



# Australia/New Zealand Data Quality Interest Group

(on behalf of the AU/NZ DQ IG) presented by

Ivana Ivánová

Curtin University

# AU/NZ Data Quality IG

## *the group*



Communities of practice

- forum for AU/NZ data providers, repository operators and data consumers;
- to discuss challenges and strategies for meeting data quality standards and procedures;
- connection to international data quality communities for exchanging experiences and best practices, incl. ESIP's IQC;
- facilitated by Australian Research Data Commons (ARDC) and open to anyone interested in data quality;
- Founded in October 2019 (by ARDC, Curtin University and ANU), meets monthly, currently 34 members;



# AU/NZ Data Quality IG

## *Workshop 6 July 2020 – overview*

- used as a starting point to:
  - help define key concepts in data quality;
  - identify current best practices in Australia/New Zealand;
  - and develop agreed quality frameworks to increase data interoperability and reusability
- 57 registered participants, 27 participated actively
- 4 breakout sessions each following the goals of pre-ESIP Workshop on ‘Developing Community Guidelines for Consistently Curating and Representing Dataset Quality Information’



# AU/NZ Data Quality IG

## *Workshop 6 July 2020 – outcomes*

1. Examine the needs for datasets quality information – *key observations and questions*:
  - How to *define quality* of a federated/aggregated dataset and the individual components that comprise it?
  - What information to provide as part of *metadata* to *assess* dataset's *fitness-for-purpose* by anyone (from any discipline)?
  - What are *the minimum (quality) machine-readable metadata* to make data reusable? And how to generate these automatically?
  - How best to provide information about datasets' *quality constraints* to users? Recording *provenance* appears to be the key.






# AU/NZ Data Quality IG

## Workshop 6 July 2020 – outcomes

- Explore approaches for evaluating dataset quality in an operational environment – **key observations and questions:**
  - *There is a lack of organisational strategy for evaluating dataset quality – some organisations follow best practice from a different discipline;*
  - *How to define organisational units for a strategy? By topic or by organisational roles?*
  - *There is a need/wish for automated procedures for quality evaluation and curation.*

# AU/NZ Data Quality IG

## Workshop 6 July 2020 – outcomes

3. Define the needs and scope of community guidelines – *key observations and questions*:
- Need for guidelines for ‘*common quality language*’;
  - Guidelines on *managing data provenance*, which appeared as the crucial (minimum) quality information to ensure the ‘I’ and ‘R’ in FAIR data;
  - Need for guidelines on *data creation, ownership and curation*, and for *provision of at least minimum metadata* at various level of detail;
  - Need for guidelines to ‘quality label’ data – similar to the 5\*open data model or GEOLabel: 



# AU/NZ Data Quality IG

## *Workshop 6 July 2020 – outcomes*

4. Discuss challenges for consistently curating and representing dataset quality information – *key observations and questions*:
- Problem with extracting *users' requirements for DQ information* to determine datasets' fitness for use;
  - *Missing metadata* longstanding problem – need for guidelines/strategy to ensure metadata are generated and curated in an automated way;
  - Which are the *DQ metrics applicable across disciplines*?
  - Where to *store metadata* – which with data, and which in separate metadata file?



# AU/NZ Data Quality IG

## *Workshop 6 July 2020 – summary*

- Help users answering question ‘should I use this piece of data or not?’ by helping producers to describe their data;
- Request for provenance and high-level metadata information that provide indicators on quality came up often in various sessions;
- People are insisting on online data, quality information or quality evaluation procedures as (web) accessible, machine readable resources more and more;
- Understanding by users (but also producers) of quality information is unclear;
- Understanding of what happens if data quality information is not available?
- Quality information in standards not mandatory – can/should we change this? The group tends to think: **Yes!**



# AU/NZ Data Quality IG

## *Workshop 6 July 2020 – path forward*

- Need for defining common language on the meaning of ‘data quality’ (e.g. the meaning of what constitutes a ‘provenance’ varies);
- Need for educating people on what to look for and where in quality information (e.g. some disciplines are familiar with expression of accuracy stated as RMSE, others are not);
- Need for FAIR community agreed profiles for quality (per discipline);