The Standardization Survival Kit (SSK)

Bringing best practices to research communities in the Humanities

PARTHENOS

Why using standards?

Manage your data, don't let your data manage you...

... and produce interoperable, sustainable & reusable results for research.

Why using **standards**?

*"Standards are a **key to great digital research**, which helps 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)

What are standards?

3 keywords

- They express a **consensus**
- They are published and easily **accessible**
- They are **maintained**

Data formats

- XML : https://www.w3.org/TR/xml/
- TIFF : ISO 12639:1998

Protocols

- ISO 11554:2017 : Test methods for laser beam power, energy and temporal characteristics

Standards?

Present at each step of the research process (examples)

- **Production**: ALTO XML, standard for recording layout and logical structure of OCRed text;
- **Processing**: Linguistic annotation encoded in XML TEI (Text Encoding Initiative);
- **Archiving**: OAIS (Open Archival Information System, ISO 14721:2012) => conceptual model dedicated to the management, archiving and long-term preservation of digital documents.

The SSK and PARTHENOS

Initial goal of the ***Standardization Survival Kit*** in **PARTHENOS** :

Support and provide expertise to researchers in their use of standards

Which communities?

- Humanities
- Social Sciences
- Heritage Science

The SSK in PARTHENOS

Positive context of an European project:

- **Diversity**: many experts from different disciplines
- **Synergy**: willingness to cooperate

⇒ Opportunity to build use cases, inspired from real life.

The SSK in PARTHENOS

Development team

Inria ALMAnaCH

- Laurent Romary (DR, WP leader, supervisor)
- Marie Puren, Charles Riondet (project management, data model)
- Dorian Seillier (UI/UX Design)
- Lionel Tadjou, Damien Biabiany (development web)

Iterations, beta tests

PARTHENOS WP4 on Standardization:

- Klaus Illmayer (OEAW),
- Karolien Verbrugge (NIOD),
- Roberta Giacomi (SISMEL),
- Panos Siozos (FORTH),
- and many more

Evolution of the concept (1)

1. **Support and provide expertise to researchers in their use of standards;**
2. Give context to standards;
3. Link them to a concrete research activity (TaDirah – Taxonomy of Digital Research Activities in the Humanities);
4. Link activities between them –> Describe a research process built on the use of standards.

Evolution of the concept (1)

Not every research step is related to a standard:

- Ethics and legal issues;
- Evaluation and comparison of results;
- ...

But the notion of **Best practices** is always relevant

Evolution of the concept (2)

When standards are protocols :

- Normative texts are very formal documents, difficult to read;
- A protocol = A suite of tasks using tools and techniques.

Evolution of the concept (1)

1. **Support and provide expertise to researchers in their use of standards;**
2. Give context to standards;
3. Link them to a concrete research activity (TaDirah – Taxonomy of Digital Research Activities in the Humanities);
4. Link activities between them –> Describe a research process built on the use of standards.
5. Platform for
    a. Documenting research best practices in digital environment
    b. "Human readable" expression of protocol standards

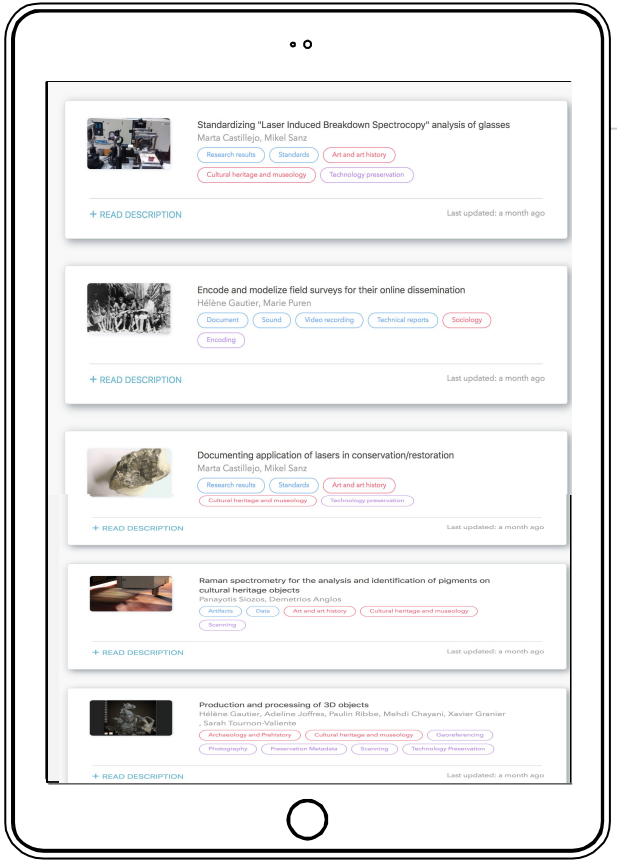
The SSK in PARTHENOS, and beyond

Propose a pleasant, sustainable and adaptable service

- UI/UX designer
- "agile" method: brainstorming, user/usability tests
- "soft" modeling of data
- Meet the potential user communities.

The scenarios

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



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.

3-level Structure

Scenario description (*techniques, disciplines, objets)*

1. Step description *(activities, standards)*
    a. *resource*
    b. *resource*
2. Step description *(activities, standards)*
    a. *resource*
    b. *resource*
3. Step description *(activities, standards)*
    a. *etc.*

Resources about standards

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

A high-level research guide

Collaborative Digital Edition of a Musical Corpus

Vincent Besson , Adeline Joffres , Hélène Gautier

Last updated: 4 months ago

Musicology And Performing Arts

Sound · Score · Multiple Score Formats

A project aims to do a digital edition of a musical corpus. The researchers need to be able to encode a broad range of musical documents in a machine-readable structure. The data to be encoded may include the musical content as provided by the composer (notes, pitches, durations, dynamics, etc.), information on the score (incipit, lyrics writer, etc.), information added by a performer when interpreting the content (timing, phrasing, various annotations, etc.), information on the visual appearance of the score (page layout, musical font, etc.) and analyses of the content in any of the other domains. The edition will be structured around a database in order to allow the users to explore it more easily. Furthermore, the project intends to be collaborative, which means it will offer anyone interested the possibility to contribute.

1

Create a digital corpus of musical compositions.

MEI · Transcription · Editing

Select resources to be included in the corpus. After collecting original musical sources, transcribe them adding critical editorial signs and normalizing, where applicable, ancient poetic texts to modern usage. To get directly MEI files, use MEISE (MEI Score Editor).

RESOURCES (specifications, papers, tutorials, etc.)

2

Change format into MEI files if necessary.

MEI · Conversion

Convert the Sibelius files into MEI files through the plugin SibMei.

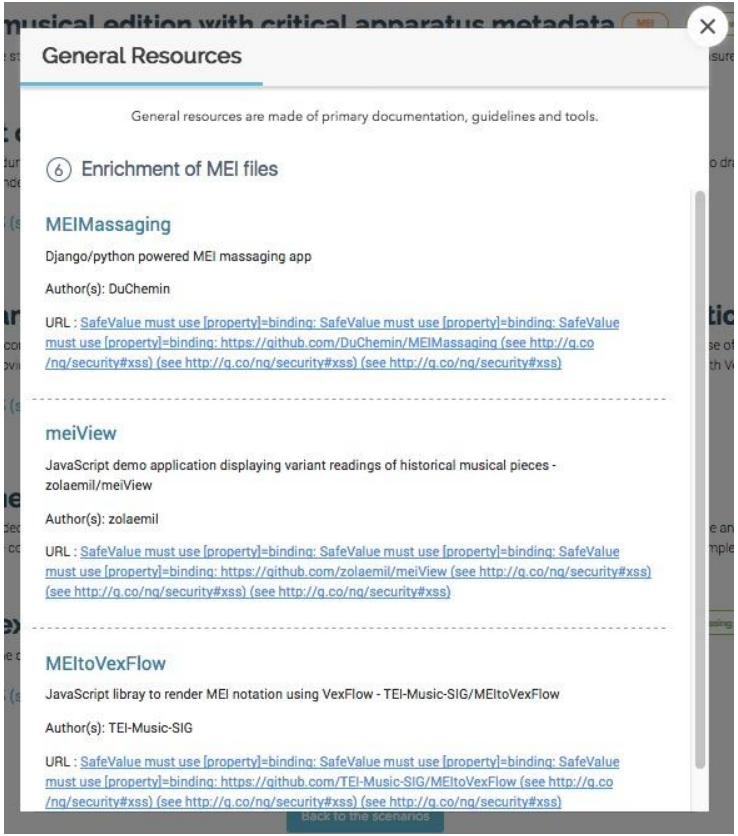
RESOURCES (specifications, papers, tutorials, etc.)

3

Collect and organize metadata

Annotating

Resources and best practices



Easy-to-use & collaborative platform

- Consult and follow the guidelines expressed in the scenarios
- Propose new scenarios

Why contributing ?

- Make your research project align with the best practices in your community
- Get peer review and visibility
- Share a project in another form than the usual blog/article (a new way to disseminate your work)

Coherence with our principles

The SSK, a 100% standards web app

Scenarios in TEI

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

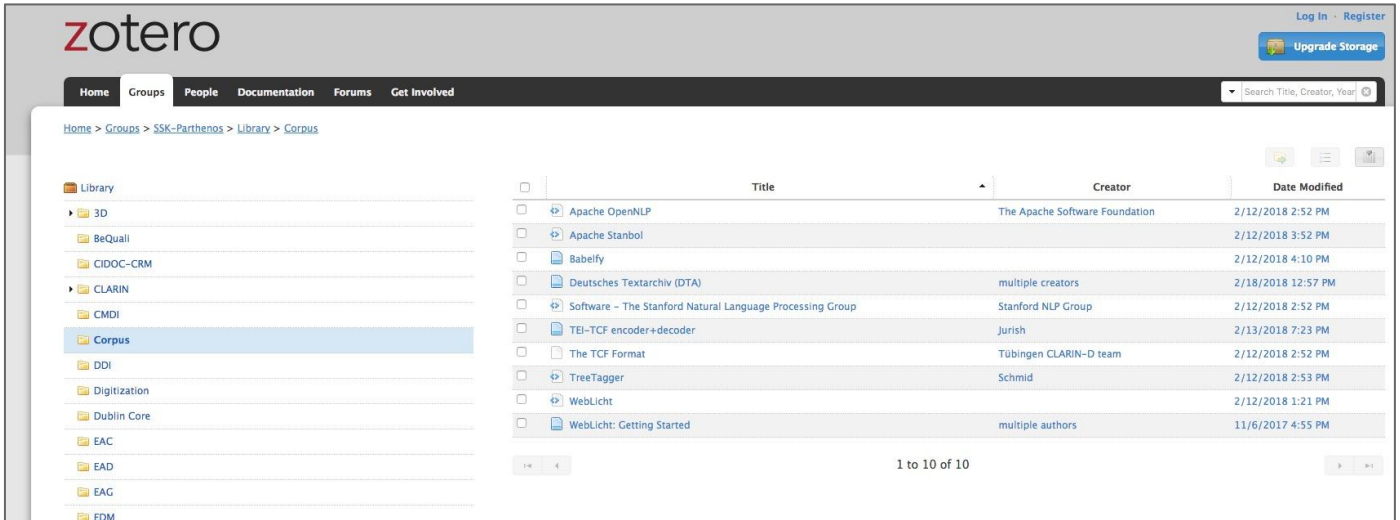
http://github.com/ParthenosWP4/SSK/tree/master

```
<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,
```

```
<listEvent>
 <event ref="step_OtICoP_171117" type="researchStep" xml:id="s1"/>
 <event ref="step_CaC_171117" type="researchStep" xml:id="s2"/>
 <event ref="step_SaD_171117" type="researchStep" xml:id="s3"/>
 <event ref="step_A_171117" type="researchStep" xml:id="s4"/>
 <event ref="step_CiSF_171117" type="researchStep" xml:id="s5"/>
 <event ref="step_TtI_171117" type="researchStep" xml:id="s6"/>
```

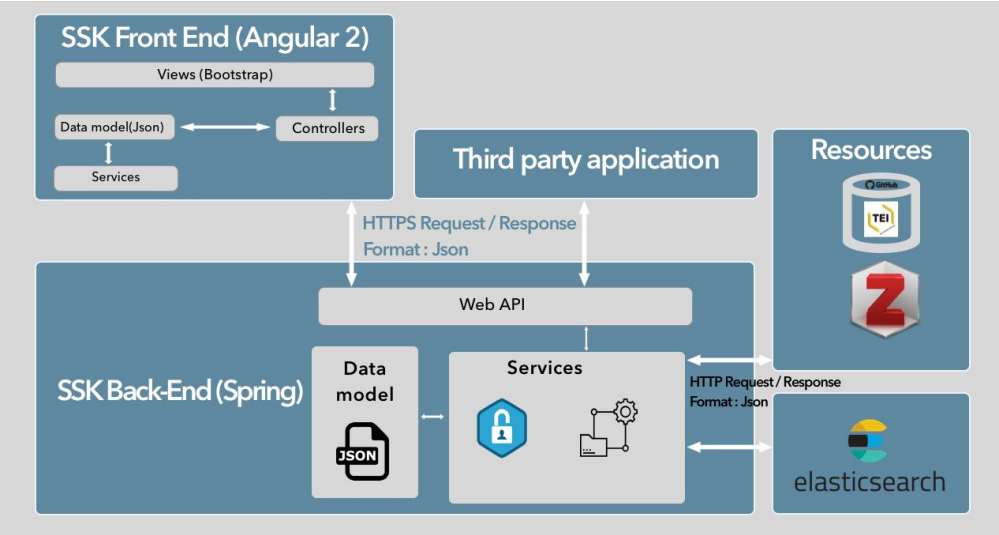
Managing resources with bibliographic standards

All the references are managed by the open source management software Zotero, and can be found in the a dedicated library.
https://www.zotero.org/groups/427927/ssk-parthenos/



RESTful architecture

http://github.com/ParthenosWP4/SSK/tree/dev



- Flexible, easy to deploy and maintain architecture
- Independent entities communicating via REST services.

Sustainability

- Open-source code and GPL licence.
- Underlying data described in TEI and hosted on GitHub under the licence CC-BY
- Bibliographical resources are part of a Zotero open library.
- Double hosting:
  - http://ssk.parthenos-project.eu
  - http://ssk.huma-num.fr
- DARIAH working groups willing to take over intellectual maintenance.

What's next? (early 2019)

- Browsing **bibliography**
- Creating an **account** :
  - to manage **bookmarks**
  - to customize **scenarios** (by combining existing steps from SSK's research scenarios)
- **Contributing** directly on the interface:
  - creation
  - edition
  - customization
- Accessing a **multilingual** interface

Test the SSK!

http://ssk.huma-num.fr

http://ssk.readthedocs.io

ssk@inria.fr