The Standardization Survival Kit: bringing best practices to research communities in the Humanities

Marie Puren[1], Charles Riondet[1], Laurent Romary[12], Dorian Seillier[1], Lionel Tadjou[1]

As an increasing number of primary and secondary legacy sources become digital, more born-digital content is being produced (Hilbert 2012) and more digital tools are being deployed, it is clear that a new generation of digitally-aware scholars in the Humanities is emerging (Latour 2014). The role of the different research infrastructures represented in PARTHENOS is to connect these resources, tools and scholars. But it is also crucial to understand the role that proper data modelling and corresponding standards could play to make digital content more sustainable and reusable. The experience gained within the research infrastructures represented in PARTHENOS, is that there is always an initial phase during which researchers should be made aware of some core standards. Therefore, elementary research projects should be encouraged to refrain from defining their own local formats, and to first demonstrate that their needs are not covered by the wide variety of already existing initiatives in the Digital Humanities landscape. This is basically what has led PARTHENOS to identify the notion of the **Standardization Survival Kit or SSK[3].**

The SSK intends to show that we can and should integrate standardization issues at the core of our infrastructural work. This work may in turn contribute to a wider understanding of the role of scholars within a digital infrastructure, and consequently on how infrastructures could better take into account the variety of research communities in the Arts and Humanities.

Standards usually take the form of documents providing information about practices, protocols, or data formats that can be used as a reference for two parties working in the same field of activity so that they may produce interoperable results. Nevertheless, there is no obligation to follow standards except when one actually wants to produce results that can be compared with those of a wider community. This is why a standardization policy for an infrastructure in the Arts and Humanities should include recommendations as to what approach the scholarly communities could or should adopt with regard to specific standards.

It is in fact essential to provide potential users with i) an awareness of the appropriate standards and the advantages to be gained by adopting them (Romary 2011), and ii) the cognitive tools to help them identify the optimal use of standards through the selection and possibly customisation of a reference portfolio.

The work carried out by the SSK covers four types of activities related to the deployment and use of standards in the Humanities and Cultural Heritage fields:

- ***Documenting*** existing standards to provide reference material for scholars who want to find out more about their role and content.

[1] Inria, team ALMAnaCH, Paris
[2] DARIAH
[3] A demo version is available here: https://ssk-application.parthenos.d4science.org/ssk/.
[3] A demo version is available here: https://ssk-application.parthenos.d4science.org/ssk/.

- ***Supporting*** the actual adoption of standards by identifying how they relate to research scenarios and gathering the essential materials for controlling their deployment (e.g. schemas);
- ***Communicating*** with research communities so that they can be made aware of both the need to apply standards in their digital scholarly practices but also be informed of the essential standards for their own fields.
- ***Training*** for researchers, by giving them access to complete frameworks so that they may acquire knowledge and know-how on standardized methodologies.

In order to apply these four principles, the SSK focuses on giving researchers access to standards in a meaningful way. That is why it is built around research scenarios. These scenarios are the core of the SSK because they aim at providing **contextual information** and relevant **examples** on how standards can be applied in a given research project. They intend to cover **most domains of the Humanities**, from Literature to Heritage science, including History, Social sciences, Linguistics, etc. They have been created and they are added to by domain experts, from **real life researcher-oriented use cases** (PARTHENOS, 2016), divided into different steps, and involving specific tasks. For that reason, the SSK can be considered as a **complete framework** showing concrete use of standards, rather than simply a catalog of resources, such as the DiRT Directory[4] or the CLARIN Standards Information System[5].

Each scenario within the SSK works like a high-level research guide for scholars. They are made up of successive steps or tasks, and can be followed as a complete process to solve a given problem with the most standardized means.

For each step, the appropriate resources to perform the given task are proposed, divided into two categories : the "general resources" that include the primary documentation and tools; and the "project-specific resources" that point to concrete use cases in which a similar task was accomplished. The material contained in these sections is of various kinds:

- the most important is the **state-of-the-art bibliography**, which includes all the documentation needed to carry out a given task. The bibliographical references are up-to-date and gathered within a Zotero library[6], which was specially created for this project.
- the SSK also offers more **technical resources**, such as stylesheets, code samples, software or services: a significant part of these resources has been brought together in a dedicated GitHub environment fed by expert partners of the PARTHENOS project.
- **Training materials** like tutorials.

[4] https://dirtdirectory.org/
[5] https://clarin.ids-mannheim.de/standards/
[6] https://www.zotero.org/groups/427927/parthenos-wp4/items

To represent and modelize the research scenarios, we use the **Text Encoding Initiative**[7]. Each scenario and each step is encoded in TEI documents that are linked together with referencing mechanisms. This choice was made in order to ensure that the scenarios and the steps can be easily extended, reused and customized. The data model allows scenario creators to modify the structure of their research scenarios on the fly, by creating, removing or reordering steps. As steps are considered as autonomous objects in the architecture, they can be used in several scenarios. Customisation mechanisms are added to make sure that the information displayed is linked to the context of the scenarios as much as possible, namely according to disciplines, research objects and techniques.

Bibliography

- Martin Hilbert *How much information is there in the "information society?"* Significance, 9 (4) 2012, 8–12.
- Bruno Latour *How Better to Register the Agency of Things: Ontology*, 2014. http://www.bruno-latour.fr/node/563
- Charles Percy Snow *The Two Cultures* London, Cambridge University Press, 1959
- DeNardis, Laura. *Opening Standards: The Global Politics of Interoperability*. Cambridge, Mass.: MIT Press, 2011.
- Laurent Romary, "Stabilizing knowledge through standards - A perspective for the humanities", *Going Digital: Evolutionary and Revolutionary Aspects of Digitization*, Science History Publications, 2011. ⟨inria-00438724⟩
- Laurent Romary, et al.. Standardization survival kit (Draft). [Research Report] Deliverable 4.1, Inria. 2016. <hal-01513531>
- Laurent Romary, et al.. Report on Standardization (draft). [Technical Report] Deliverable 4.2 Inria. 2017. <hal-01560563>

[7] The underlying data model of the SSK is in XML/TEI, and it is publicly available in a Github repository : https://github.com/ParthenosWP4/SSK