Summary of Dataset Contextualisation workshops

- Positive reactions and engagement

- United in need for contextualisation descriptions

- Differences in preferred information and presentation of descriptions

- Emphasis on standardisation and machine-readability

- Combining flexibility with standardisation is a challenge for future development

Further work is needed to improve contextualisation descriptions and to answer open questions about their practical implementation

Which fields are needed?

All 66 current fields are essential for at least **ONE** person

Biases, Errors and Access most essential

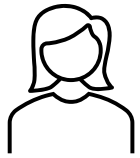
Agreement on only **7** essential fields across workshops

More fields and more detail are wanted

Using the example of data-envelopes

Where does information come from?

Myself 20



Colleague 14

Automatic extraction 6

Don't know 2

Filled in for 42 of the 66 fields

Which vocabularies?

Dataset features

Languages

Dates (e.g. CIDOC)

Topics

Temporal

Genres

Data subjects

Geographical locations (e.g. Geonames)

Domain

Termennetwerk

ELSST

ISIL

Institutional

Others

ORCID

Licenses (e.g. ODRL)

Which export formats?

Formatted text

Markdown

PDF

Linked data

RDF

Tabular

CSV

TSV

Interchange

JSON

"Anything in common use and machine-readable"

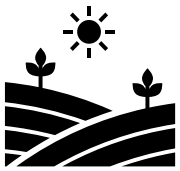
Questions to answer to empower future uptake

'BEFORE': Preliminary questions

- When should a dataset have a contextualisation description?
- Who fills the description in?

'PROCESS': Creating a dataset contextualisation description

- Guidance
  - Clarify fields/structure
  - Add explanatory information/instruction
  - Investigate how to keep the balance between information about contributor and her/his privacy
- Usability
  - Flexible presentation
  - Multiple languages



'AFTER': Role in dataset ecosystem

- Relationship to other documentation
- Integration in own system