Registration of datasets in Pure

Workflow Wageningen University & Research - Library

Hilda van der Pol, 2022 June 28th

Why data registration?

- Ownership: where is our data?
- Monitoring data policy
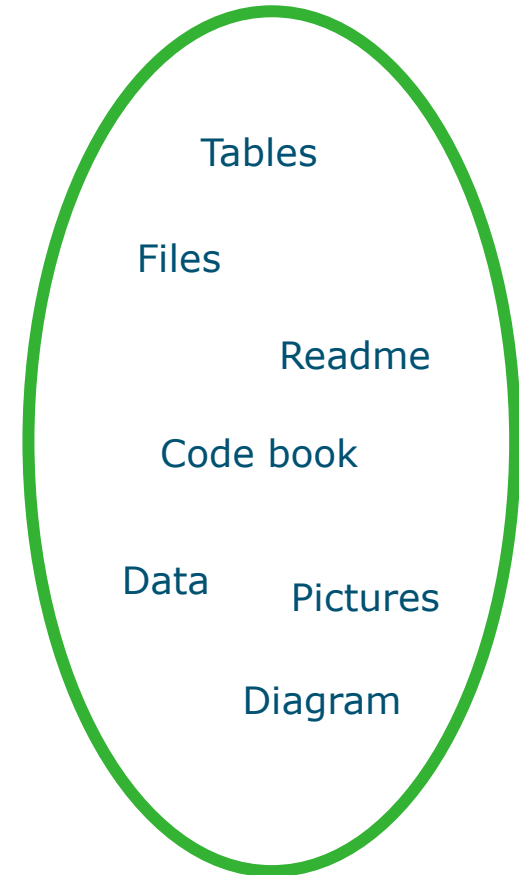- Evaluations and assessments

What to register?

Data

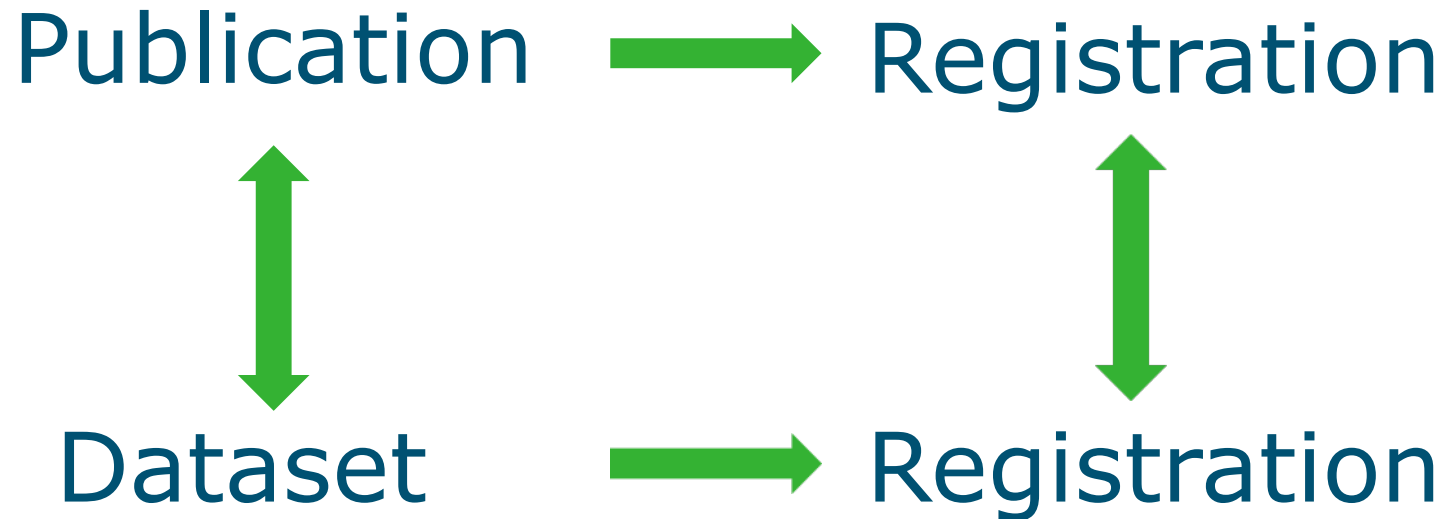- Research data
- No supplementary information

Set

- Package
- DOI, link or identifyer

No quality check

Tables

Files

Readme

Code book

Data

Pictures

Diagram

Dream of the (data) librarian

Publication ➡ Registration

⬍   ⬍

Dataset ➡ Registration

Challenges

- Multiple datasets
- Multiple publications
- No publication
- Versions
- Collections
- Missing metadata
- Datasets that are not datasets
- Datasets without WUR authors
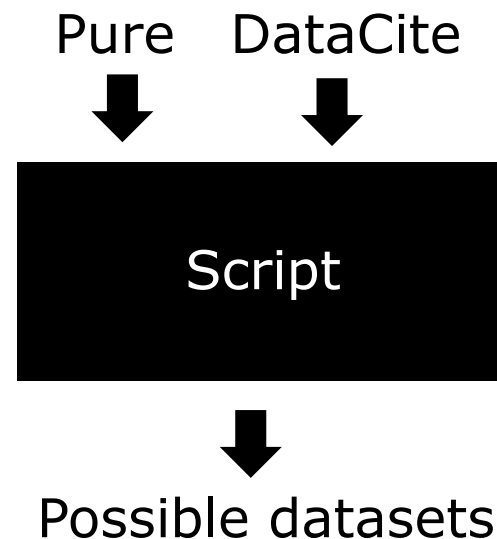- Time span between publication and dataset

How to find datasets?

- Deposited by data librarian

- Reported by researcher

- Data availability statement

- Search engine

- Search in repository

- DataCite

- Elsevier Data Monitor?

Workflow at WUR Library

Find possible datasets

- Get metadata from DataCite (2019-present)
- Find article DOI's in metadata and match with Pure
- Find WUR affiliation in metadata
- Match to article based on title
- Clean up versions
- Clean up collections
- Filter out what's already in Pure
- Filter out earlier lists

Pure   DataCite

Script

Possible datasets

Workflow at WUR Library

Imported in Pure:

- Dataset with publication
- Multiple datasets for one publication

Checked manually:

- Dataset with multiple publications
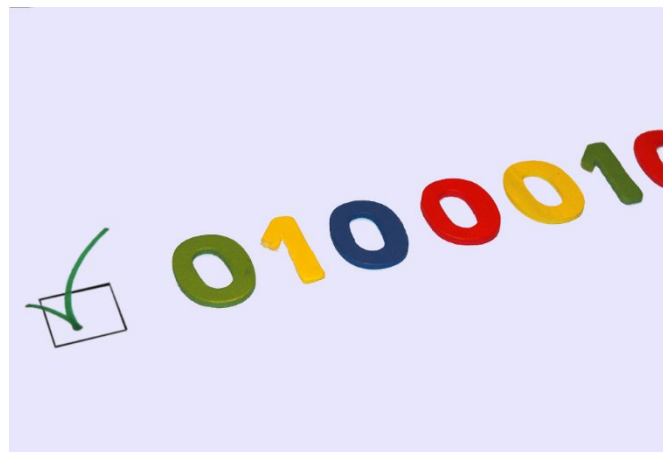- Dataset without publication

Possible datasets

⬇

Script

⬇

Pure

WAGENINGEN
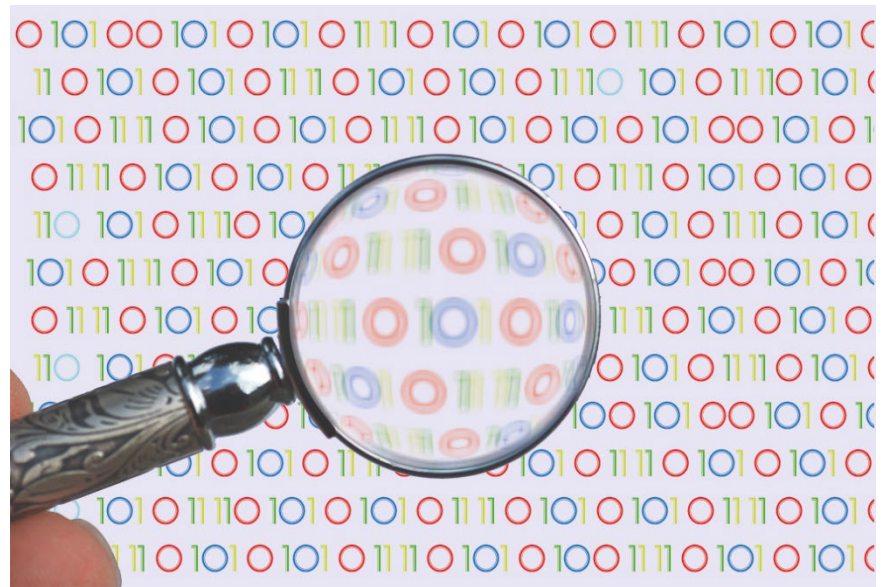UNIVERSITY & RESEARCH

8

Workflow at WUR Library

Validation of records

- Check dataset / supplementary information
- Check extra / other datasets
- Check authors and affiliations
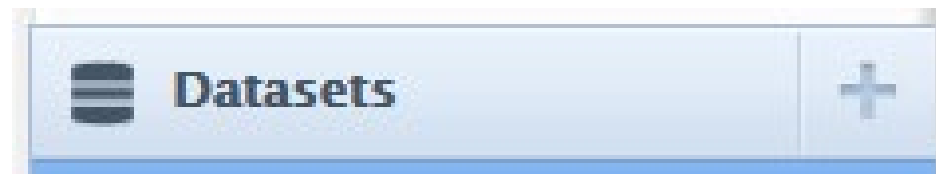- Correct / complement metadata

Genetic datasets

- Gene Expression Omnibus, ENA, Array Express, ...
- Accession number
- Not in DataCite
- Search in repository
- Registered manually
- Text search?
- Import?

Dataset template in Pure

Improvements necessary

- Licence field missing
- Separate fields for publisher and depositing organisation
- Contact person must be one of the creators
- Multiple DOI's not possible
- File formats missing
- Bibliographical field missing

Elsevier Data Monitor

- Import source

- Metadata problems

- Problems with versions and collections

- Pilot after the summer

- Replacement, addition or neither?

- Constant flow

- Costs?

Data Monitor

Import from Data Monitor
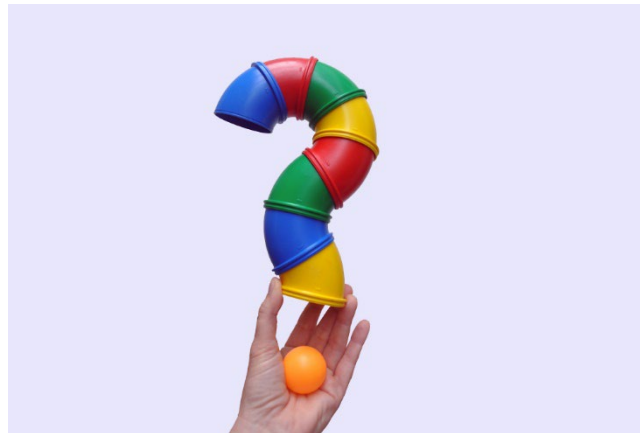
Query

Title

Author name(s)   DOI

The future…

- More regular (now twice a year)
- Datasets without DOI
- Unpublished datasets (W-drive, YoDa)
- Data Monitor?
- Other import sources in Pure?
- Software / source code

Questions?

Hilda van der Pol, hilda.vanderpol-sturm@wur.nl