The Future of PRONOM
A review
#OPFCon

David Clipsham
10 June 2020
PRONOM – a potted history

- 2001: 1st internal version
- 2005: Version 3 – 1st public version. DROID released!
- 2008: Version 6 – the current version
- 2008-10 PRONOM 7 requirements documents produced
- 2010-11: Linked data prototype
- ...
- ...
- 2019-20: ‘Future of PRONOM’ project
Project background

• Timeboxed project ran December 2019-March 2020

• Modernise the infrastructure – looking specifically at graph data tech
• Refine the data model – Strip out unused data tables, improve extensibility
• Improve data linking/sharing – Including API definition
• Extend the PRONOM signature syntax – incorporate container signatures into PRONOM core

• Continue to produce DROID-compatible signatures
• Make it easy to add/update data
• Cloud-first!
January requirements virtual workshop

• 30 participants from around the world
• Active, lively discussions strong participation, prioritisation voting
• Huge number of requirements gathered!
• Use of Mentimeter tool for voting:
Mentimeter results

Session 1 - Refining the PRONOM Data Model
Session 2 - The PRONOM API: slider

1. Establish hyperlinks to external resources
2. Bulk download of PRONOM XML files
3. RESTful API
4. Standard based (e.g., OData)
5. PAR API
6. Best-practice-ish (SPARQL)
7. Listing of formats
8. Display draft PUIDs with a draft status
Mentimeter results

Session 3 - Other requirements: slider

1. Provide ‘Trusted Partner’ access to update PRONOM.
2. Formal release Schedule
3. Backlog visibility
4. Public facing signature workflow progress etc
5. Tool to infer PRONOM sigs from sample files (like TRiD does)
6. Ease the process of allowing the community to feed back to TNA re PRONOM
7. Do more research weeks
8. Documentation of what a file format is to PRONOM
Mentimeter results

Session 4 - Extend the PRONOM Signature Vocabulary

1. Fully embed container signatures within PRONOM
2. New signature type for XML files
3. Fix for SPSS Portable (fmt/997) file format signature.
4. (Example for 2) Fix for GML (x-fmt/227 and fmt/1047) file format signatures.
5. Multiple bytes from a byte range
6. Match repeats of a byte sequence
7. Match bytes from a non-continuous set
8. Macintosh packages
9. Dynamic & Static filenames
10. Conditional Statements
Prototypes produced – graph data model
Class pronom2020:InternalSignature
Prototypes produced – editorial service
Prototypes produced – report service
Prototypes produced – SPARQL endpoint
What’s next?

• Project itself is currently paused, but...

• We’ll publish the proposed data model for feedback
• The existing PRONOM service continues as normal for now
• We’ll be improving PRONOM research transparency, using GitHub to publish backlog
• Look out for PRONOM Research Week 2020 in October!