

Towards Resolution Services for Text URIs

Hugh Cayless (Duke) & **Matteo Romanello @mr56k** (EPFL)

Workshop Texts and APIs: the Distributed Text Services

Hamburg, 15-16 July 2019

Outline

1. Resolvable text URIs: why?
2. Use cases
3. Text resolution services
4. Wrap up

1. Why resolvable text URIs?

What is a resolvable text URI?

Places

Athens → <https://pleiades.stoa.org/places/579885> → place description as RDF [name variants, coordinates, related places, etc.]

Texts

Hom. // 1.1 → <http://purl.org/hucit/kb/urn:cts:greekLit:tlg0012.tlg001:1.1> → text description as RDF [author, title, citation unit (book/line)]

Key idea

A centralized service to create resolvable URIs for texts would **relieve publishers** from having to mint such URIs, thus lowering the barrier for the integration of textual resources into the LOD graph.

Provided that texts are made available via **standard APIs** and described by **sufficient metadata**, the creation and resolution of URIs can be delegated to a centralized service.

Resolution of Text URIs

<resolverURI>/urn:cts:greekLit:tlg0012.tlg001

- What is this ?
- What **editions and translations** exist ?
- What are its **citable units** ?
- What other **related resources** exist ?
- ...

<resolverURI>/urn:cts:greekLit:tlg0012.tlg001:1.1

- What is this ?
- Where do I find electronic **editions and translations** of this passage ?
- ...

How to make it work?

Through use of:

1. APIs (CTS, DTS) to share text collections
2. Unique identifiers for texts **and** their citable units
3. Shared metadata vocabulary to describe texts

Previous/related work

Text APIs

- Canonical Text Services
- Distributed Text Services
- Lots of custom, project-based APIs

Text Ontologies/Knowledge Bases

- LAWD, FRBRoo, SPAR ontologies
- HuCit ontology and knowledge base (Romanello & Pasin 2017)

Pelagios WG

- Community grant 2018
- Workshop at Duke (June 2018) with 12 participants on *resolvable text URIs*

2. Use Cases

Linking (1)

To link towards a text (passage) without specifying an edition/version

Bryn Mawr Classical Review 2013.01.10

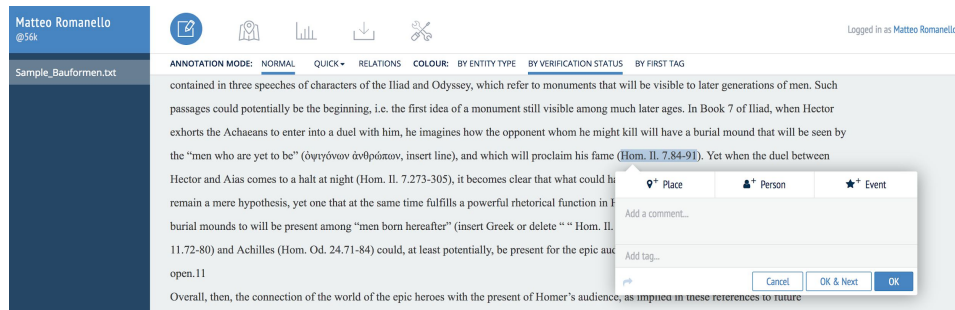
Mélina Tamiolaki, *Liberté et esclavage chez les historiens grecs classiques. Hellenica*. Paris: Presses de l'Université Paris-Sorbonne (PUPS), 2010. Pp. 503. ISBN 9782840506881. €28.00 (pb).

Reviewed by Alberto Maffi, University of Milano-Bicocca (alberto.maffi@unimib.it)

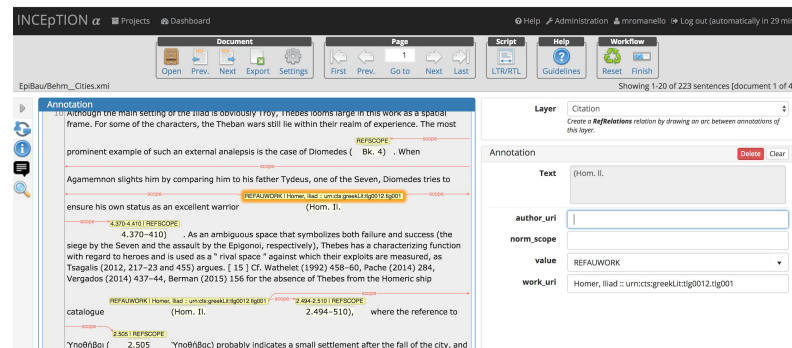
The book is divided into three parts: *Liberté et esclavage entre les cités ou les peuples*; *Liberté et esclavage à l'intérieur des cités ou des peuples*; *Liberté et esclavage en dehors de la cité*. Each of the first two parts is in turn divided into three chapters, dedicated respectively to the analysis of the work of Herodotus, Thucydides and Xenophon. The third part is divided into two chapters, and is devoted to analyzing all of the works of Xenophon: the first chapter, entitled *Les limites et les ambiguïtés de la soumission volontaire au chef charismatique*, contains an analysis of the *Cyropedia*, which is followed by a study of the *Oeconomicus* (in relation to the submission of the slaves and the master's wife), the second chapter is entitled *L'exemple de Socrate comme chef exceptionnel et la réflexion sur la liberté du monde intérieur*. The volume concludes with two appendices containing the critical exegesis of two famous passages ([Thuc. 1.101.2.](#) and [Xen. Hell. 3.3](#)) and a table containing the terminology of freedom and slavery in the works of the three historians.

Annotation (2)

Text URIs as a controlled vocabulary in **annotation tools** to create **semantic statements** about texts (ontologies: SAWS, CiTO, INTRO, etc.)



recogito.pelagios.org



inception-project.github.io

3. Text resolution services

Resolver or catalog?

<resolverURI>/urn:cts:greekLit:tlg4036.tlg006

- Proclus' *Element of Physics*
- Perseus Catalog → Bibliographic information
- SAWS project → Ritzenfeld's edition (urn:cts:greekLit:tlg4036.tlg006.saws01)
- Future editions/translations

Resolver or catalog?

Responsibilities:

- Should enable users (humans/machines) to **look up identifiers**
- Should allow **text repositories** (exposing supported APIs) to be **registered**
- Should provide **URIs** for each citable passage of registered texts **that resolve** to an RDF representation

Implementing Resolution Services

Key requirements:

1. **APIs** to access and retrieve texts over standard protocols
2. **Unique identifiers** for texts and their citable units
3. **Metadata** to describe texts, and the relations with other texts/resources

Components:

1. Registry of text services (supported APIs)
2. Resolver of text identifiers
3. Shared document metadata scheme (to enable aggregation, browsing, etc.)

APIs – Canonical Text Services

Importance:

- It's become an important piece of infrastructure:
 - Text Repositories
 - Perseus Catalog
 - Perseus Digital Library
 - Client libraries (Java, Scala, Python, GO)
 - Tools (Capitains)

Issues:

- Strong commitment to ontologies of canonical texts
- **Scalability** issues with large collection of small texts (e.g. Papyri.info)
- Development: **project-driven VS community-driven**

APIs – Distributed Text Services

What ?

- A **community-driven** standard to share (TEI) texts via HTTP queries
- Status: **First Public Working Draft**, <https://w3id.org/dts/>

Why ?

- As **generic** as possible: works with any text having a TEI representation

How ?

- 3 endpoints: 1. Collection, 2. Navigation, 3. Document
- Built upon Hydra, speaks JSON-LD

Demo: <https://dev.chartes.psl.eu/dts-demo/>

Unique identifiers (in the wild)

<u>urn:cts:latinLit:phi0978.phi001.perseus-lat1:11.4.11</u>					
<u>prefix</u>	<u>namespace</u>	<u>textgroup</u>	<u>work</u>	<u>version</u>	<u>passage</u>
		<u>work</u>			

urn:cts:greekLit:tlg0012.tlg001.perseus-grc2:1.1 (Perseus, CTS)

urn:cts:frenchLit:pos2001.pos200104.positionThese-fr1 (ENC, DTS)

urn:dts:betmasMS:BLorient718:1r (Beta maṣāḥəft, DTS)

...but also... **ark:/12148/cb31463392g** (BNF)

Document metadata scheme

Challenge:

- CTS API makes **strong assumptions** about nature of texts (canonical)
- DTS makes **no assumptions** on this respect
- How to strike a balance between two approaches?

Aggregation of text collections:

- How do I know which **copy** I want ?
- What common **metadata to navigate** this aggregated view ?
- How to determine **identity** across text repositories ?
- How to handle **versioning** (same “work”, but different versions)

Shared Document Metadata Scheme

Rationale:

- As generic as possible
- Distinction between **source/edition**
- Vocabularies: DC, CIDOC-CRM, LAWD, SPAR (CiTO, FaBiO), FRBRoo

Usage:

- In the `dts:extensions` property of DTS replies

Examples:

- Papyrus from Papyri.info (TEI/XML via DTS)
- OCR text of a printed translation of a Greek Text (TEI/XML via DTS)

Example 1: Dig. Ed. of Papyrus

```
{
  "@context": {
    "cito": "http://purl.org/spar/cito",
    "dc": "http://purl.org/dc/terms/",
    "ecrm": "http://erlangen-crm.org/current-version",
    "lawd": "http://lawd.info/ontology/"
  },
  "@id": "https://papyri.info/ddbdp/p.col;8;237/source",
  "@type": ["lawd:Edition", "fabio:DigitalItem"],
  "dc:source": {
    "@id": "https://papyri.info/ddbdp/p.col;8;237/work",
    "@type": ["lawd:Edition", "fabio:DigitalItem"],
    "cito:isCitedBy": "http://papyri.info/biblio/65200",
    "dc:source": {
      "@id": "https://papyri.info/ddbdp/p.col;8;237/original",
      "@type": "lawd:WrittenWork",
      "rdfs:seeAlso": "http://www.trismegistos.org/text/10553",
      "lawd:origin": "https://pleiades.stoa.org/places/737081",
      "lawd:foundAt": "https://pleiades.stoa.org/places/737081",
      "ecrm:P62i_is_depicted_by": "http://papyri.info/ddbdp/p.col;8;237/images",
      "dc:created": "3 June, 381 or 382"
    }
  }
}
```

Example 2: OCR of printed book

```
"dts:extensions": {
  "@context": { ... },
  "@id": "https://library.org/texts/urn:cts:greekLit:tlg4036.tlg006.boese",
  "ecrm:P1_is_identified_by": {
    "rdfs:label": "urn:cts:greekLit:tlg4036.tlg006.boese",
    "ecrm:P2_has_type": "http://purl.org/hucit/kb/types/CTS_URN"
  },
  "@type": "fabio:DigitalItem",
  "fabio:isPortrayalOf": "http://purl.org/hucit/kb/works/3822",
  "frbr:exemplarOf": "https://library.org/editions/urn:cts:greekLit:tlg4036.tlg006.boese",
  "dc:source": {
    "@id": "https://library.org/editions/urn:cts:greekLit:tlg4036.tlg006.boese",
    "@type": "lawd:Edition",
    "fabio:hasPublicationYear": "1958",
    "dc:medium": "printed edition",
    "fabio:hasPlaceOfPublication": "http://www.geonames.org/2950159/berlin",
    "dc:language": "de",
    "dc:title": "Die mittelalterliche Übersetzung der Stoicheiosis phusike des Proclus",
    "dc:bibliographicCitation": "Boese, H., Die mittelalterliche Übersetzung der Stoicheiosis phusike des Proclus (Deutsche Akademie der Wissenschaften zu Berlin, Institut für griechisch-römische Altertumskunde, Veröffentlichungen 6), Berlin: Akademie Verlag, 1958.",
    "fabio:isManifestationOf": "http://purl.org/hucit/kb/works/3822"
  }
}
```

Wrap up

- Results of the Pelagios-funded *Linked Texts* WG
- **Problem:** provide resolvable URIs for texts and their citable units
- **Idea:** centralized service to mint and resolve URIs for texts available via registered text repositories
- Service **implementation** should rely on:
 - Standard APIs (CTS/DTS)
 - Unique identifiers
 - Shared document metadata scheme

Thanks for your attention !

Hugh Cayless
Matteo Romanello

@hcayless
@mr56k

hugh.cayless@duke.edu
matteo.romanello@epfl.ch