So, what are the Distributed Text Services

Hamburg, Texts and APIs workshop

July, 15-16 2019

Thibault Clérice, École Nationale des Chartes

École nationale des chartes | PSL ★

Thanks to the Procope Program and the French ambassies of Switzerland and Germany.

What's DTS ?

1. How does the group work ?

2. Broadly : Principles, Structure and Examples

3. Questions ?

The team

Chair: Jonathan Robie

Technical Comittee: Bridget May Almas, Hugh Cayless, Vincent Jolivet, Ian W. Scott, James Tauber, Thibault Clérice

Recurrent invitees and feedbackers: E. Morlock, M. Romanello, P. Luzzio

History

- Simple statement: no community-wide accepted standard for sharing text over HTTP queries.

- IIIF is a success!

- December 2015: meeting at Tufts with European and American colleagues

- Decision taken: we need to build something

- Working group meeting funded by Pelagios in Duke University (June 2018)

- First draft presented at TEI Tokyo in September 2018

- Presentation in France in Feburary at the Ecole Nationale des Chartes

École nationale des chartes | PSL ★

The teamwork

- Issues on github, pull requests.

- Bi-monthly meetings

- Assign review, discuss from one meeting to another

- If possible, consensus. Otherwise vote.

- Open process, meetings notes are available.

General principles

1. DTS is an API to **serve texts** and a **metadata catalog** to discover these texts.

2. It has **no assumption on the technical implementation**.

3. The API should follow a well **established infrastructure**: we choose REST.

4. It should be **modular**, while still maintaining a **core of functionalities**.

5. Each response must exist in at least one specified **standardized format**.

General Architecture

1. If the API serves the same role as a human website or a physical library, what do you need ?
   - Catalog
   - Text retrieval
   - Index
2. Not everybody wants or needs everything.
3. HYDRA standardized API, with TEI when required

General architecture

Like any public building or any website, we need an entrance that will actually tell you what is there, some kind of plan :

http://tnah.chartes.psl.eu/2019/dts/

```
{
  "@context": "/dts/api/contexts/EntryPoint.jsonld",
  "@id": "/2019/dts/",
  "@type": "EntryPoint",
  "collections": "/2019/dts/collection",
  "documents": "/2019/dts/document",
  "navigation": "/2019/dts/navigation"
}
```

General Architecture : Catalog

1. Official name: Collection endpoint

2. Every project has its metadata ?

    i. Then a base requirement for real interoperability : at least Dublin Core Terms

    ii. But give freedom for projects : a specific zone in the data model for

3. Let's not impose our model of data: collections can be cyclical graph, tree, you can do FRBR, YOU'RE FREE.

4. Let's not build too much of vocabulary

5. ... But still do (citeDepth, citeStructure) when required.

6. Prepare for every size of corpus: YES to pagination!

Catalog: some examples

Developed with Lucie Vieillon & HÃ©lÃ¨ne Dartois from ENC MA

A root collection

> http://tnah.chartes.psl.eu/2019/dts/collection

A readable collection :

> http://tnah.chartes.psl.eu/2019/dts/collection?
> id=https://www.wikidata.org/wiki/Q1340251

General Architecture : Text

1. Official name : **Document Endpoint**

2. **Retrieve** full or partial text

3. **Identifier** for texts, but also for passage

4. Allow passage **range**

5. Requires **TEI** but allows **media negociation**

6. Provide normalized ways of telling where you are without imposing things on the content format: use **Link header**

Text: some examples

Full text

> http://tnah.chartes.psl.eu/2019/dts/document?id=https%3A%2F%2Fwww.wikidata.org%2Fwiki%2FQ1340251

Passage

> http://tnah.chartes.psl.eu/2019/dts/document?id=https%3A%2F%2Fwww.wikidata.org%2Fwiki%2FQ1340251&ref=1

Passage Range

> http://tnah.chartes.psl.eu/2019/dts/document?id=https%3A%2F%2Fwww.wikidata.org%2Fwiki%2FQ1340251&start=2&end=3

General Architecture : Index

1. Official name : **Navigation Endpoint**

2. Provide the list of passages in a text

3. Texts can have multiple levels, multiple citation schemes
   - But, one string can identify only one passage

4. Citation scheme can be defined

5. Passage can have their own metadata: author, date, etc. using DCT

Index examples

All

> http://tnah.chartes.psl.eu/2019/dts/navigation?
> id=https%3A%2F%2Fwww.wikidata.org%2Fwiki%2FQ134025
> 1

Within a range

> http://tnah.chartes.psl.eu/2019/dts/navigation?
> id=https%3A%2F%2Fwww.wikidata.org%2Fwiki%2FQ134025
> 1&start=1&end=5&level=0

Thanks !

Email : thibault.clerice@chartes.psl.eu

Twitter : ponteineptique