

The Future of PRONOM

A review

#OPFCon

David Clipsham

10 June 2020

THE
NATIONAL
ARCHIVES

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

THE

NATIONAL

ARCHIVES

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!

THE

NATIONAL

ARCHIVES

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:

What's your favourite file format?

Mentimeter

A well document one

png

png

FLAC

Pdf

jpeg

Gif

tiff

rom =)

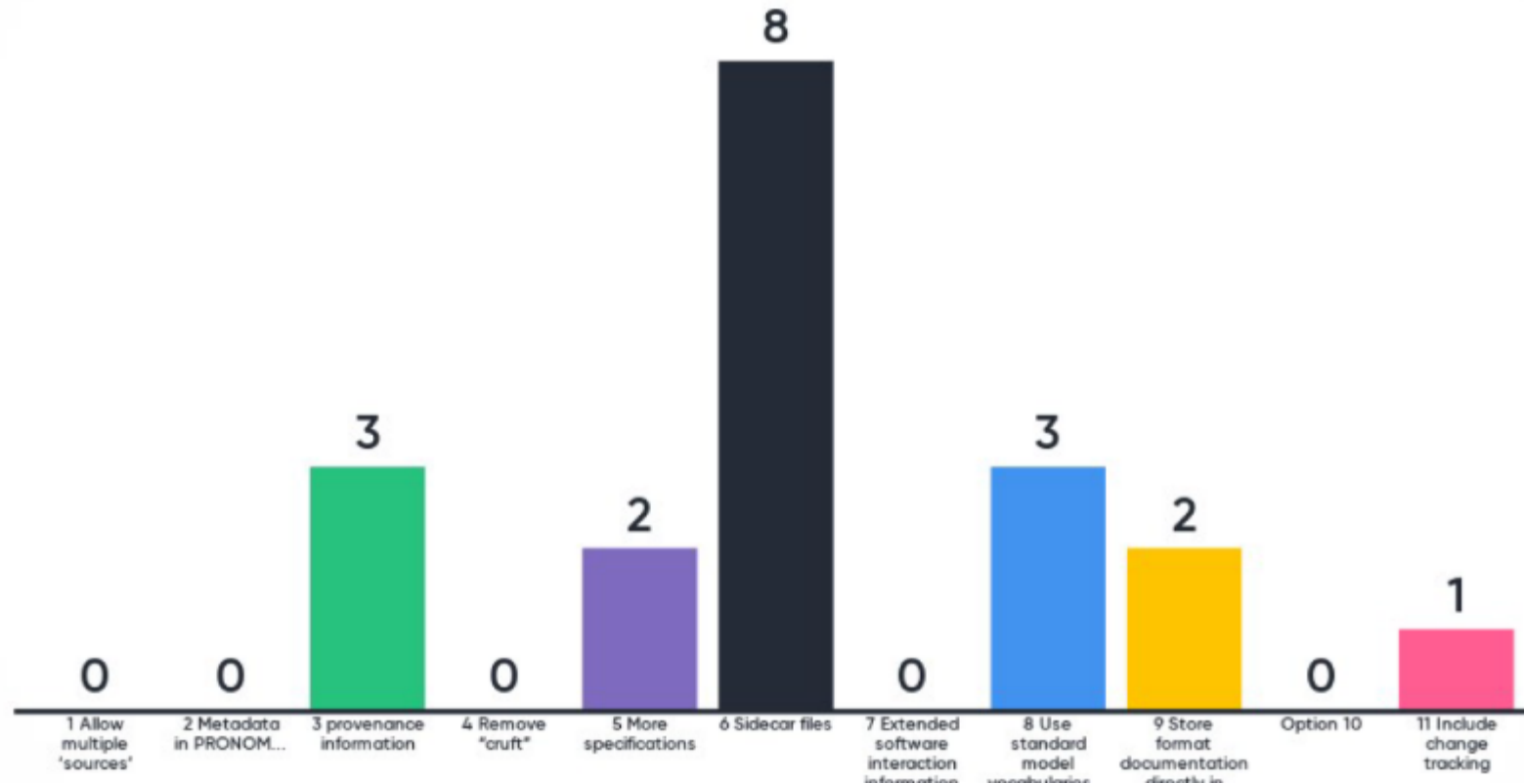
THE

NATIONAL

ARCHIVES

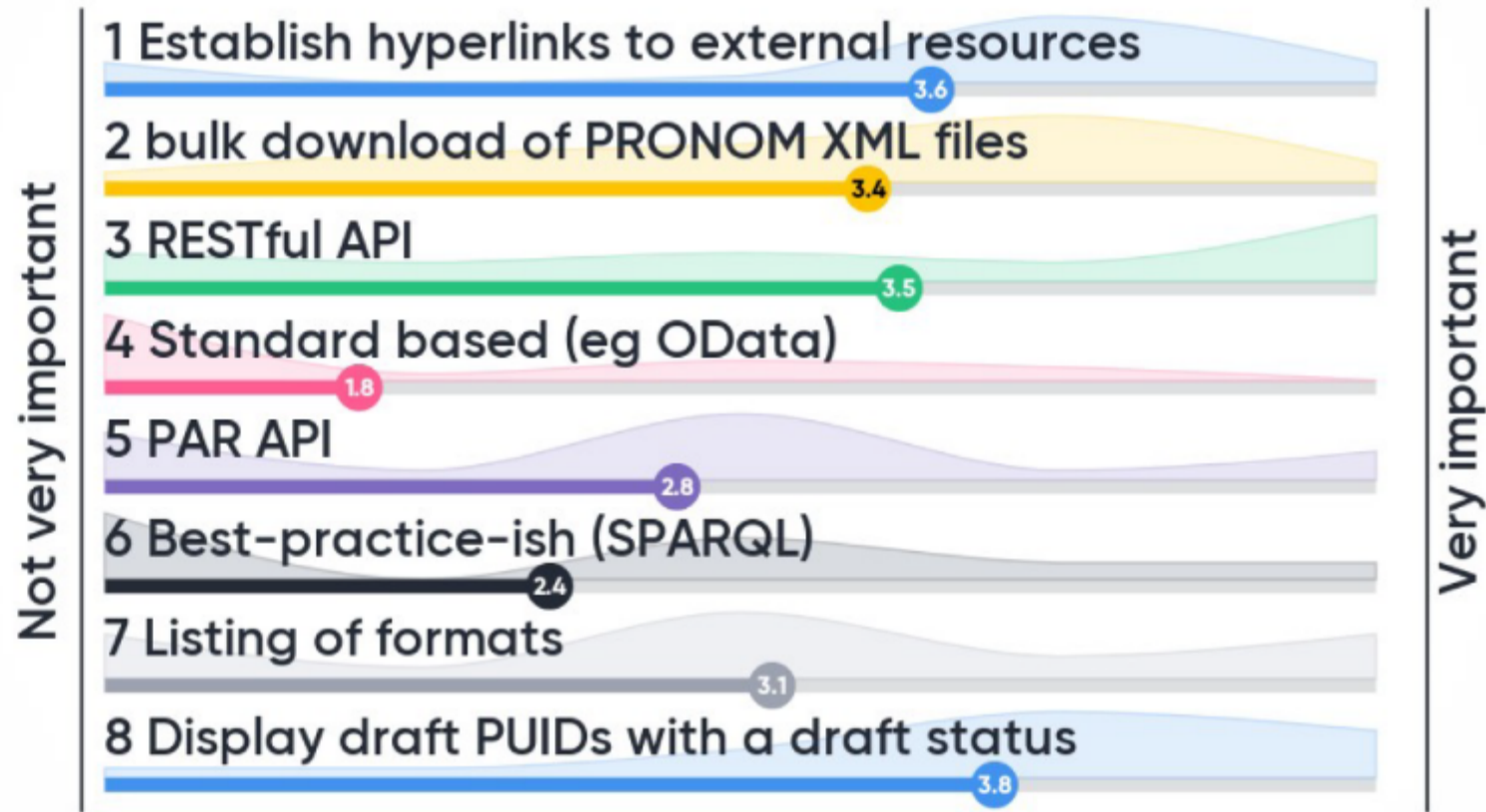
Mentimeter results

Session 1 – Refining the PRONOM Data Model



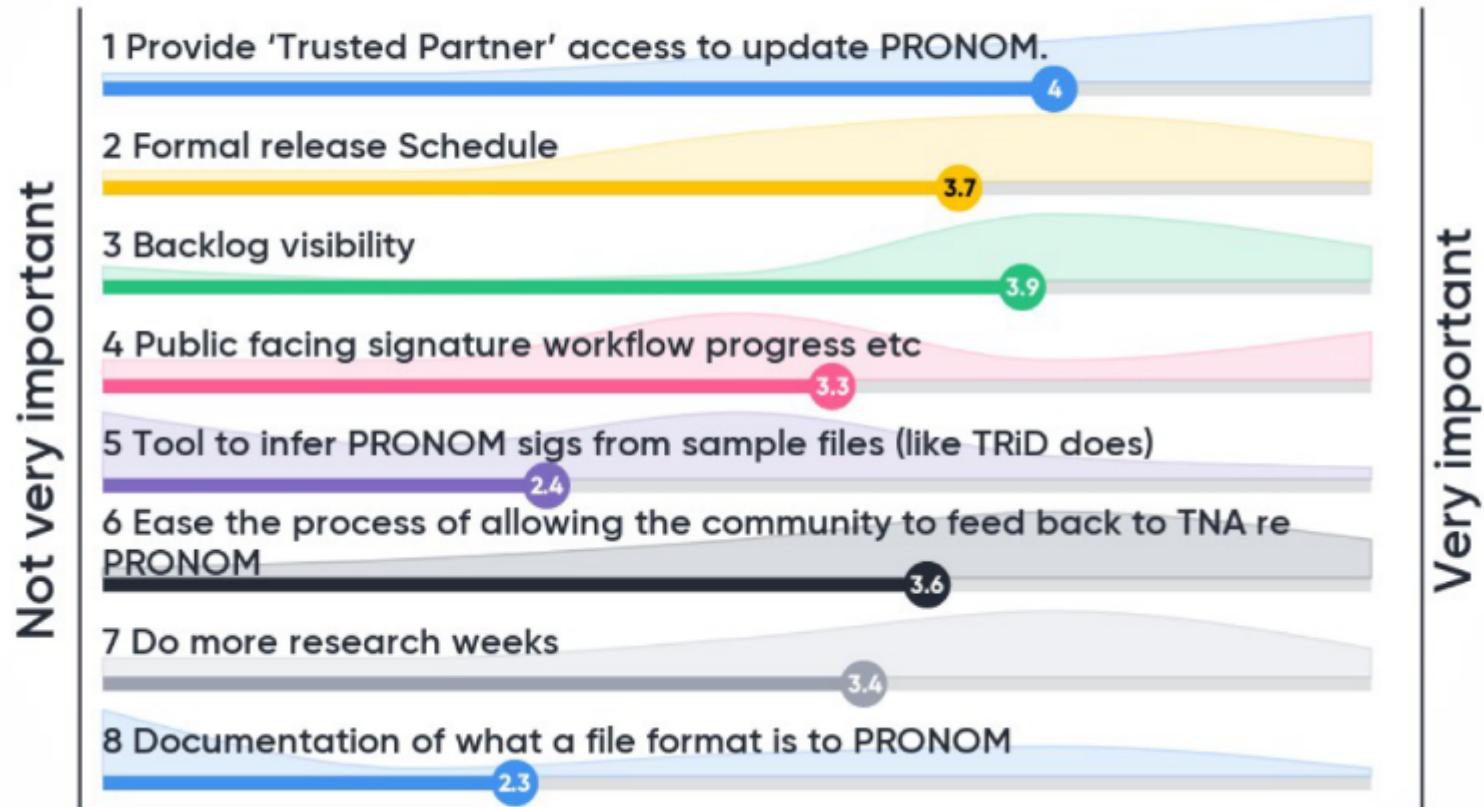
Mentimeter results

Session 2 – The PRONOM API: slider



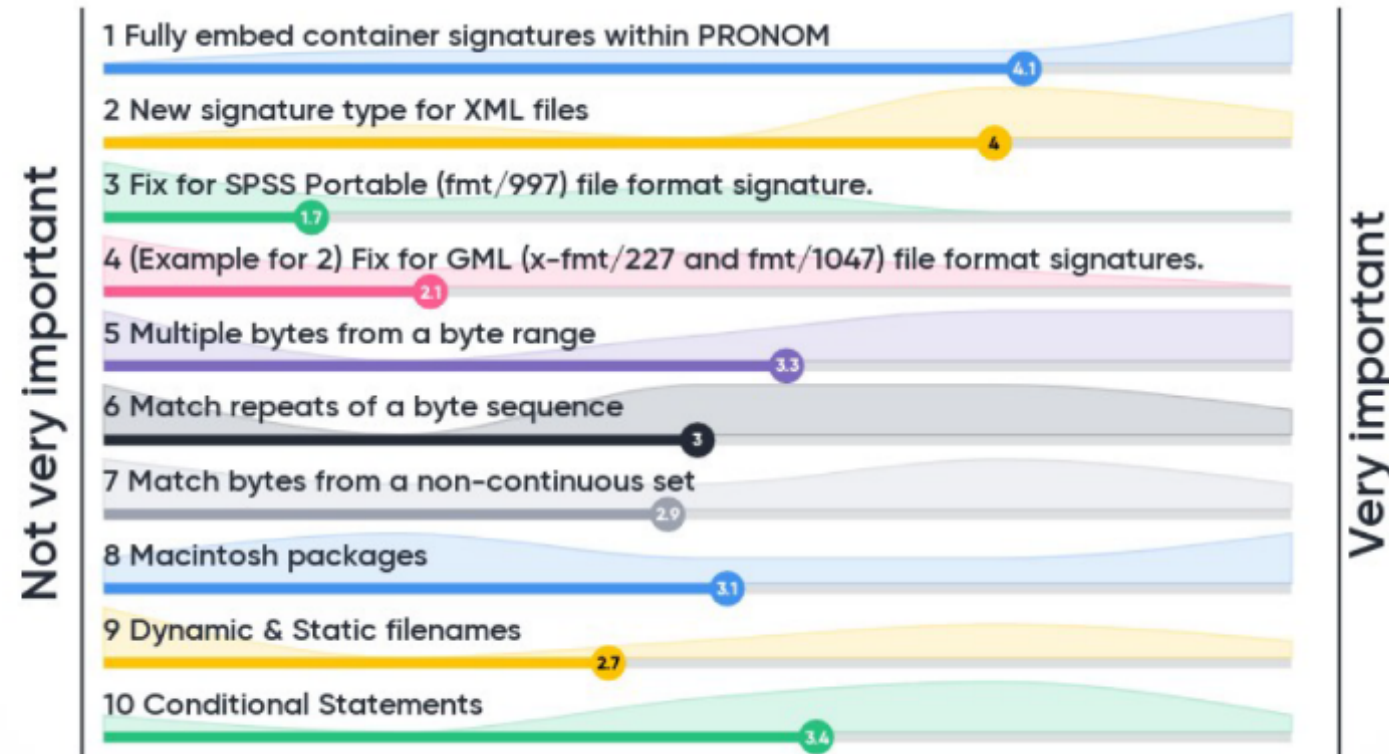
Mentimeter results

Session 3 - Other requirements: slider



Mentimeter results

Session 4 – Extend the PRONOM Signature Vocabulary



Prototypes produced – graph data model

Navigation

[Summary](#)
[Ontologies](#)
[Class Hierarchy](#)
[Classes](#)
[Annotation Properties](#)
[Datatype Properties](#)
[Object Properties](#)
[RDF Properties](#)
[Individuals](#)
[All Resources](#)
[Glossary](#)

[pronom2020:ContainerFile](#)
[pronom2020:ContainerSignature](#)
[pronom2020:ContainerType](#)
[pronom2020:Country](#)
[pronom2020:Document](#)
[pronom2020:DocumentType](#)

Summary

Class [pronom2020:ActedUpon](#)
subClassOf: [pronom2020:Family](#)

Class [pronom2020:Activity](#)
subClassOf: [pronom2020:Relation](#)

Class [pronom2020:ActivityType](#)
subClassOf: [pronom2020:Reference](#)

Class [pronom2020:Actor](#)
subClassOf: [pronom2020:Related](#)
subClassOf: [pronom2020:ActedUpon](#)

Class [pronom2020:ActorType](#)
subClassOf: [pronom2020:Reference](#)

Class [pronom2020:Alias](#)
subClassOf: [pronom2020:Related](#)

Class [pronom2020:ByteOrder](#)
subClassOf: [pronom2020:Reference](#)

Class [pronom2020:ByteSequence](#)
subClassOf: [pronom2020:byteSequence.ByteOrder](#) max 1
subClassOf: [pronom2020:byteSequence.ByteSequencePosition](#) exactly 1
subClassOf: [pronom2020:byteSequence.Offset](#) max 1
subClassOf: [pronom2020:byteSequence.IndirectOffsetLength](#) max 1
subClassOf: [pronom2020:byteSequence.MaxOffset](#) max 1
subClassOf: [pronom2020:byteSequence.Value](#) exactly 1
subClassOf: [pronom2020:Related](#)
subClassOf: [pronom2020:byteSequence.IndirectOffsetLocation](#) max 1

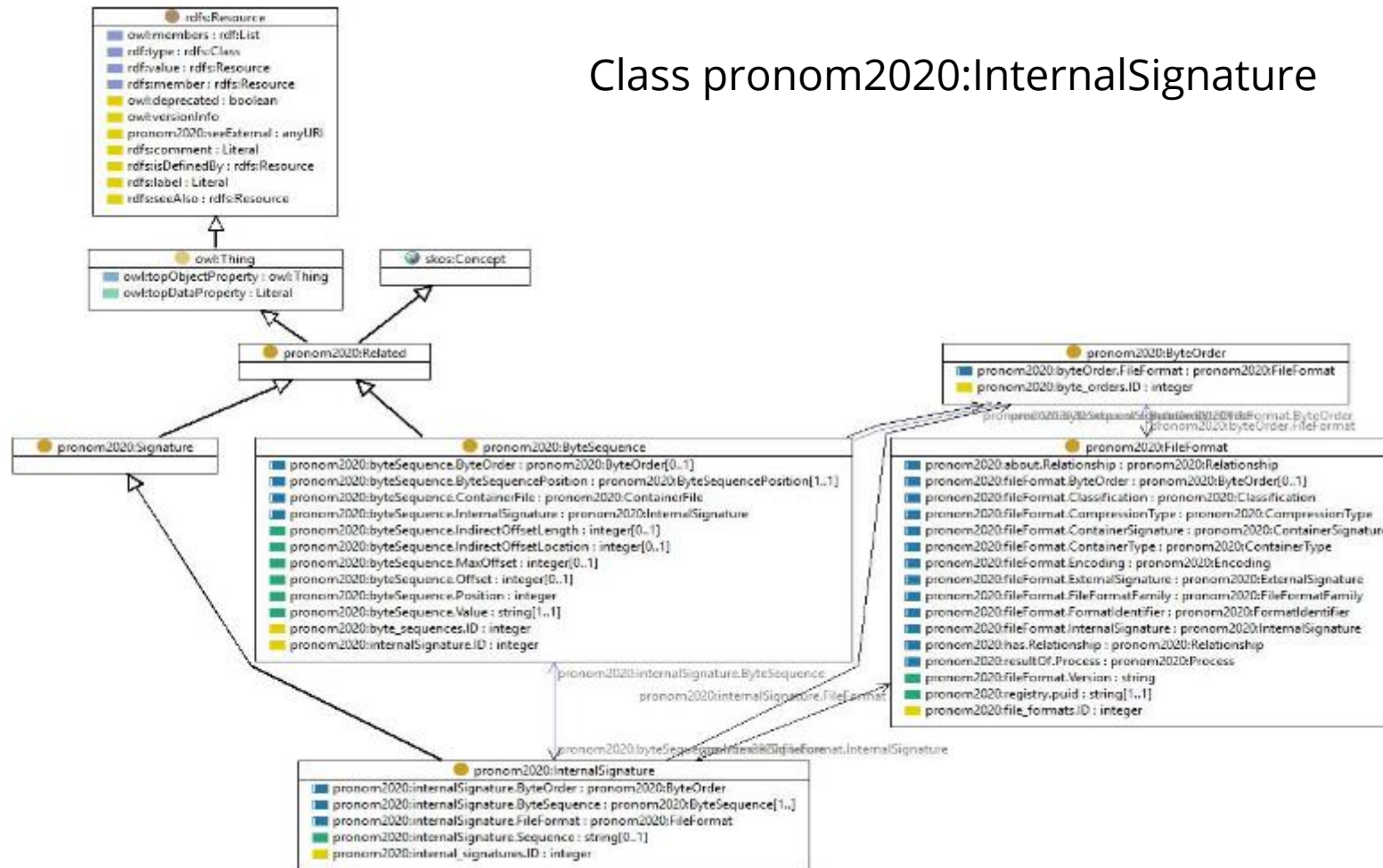
THE

NATIONAL

ARCHIVES

Prototypes produced – graph data model

Class pronom2020:InternalSignature



THE
NATIONAL
ARCHIVES

Prototypes produced – editorial service

Resources	>	FileFormats	Blitz3D File Format
Familys	>	<	
References	>		
Registrys	>		
Relateds	>		
Relations	>		

File Format (Audio/Video Interleaved Format)

Summary

Classification

PUID:*

fmt-5

comment:

Audio/Video Interleaved Format (AVI) is a chunk-based video format developed by Microsoft, based on the generic Resource Interchange File Format (RIFF) specification developed by Microsoft and IBM. Structurally, an AVI file is composed of a number of chunks, each

version:

Note:

see external:

is classified as (2)

Audio Video

belongs to file format family (0)

has format identifier (2)

video/x-msvideo public.avi

has alias (1)

AVI

uses byteOrder:

Little Endian

is identified with containerSignature (0)

Internal Signature

has internal Signature (1) AVI

External Signature

associated with externalSignature (1) avi

Prototypes produced – report service

```
▼<ContainerSignatureMapping Version="100" DateCreated="2020-06-03T13:45Z">
  ▼<ContainerSignatures>
    ▼<ContainerSignature Id="1000" ContainerType="OLE2">
      <Description>Microsoft Word 6.0/95 OLE2</Description>
      ▼<Files>
        ▼<File>
          <Path>CompObj</Path>
          ▼<BinarySignatures>
            ▼<InternalSignatureCollection>
              ▼<InternalSignature Id="1000">
                ▼<ByteSequence Reference="BOFOffset">
                  ▼<SubSequence Position="1" SubSeqMaxOffset="1024" SubSeqMinOffset="40">
                    <Sequence>10 00 00 00 'Word.Document.' ['6'-'7'] 00</Sequence>
                  </SubSequence>
                </ByteSequence>
              </InternalSignature>
            </InternalSignatureCollection>
          </BinarySignatures>
        </File>
      </Files>
    </ContainerSignature>
    ▼<ContainerSignature Id="10000" ContainerType="ZIP">
      <Description>Java Archive Format</Description>
      ▼<Files>
        ▼<File>
          <Path>META-INF/MANIFEST.MF</Path>
          ▼<BinarySignatures>
            ▼<InternalSignatureCollection>
              ▼<InternalSignature Id="10000">
                ▼<ByteSequence Reference="BOFOffset">
                  ▼<SubSequence Position="1" SubSeqMaxOffset="0" SubSeqMinOffset="0">
                    <Sequence>'Manifest-Version:'</Sequence>
                  </SubSequence>
                </ByteSequence>
              </InternalSignature>
            </InternalSignatureCollection>
          </File>
        </Files>
      </ContainerSignature>
    </ContainerSignatures>
  </ContainerSignatureMapping>
```

THE

NATIONAL

ARCHIVES

Prototypes produced – SPARQL endpoint



RDF4J Server

Repositories

New repository

Delete repository

Explore

Summary

Namespaces

Contexts

Types

Explore

Query

Saved Queries

Export

Modify

SPARQL Update

Add

Remove

Clear

System

Information

Query Repository

Query Language: SPARQL

1

Query:

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!

THE

NATIONAL

ARCHIVES