

DATA FAIRNESS PRACTICE

Dr Martin Uhrin

Dr Noel Vizcaino MIET IEEE

SYNOPSIS

- Accessible PIDs: graph structure preserved
- Parsable metadata payload (fetched)
- List of expected good practices
- Use international standards

HIGH LEVEL FAIR SUMMARY ...

FINDABLE

- All IDs are global (eg UUID, hash)
- All IDs are permanent IDs (PIDs)
- Belong to a formal PID scheme
- Self ID: Dataset and Distribution (files)
- Can metadata be retrieved?
- Serialisation parsable?
- It includes core elements for findability e.g. creator
- Optional multiple online sources

ACCESSIBLE

- Access level
- Metadata and data accessible by standardised protocols
- Separate access for data and metadata

INTEROPERABLE

- Using formal knowledge representation language
- Using semantic resources (eg namespaces)
- Metadata and data formally associated

REUSABLE

- Metadata for files? (distribution)
- Licensing info
- Using semantic resources (eg namespaces)
- Attribution/Provenance processable (PROV-O or PAV)
- Per community, metadata standards (list)
- (List) Open standards for files (eg CSV) or long term (ISO/TR 22299) or scientific formats

FAIR ANALYSIS AND SCORING

- Take results/score with a grain of salt!
- QA automation (aiding) should be the goal.
- Accounting only deployed data/scenarios
- Not progressive (FAIR/ER). And cascading failure on any issue.
- Note tools checking PIDs and other items by string methods and/or regexing.
- More semantic resources is not necessarily better. Harcoded Sets.
- Checking the unkown (absurd requirement), scope problems (software engineering).
- Automation is challenging
- Different tools have different aims.
- DCAT or schema.org serialised as any RDF FAIR by design (now much?): issue detectors

SOME FAIR ASSESSMENT TOOLS (SOMEWHAT RELATED)

- F-UJI
- FOOPS!
- O'FAIRE

EXERCISE OPTIONS

- Make a simple data example FAIRer
- Make DCAT dataset in JSON-LD
- Must validate in playground

THE END