Linked (Open) Data for Research Infrastructures

Luc Boruta, Thunken Inc.
luc@thunken.com — @thunkenizer
MERIL Conference, Lisbon, 2019/06/25

Sustainability for RI Metadata Projects

- MERIL is coming to a close
  - Who will use the outputs, and how?
  - Who will maintain the data, and how?
  - Will newer projects contribute back to maintain a core DB?
- How do we link RI data with the rest of the scholarly web?
- Sustainability = FAIRness + Governance + Engagement?

SCHOOL BUS
Yeah Buffy, what are we gonna do now?

FAIR Principles for Digital Assets

- Findable
  - Globally unique and persistent IDs (= PIDs), rich metadata, indexing
- Accessible
  - Open, free, and standardized protocols
- Interoperable
  - Standardized vocabularies, references to other metadata
- Reusable
  - Detailed provenance, clear usage license

https://www.go-fair.org/fair-principles/

Metadata 2020 Principles

- Compatible
  - Provide a guide to content for machines and people
- Complete
  - Reflect the content, components and relationships as published
- Credible
  - Be of clear provenance, trustworthy and accurate
- Curated
  - Reflect updates and new element

http://www.metadata2020.org/blog/2019-05-13-principles-call-for-comment/

European RI Metadata before MERIL

😕 Findable

😕 Accessible

😕 Interoperable

😕 Reusable

😕 Compatible

😕 Complete

😕 Credible

😕 Curated

European RI Metadata with MERIL

- 🤔 Findable
- 🥳 Accessible
- 🤔 Interoperable
- 🤔 Reusable
- 🤔 Compatible
- 🥳 Complete
- 🥳 Credible
- 🤔 Curated

MERIL: Metadata 20/20 Hindsight

- Findability
  - No PIDs for RIs
- Interoperability
  - No PIDs for other fields in the model (organizations, people, works, etc.)
- Reusability/Compatibility
  - The CC BY-NC-ND license is “viral” and (thus) restrictive
- Curation
  - Curation is expensive and tedious

Quick Introduction to Linked (Open) Data

- Semantic web = web of linked data
  - Rather than a web of documents
  - Use URIs (≃ PIDs) as names for things
  - Include links to other URIs so that users can discover more things
- Benefit: query data + draw inferences using vocabularies
- Example: Wikipedia vs. Wikidata

https://www.w3.org/standards/semanticweb/data

Quick Introduction to Linked (Open) Data

Not log

Article Talk Read Edit View

Alan Turing

From Wikipedia, the free encyclopedia

"Turing" redirects here. For other uses, see Turing (disambiguation

Alan Mathison Turing OBE FRS (/ˈtjʊərɪŋ/; 23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist.[6] Turing was highly influential in the development of theoretical computer science, providing a formalisation of the concepts of algorithm and computation with the Turing machine, which can be considered a model of a general-purpose computer.[7][8][9] Turing is widely considered to be the father of theoretical computer science and artificial intelligence.[10] Despite these accomplishments, he was never fully recognised in his home country during his lifetime, due to his homosexuality, which was then a crime in the UK.

During the Second World War, Turing worked for the Government Code and Cypher School (GC&CS) at Bletchley Park, Britain's codebreaking centre that produced Ultra intelligence. For a time he led Hut 8, the section that was responsible for German naval cryptanalysis. Here, he devised a number of techniques for speeding the breaking of German ciphers, including improvements to the pre-war Polish bombe method, an electromechanical machine that could find settings for the Enigma machine.

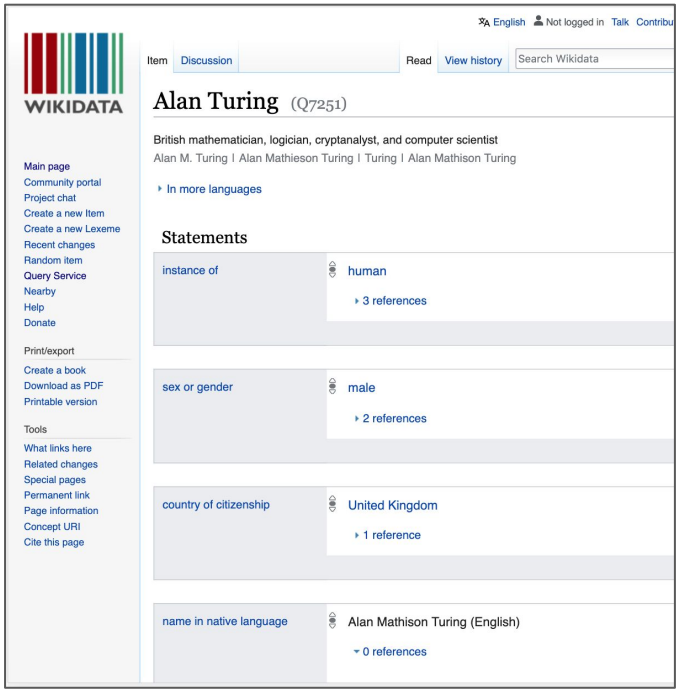
Wikipedia
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

In other projects
Wikimedia Commons
Wikinews



English Not logged in Talk Contribu

Item Discussion Read View history Search Wikidata

WIKIDATA

Alan Turing (Q7251)

British mathematician, logician, cryptanalyst, and computer scientist
Alan M. Turing | Alan Mathieson Turing | Turing | Alan Mathison Turing

In more languages

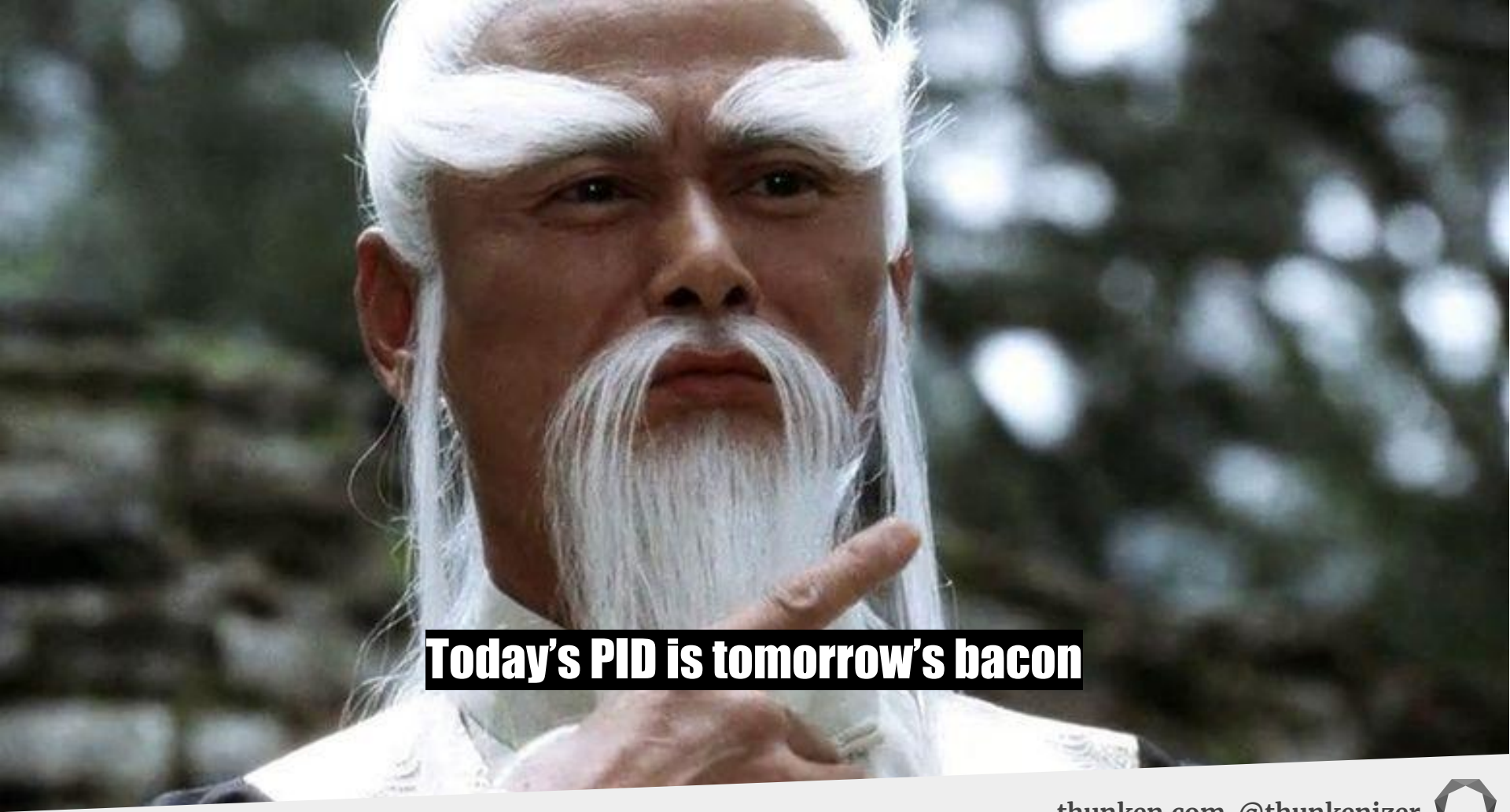
Statements

instance of	human 3 references
sex or gender	male 2 references
country of citizenship	United Kingdom 1 reference
name in native language	Alan Mathison Turing (English) 0 references

Main page
Community portal
Project chat
Create a new Item
Create a new Lexeme
Recent changes
Random item
Query Service
Nearby
Help
Donate

Print/export
Create a book
Download as PDF
Printable version

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page


Today's PID is tomorrow's bacon

PIDs for RIs

- No standard (global, unified) PID for RIs yet
- MERIL: current IDs are not globally uniques nor persistent
- Recommendation: use the Handle System
  - Efficient, extensible, and secure identifier and resolution services
  - Integrations: digital libraries, scholarly web, internet of things, defense systems, grid computing, etc.
  - Backbone of the DOI System (DOIs are Handles)

Knowledge Graphs

http://bit.ly/2rfgwAm

Knowledge Graphs

- Knowledge graph ≈ knowledge base ≈ ontology
- Non specialized graphs, e.g. Wikidata and DBpedia
- Knowledge graphs for the scholarly web
  - **FREYA** (H2020): PIDs only
  - **Cobaltmetrics** (Thunken): PIDs and non persistent IDs
- No PIDs for RIs = no RIs in these graphs
  - Applications: content discovery, scientometrics, publisher analytics, etc.

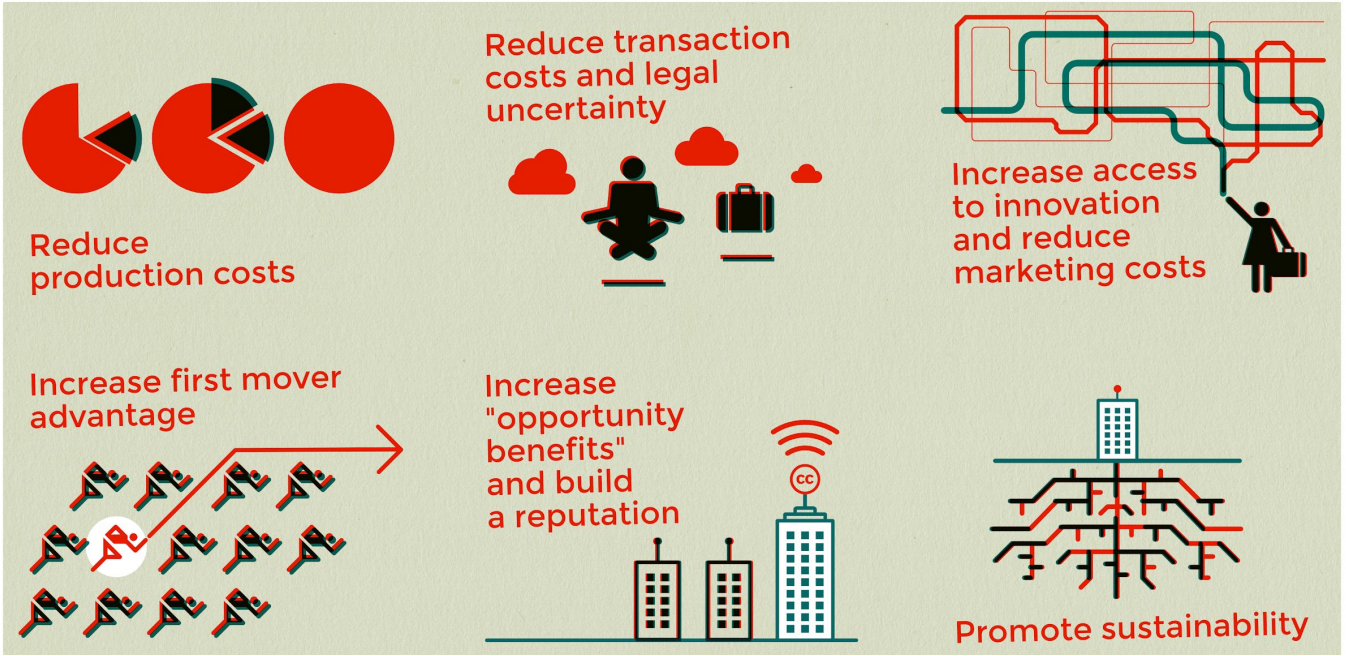
Wishes to the FAIRy Godparent

- Data model and API: PIDs, PIDs everywhere!
  - PIDs for RIs
  - Other PIDs, e.g. DOI, GRID, ORCID
- User interface: use PIDs to autocomplete forms
- License: CC0 or CC BY

Sustainability for RI Metadata: What's Next?

- FAIRness
  - PIDs and a more permissive license
- Governance
  - MERIL as the central registration authority
- Engagement
  - Include private/for-profit RIs?
  - CC Open Business Model

Sustainability: CC Open Business Model



https://drive.google.com/drive/u/0/folders/0B2KL8PtO4xD-bEVGU1Q0Qm8xejA

Findable?
Accessible?
Interoperable?
Reusable?
Compatible?
Complete?
Credible?
Curated?
Yeah, duh.

THUNKEN