The changing scope of data quality and fit for purpose: evolution and adaptation of a CRIS solution

15th International Conference on Current Research Information Systems (CRIS2022), Dubrovnik, Croatia, May 2022

Thomas Gurney – Product Manager – Pure
Thomas Gurney

- Product Manager Data Models
- Bibliometrics & Scientometrics
- Knowledge transfer and Science policy (NL – Rathenau Instituut & VU)
Fit for purpose?
How we help

- COLLECT YOUR DATA
- MANAGE YOUR DATA QUALITY
- GROW YOUR COLLABORATIONS
- TELL YOUR RESEARCH STORIES
- OPTIMIZE YOUR PROCESSES
Fit for purpose (my purpose)

• Identify potential funding
• Evaluate research funding
• Avoid duplication of research activity
• Establish priorities for research activities
• Source for scientific publishing
• Inform educational establishments, industry and the general public
• Facilitate information access and exchange
• Facilitate exploitation of research
• Internal administrative functions
• Analyse research trends within and across countries
• Compare institutions within countries in R&D
• Promote international cooperation
• Contacts and networks
• Locate new market for products
• Locate new products for market
• Locate persons/organisations with desired skills
• Administration tool for research projects
Example: data provenance
Data provenance and enrichment
More than deduplication
Data quality management creates both technical challenges and trust problems.

Factors affecting data quality
- diversity of data sources
- depth and relative quality of import sources
- how content is imported and processed
- More (…much, much more)
Framework supporting Data Quality

The 5Cs framework extends beyond deduplication

**Complete** - a record has as many fields filled in as possible

**Correct** - all fields with values have the right value

**Connected** - all records are connected, where possible, to other records

**Current** - all fields with values have the most up to date values

**Compliant** - all fields/records comply with my institution's requirements
Data quality survey

Focus on:
• Context
• Perceptions of data quality
  • Overall
  • By content type
  • By 5Cs
• Sources of errors
• Resource limitations
• Limits of automation

Why?
• Understand institutions’ needs when it comes to data quality
• Prioritise most pressing issues

Survey Oct 2021: n=84, 11 countries
Perceptions of data quality

How important is each of the 5Cs of data quality to your institution?
What are the biggest barriers you face to improving data quality in your Pure?

Other?
- A&H not covered in import sources
- Sense of importance/urgency for faculties
- Interface confusion
- Lack of resources to connect content types
- Hierarchies in import sources
Automation & Trust

How to balance opportunities that automation can address vs opportunities with manual work on import and on current data?

- Available updates
- Match import source IDs for Publications
- Merge duplicate journals
- Person maintenance
- SHERPA/RoMEO synchronization
- Organisation disambiguation

Should data quality be improved by more automation?
Conclusion

CRIS plays a central and significant role in helping institutions achieve their goals and purposes – no matter their context.

To add more value, CRIS must adapt and evolve how they collect, process and manage data, and data quality.

Due to a lack of trust in source data, CRIS can struggle to build trust.

Automation can help, but manual curation is still desired (until trust reaches acceptable levels).
Thank you

Thomas Gurney
t.gurney@elsevier.com