Data registration and Data Monitor pilot at Leiden University

Fieke Schoots | Leiden University Libraries | Slides by Kristina Hettne

UKB wg session on Data registration

28 June 2022

Universiteit Leiden
The Netherlands

CC BY

Why register data?

Section 4 After the research

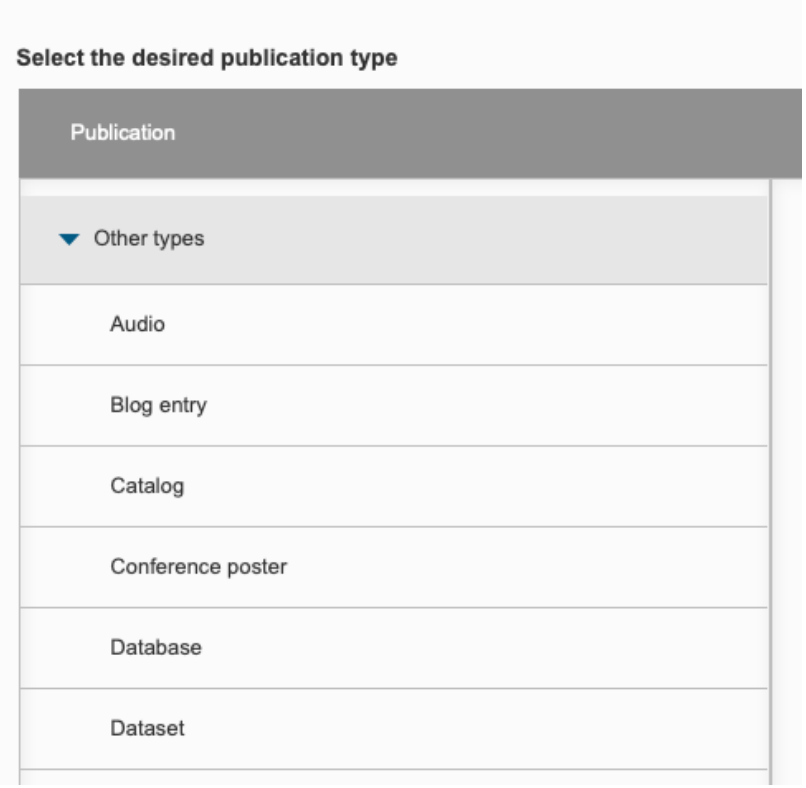
Article 11.

Digital research data that form the basis for a scientific publication are registered at the time of publication and managed according to the FAIR principles.

2021 Leiden Data Management Regulations
https://www.organisatiegids.universiteitleiden.nl/binaries/content/assets/ul2staff/reglementen/onderzoek/research-data-management-regulations-leiden-university.pdf

But…

- Data registration is not yet required nor encouraged
- No clear guidelines provided
- Data sets not visible as such in profiles on web site
- LUCRIS (=Converis) considers data sets as other publications



Example…

Marian Klamer

Directeur PhD-programma's en hoogleraar

Naam	Prof.dr. M.A.F. Klamer
Telefoon	+31 71 527 2783 ›
E-mail	m.a.f.klamer@hum.leidenuniv.nl ›
ORCID iD	0000-0002-2905-7161 ›

Overzicht | Profiel | Contact | **Publicaties** | Nevenwerkzaamheden

Uitgelichte publicaties | Wetenschappelijke publicaties | Boekbesprekingen | Vakpublicaties | Populariserend | Overig

Sorteer op: Categorie | Jaar

Overig

Klamer Marian, Fricke H.L.M., Moro F.R., Saad G. & Visser Eline (2019), Language Collection “Eastern Indonesia and Timor Leste” . [dataset].

overig

However in Narcis…

Leiden Data Monitor project

Kristina Hettne: Project lead

Ben Companjen: Data analysis

Project team Centre for Digital Scholarship: Saskia Woutersen, Laurents Sesink, Fieke Schoots, Ben Companjen

*** * All slides and results presented here created by Kristina Hettne * * ***

Can Elsevier data monitor help to improve registration of data sets?

- Project with Elsevier: the Elsevier Data Monitor
- Main goal: Investigate if the Data Monitor contains the metadata fields needed to connect to LUCRIS for registration
- Method:
  - Elsevier has mined several data sources for data sets connected to Leiden University
  - For some a coupling to ORCID could be made -> match with Faculty and Institute
- Pending decision:
  - Automatic registration or manually by researchers/faculties or a combination?
- Preliminary results in this presentation – feedback is very welcome!

A big thank you to Ben Companjen for facilitating the data analysis!

Preliminary results: Data Monitor metadata fields allow for connection with LUCRIS

Name	Explanation
Dataset DOI	The Digital Object identifier of the dataset
Persistent Identifier	The persistent identifier of the dataset
Repository name	The name of the repository
Dataset title	The title of the dataset
Dataset URL	The URL of the dataset
Description	The description of the dataset
Contributors	The names of the auteurs of the dataset
Publication date	The date of publication of the dataset
Linked	Denotes if there is a scientific article linked to the dataset
Article DOIs	The Digital Object identifier of the scientific article
Journals	The scientific journal
Data Types	The different types of data in the dataset
Institutional Original Contributors	The names of the Leiden authors
Institutional Enriched Contributors	The names and unique identifiers of the Leiden authors

Number of datasets

Year published	Records
2016	239
2017	349
2018	11.261*
2019	554
2020	684
2021	271
Total	**13.358**

*Most are spectra within the same study that have been registered as separate datasets and not as a collection in the repository

Number of datasets with link to a scientific publication (roughly 50%)

Year published	Records with link to publication
2016	158
2017	269
2018	277
2019	347
2020	358
2021	152
Total	**1561**

Top 15 repositories 2016-2021

Leibniz-Institut für Astrophysik Potsdam (AIP)	10852
4TU.Centre for Research Data	698
figshare Academic Research System	665
Strasbourg Astronomical Data Center	318
figshare SAGE Publications	208
The Cambridge Structural Database	156
EASY	99
DRYAD	90
DataverseNL	69
Harvard Dataverse	41
Zenodo	34
Mendeley Data	32
RCSB-PDB	25
Humanities Commons	14

(Meta)data quality

1. Manual (meta)data analysis 5% of records per year
2. Most frequent/significant problems:
    1. Dataset not from Leiden University
    2. Enriched contributor field empty (states detailed author information, like ORCID)

    Only 82 of 668 records had information about LEI author and identifier (Scopus ID or ORCID). This problem is related to the metadata of the data repositories: only 7 of 16 repositories (in this sample of 5 % of records per year) passes on author ID information to Data Monitor.

Elsevier is working on these problems. Most of the datasets that were wrongly accredited to Leiden are now removed. Enriching author information will be done for example by analysing the linked publication.

Next steps

- Evaluate the project results - ongoing
- Reach go or no-go decision for integration of the Elsevier Data monitor with LUCRIS
- Update the Research Data Services catalogue with new repositories
- Questions or ideas? Contact Kristina Hettne <k.m.hettne@library.leidenuniv.nl>

https://digitalscholarship.nl/rds/

Universiteit
Leiden
The Netherlands