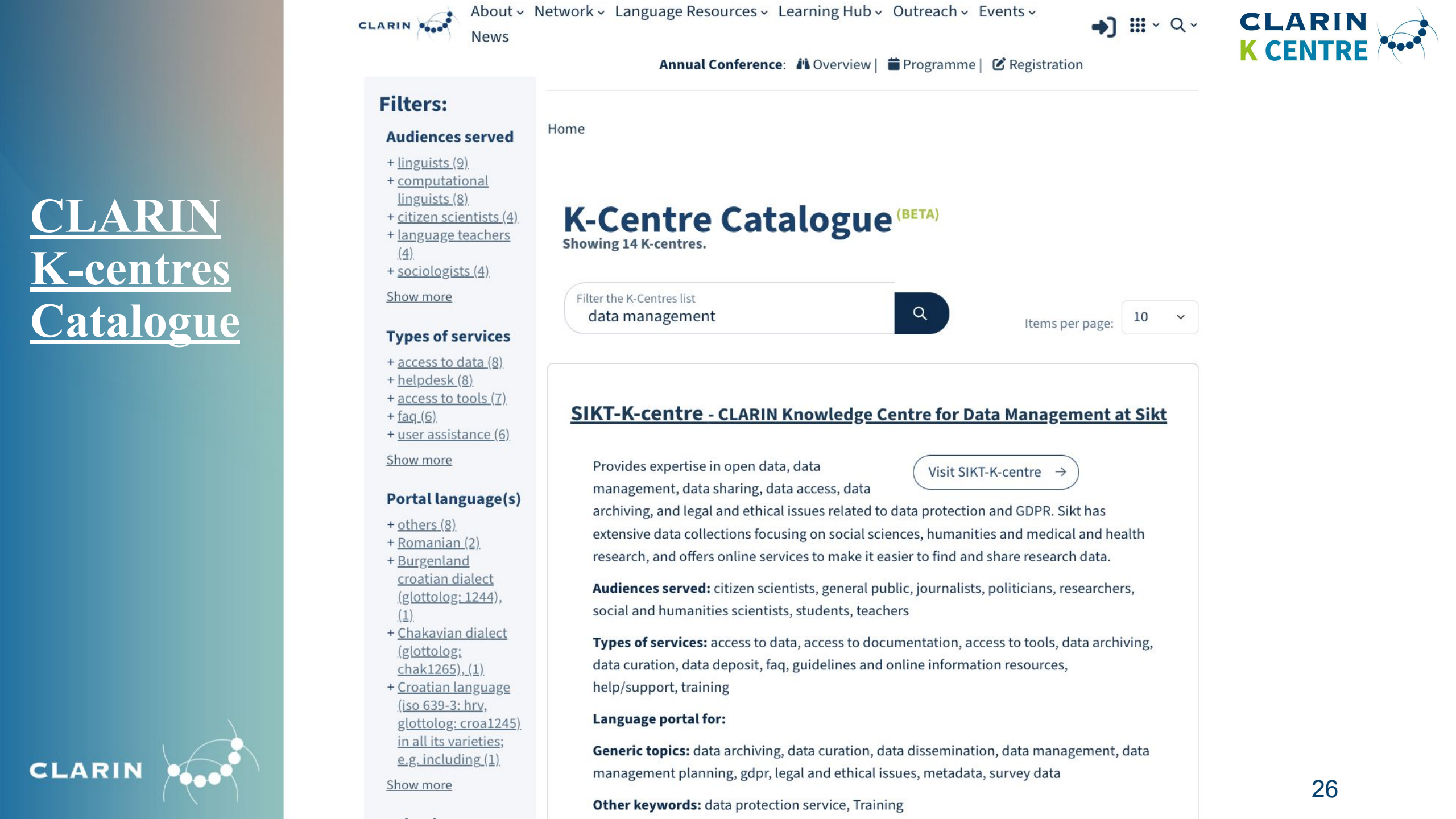
CLARIN Knowledge centres (K-centres)

- CLARIN ERIC, European Research Infrastructure, has several Knowledge centres which consult researchers on their DMPs.
- The K-centres usually reply within two working days.

Examples:

- CLARIN-D
- CLARIN Knowledge Centre for Data Management at NSD

CLARIN K-centres Catalogue



Discussion

- What is your experience with DMPs?
- Did you have to write one when submitting a research proposal or depositing a new dataset?
- What would you like to know about DMPs? Would practical tutorials be useful?
- Do you agree that there is a lack of information about funder requirements and advice on DMP preparation or resource deposition?
- Do you have any best-practice examples that you would like to share?
- Do you think these topics should be integrated into MA study curricula?
- What next steps would you suggest? What could CLARIN-LT do for you?

References

Atvirosios prieigos prie mokslo publikacijų ir duomenų gairės. 2016 m. vasario 29 d. nutarimas Nr. VIII-2. https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/8113c930e0b811e5b18181b790158f61?jfwid=lbwuxeb1i

Atvirosios prieigos prie mokslinių tyrimų ir eksperimentinės plėtros rezultatų tvarkos aprašas. 2024. https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/a1a39ae1696211efad96fef38de4d94a?jfwid=72zogdbfw\

Data Management Plan. CLARIN-D. https://www.clarin-d.net/en/preparation/data-management-plan (accessed 2025 01 12)

Data Management. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Helbig, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y., Kryger Hansen, K., Kuusniemi, M. E., Letizia, V., … Zhou, B. 2022. D7.4 *How to be FAIR with your Data. A Teaching and Training Handbook for Higher Education Institutions (V1.2.1)* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6674301

Jones, S. 2011. *How to Develop a Data Management and Sharing Plan. DCC How-to Guides*. Edinburgh: Digital Curation Centre. http:// www.dcc.ac.uk/resources/how-guides

Pedonese, G., König, A., Wynne, M., and Bushnell, M. (2025, June 29). *Applying FAIR Data Principles in Corpus Linguistics*. Corpus Linguistics (CL2025), Birmingham. Zenodo. https://doi.org/10.5281/zenodo.16407468

Spichtinger, D. Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience [version 2; peer review: 2 approved, 1 approved with reservations]. *Open Res Europe* 2022, 1:42, https://doi.org/10.12688/openreseurope.13342.2

Trippel, T. and Zinn, C. 2015. *DMPTY – A Wizard for Generating Data Management Plans*. Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wroclaw, Poland. Linköping Electronic Conference Proceedings 123:6, p. 71-78.

Wilkinson, M., Dumontier, M., Aalbersberg, I., et al. 2016.The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data 3*, https://doi.org/10.1038/sdata.2016.18

CLARIN