HAL
archives-ouvertes.fr

The Standardization Survival Kit : For a wider use of standards within Arts and Humanities

Marie Puren, Charles Riondet, Dorian Seillier, Lionel Tadjou

▶ **To cite this version:**

Marie Puren, Charles Riondet, Dorian Seillier, Lionel Tadjou. The Standardization Survival Kit : For a wider use of standards within Arts and Humanities. Journée de formation : "Gérer et explorer les données textuelles", Apr 2018, Le Mans, France. 2018. hal-01763688

HAL Id: hal-01763688
https://hal.inria.fr/hal-01763688

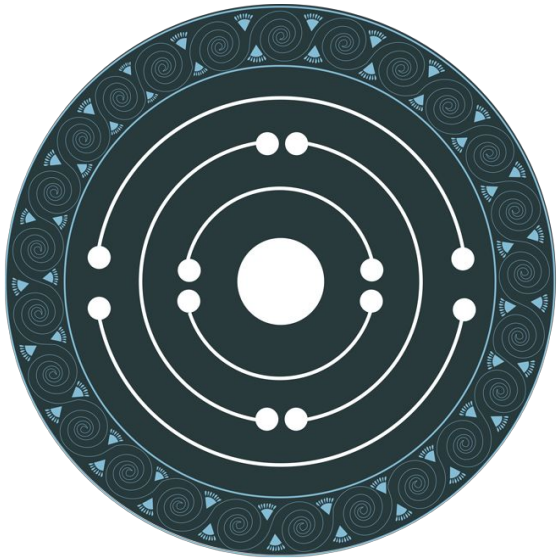
Submitted on 11 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Standardization Survival Kit

For a wider use of standards within Arts and Humanities



PARTHENOS

Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

http://www.parthenos-project.eu/

Standards ?

They are non legally binding documents produced by an organisation ensuring :

- International **consensus** building (not a one person's/one group's work)
- **Communication** (standards cannot be internal to an organisation)
- **Maintenance** (evolution according to users' needs, technology change, etc.)

Some examples

- ISO standards, W3C recommendations or TEI guidelines comply to these principles

- Horizontal standards (cross-domain: ex. XML, ISO/CEI 10646/Unicode) vs. Vertical standards (specific to one domain: ex. ISO 24611 for morpho-syntactic annotations, EAD for archival collections)

Research data ?

Data created, collected and/or transformed during a research project.

Need for an active management during the whole Research process, and even after.

Good data management covers:

- **Description**
- **Hosting / archiving**
- **Access**
- **Reuse**

“Standards are a key to great digital research, which helps to discover and understand our cultural and societal life.”

Laurent Romary

Chairman of the Technical Committee "Terminology and other language and content resources" of the International Organization for Standardization (ISO)

Why using standards ?

To produce interoperable, sustainable & reusable results for research.

Here the SSK comes into play ...

LES AVENTURES DE
MORK ET TORK!

The SSK : a toolkit for Humanities scholars

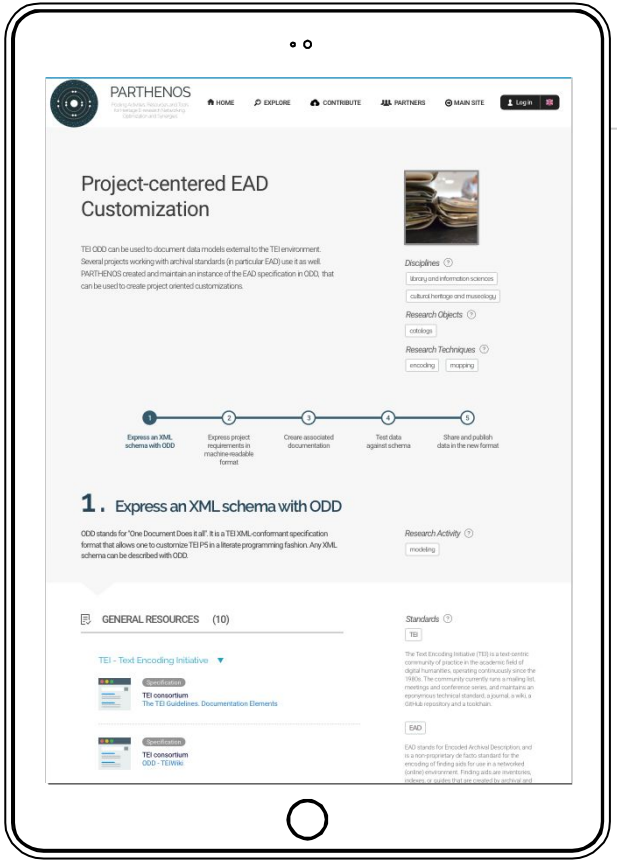
- Documenting
- Supporting
- Training
- Communicating

What the SSK is not

- A software where you can perform actions
- A data repository
- A catalog of standards or resources

The scenarios

Providing **contextual information** and **relevant examples** on how standards can be applied in a given research project.



What Tork & Mork tell us

Hey Mork ! For my project, I need to create an annotated linguistic corpus, but I don't know where to start and which format to use...

Well, just check the SSK ! You will find the perfect scenario "*Creation of a TEI-based corpus*". It will guide you through all the necessary steps and give you the best practices !

Three layers

Scenario

A complete and generic research use case composed of several steps to be followed.

Step

A unique task to be performed inside a scenario with the help and recommendation of one or several resources.

Resource

A standardized tool / service / document guiding the researcher in her/his tasks completion.

The TEI model

The scenarios are described using the TEI format (Text Encoding Initiative). All the information displayed within the SSK proceed from TEI files.

```
<text>
   <body>
      <div type="scenario" xml:id="sc_schemaCustomization">
         <head xml:lang="en" type="scenarioTitle">Harmonization of digitized textual resources
            with the DTABf</head>
         <desc type="definition" xml:lang="en">Currently, initiatives for the digitization of
            textual resources and their provision to the interested community are manifold and
            various. Hence, scholars who want to base their research on digitized texts,
```

What is the TEI ?

The TEI is made up of recommendations for the creation and management of all types of data coming from researchers in the Arts & Humanities (manuscripts, archival documents, old inscriptions, etc.)

Three great principles :

- fostering meaning before format
- ensuring software independency
- being supported by a strong community

How the TEI works in the SSK ?

- A scenario is a list of events (**<tei:listEvents>**), each step in a scenario is an event (**<tei:event>**).
- TEI files are all available on **Github** (https://github.com/ParthenosWP4/SSK).
- Documentation: ***https://parthenoswp4.github.io/SSK/SSKspec.html***

Handling the resources

Standards documentation (ISO, TEI)

Official publications and reports (D4Science, HAL, Zenodo)

Documentation

Reference libraries organised by domains and standards

Maintenance with Zotero

Bibliography

Code snippets (GitHub)

Tools & services (D4Science)

Technical resources

Wikis

Blog posts (Hypotheses.org)

Discussion lists

User communities

Our ontologies

- Disciplines (AureHAL)
- Research objects (TaDiRAH, NeDiMAH)
- Research activities (TaDiRAH)
- Research techniques (TaDiRAH)
- Standards (our own knowledge base)

Next features

- Browsing **vocabularies** and **bibliography**
- Having a profile to manage one's personal space :
  - handle **bookmarks**
  - customize scenarios
- **Contributing** (in TEI or directly on the interface) :
  - creation
  - edition
  - customization
- **Multilingual** interface

Let’s test it !

https://ssk-application.parthenos.d4science.org/ssk/

*and mail us : **ssk@inria.fr***

Exercises

- Find a research scenario in your academic field and follow the steps of the scenario
- You want to create a digital edition of your research corpus. First you need some guidelines on how you can digitize your (very fragile) source. Use the SSK to find the best way to go.
- You want to find the research steps where the use of the Dublin Core standard is mentioned. Use the SSK to find a corresponding result.

Exercises

Create a scenario corresponding to your research field to document your practices

- Group work
- Based on the steps and resources found in the SSK
- Use available vocabularies
- result in a simple text or in TEI

Use the documentation