Standardization Survival Kit (Final)

Charles Riondet, Dorian Seillier, Lionel Tadjou, Laurent Romary

PARTHENOS

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

STANDARDIZATION SURVIVAL KIT

(Final)

Inria

CLARIN

KNAW

OEAW

SISMEL

(And Contributions from all WP4 partners)

HORIZON 2020 - INFRADEV-4-2014/2015:

Grant Agreement No. 654119

PARTHENOS

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

STANDARDIZATION SURVIVAL KIT (Final)

Deliverable Number	D4.3
Dissemination Level	Public
Delivery date	31 October 2018
Status	Final
Author(s)	Charles Riondet (Inria) Dorian Seillier (Inria) Lionel Tadjou (Inria) Laurent Romary (Inria) Reviewers: Mikel Sanz (CISC), Klaus Illmayer (OEAW), Sheena Bassett (PIN)

Project Acronym	PARTHENOS
Project Full title	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
Grant Agreement nr.	654119

Deliverable/Document Information

Deliverable nr/title	D4.3 Standardization Survival Kit (final)
Document title	Standardization Survival Kit (final)
Author(s)	Charles Riondet (Inria)
	Dorian Seillier (Inria)
	Lionel Tadjou (Inria)
	Laurent Romary (Inria)
	Reviewers: Mikel Sanz (CISC), Klaus Illmayer (OEAW), Sheena Bassett (PIN)
Dissemination level/distribution	Public

Document History

Version/date	Changes/approval	Author/Approved by
V 0.1 30.10.18	First edited version	Charles Riondet
Final		Sheena Bassett

Table of contents

Executive Summary

This deliverable finalises the release of the Standardization Survival Kit (SSK), a web platform aiming at giving access to standards and best practices in a meaningful way, by the mediation of research scenarios. This report also compiles all the documentation related to the SSK, composed of three parts: Users documentation, data model specifications and technical features.

Section 1 provides the URLs for the SSK with explanations for each of these (i.e. hosting, development and data on GitHub, documentation). Section 2 presents the rationale behind the SSK arising from the need to promote the use of standards in Humanities research. The aim has been to build an **easy-to-use** online and collaborative platform with a **user-friendly** design. The idea of having general, end-to-end scenarios to help researchers carry out their project by following best practices and clear methods in their area of expertise is the most important design principle for the SSK. The user testing is also described with the issues that this highlighted and the long term sustainability of the SSK is discussed.

Section 3 addresses the actual components that make up the SSK which is a web platform built on three main layers nested within each other following a specific order: Research scenarios, steps and resources. How users can set up their own scenarios is explained in detail. Section 4 describes the SSK data model. This consists of the structure of the scenarios and the steps, their content, the taxonomies, and some advanced features.

The SSK data model is presented in Section 5. Users can contribute scenarios using TEI on GitHub. The main elements and common attributes are described along with the resources (Zotero and GitHub). Further details are also supplied about scenarios and steps. Section 6 covers the technical infrastructure required to install and deploy the SSK. The final section 7 provides some conclusions about the future development of the SSK and the reports concludes with references.

Introduction

To support the digital evolution within Social Sciences and Humanities research, it is necessary to stabilize knowledge on standards and research good practices. The goal of the **Standardization Survival Kit (SSK)**, developed within the PARTHENOS project, is to accompany researchers along this route, giving access to standards and best practices in a meaningful way, by the mediation of research scenarios. A research scenario is a (digital) workflow practiced by researchers, that can be repeatedly applied to a task that will help to gain material or insights in view of a research question. These scenarios are at the core of the SSK, as they embed resources with contextual information and relevant examples on standardized processes and methods in a research context. The SSK is an open tool where users are able to publish new scenarios or adapt existing ones. These scenarios can be seen as a living memory of what should be the best research practices in a given community, made accessible and reusable for other researchers.

1. Releasing the Standardization Survival Kit

The Standardization Survival Kit can be accessed at the following URLs:

- http://ssk-application.parthenos.d4science.org/ssk/ (development version)
- https://ssk.parthenos-project.eu/ssk/# (Public version on PARTHENOS website)
- http://ssk.huma-num.fr (Sustainable version, See the *Sustainability plan* section)

These three versions are exact replicas, deployed from the same sources and automatically updated with the latest data, via webhooks harvesting the SSK GitHub repository.

The sources of the SSK are hosted on GitHub:

- The application on the *dev* branch:
  https://github.com/ParthenosWP4/SSK/tree/dev
- The data on the *master* branch:
  https://github.com/ParthenosWP4/SSK/tree/master
-

The documentation of the SSK (for users and developers) is available here:

- http://ssk.readthedocs.io

2. The Standardization Survival Kit: presentation

2.1 Why standards after all?

The main issue in defining a policy about standards is to understand what they actually are. In the context of research, standards usually take the form of documents informing about practices, protocols, artefact characteristics or data formats that can be used as reference for two parties working in the same field of activity to be able to produce comparable (or interoperable) results. This will also foster innovative, cross-disciplinary research paths, and eventually contribute to bridge the gap between the different cultures that are represented in the wide landscape of the Arts, Humanities and Cultural Heritage studies.

Standards are usually published by standardization organisations (such as ISO, W3C or the TEI Consortium) which ensure that the following three requirements for standards are actually fulfilled:

- Expression of a consensus: the standard should reflect the expertise of a wide (possibly international) group of experts in the field
- Publication: the standard should be accessible to anyone who wants to know its content
- Maintenance: the standard is updated, replaced or deprecated depending on the evolution of the corresponding technical field.

Standards are not regulations. There is no obligation to follow them for research 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 attitude the scholarly communities could or should adopt with regard to specific standards.

The preceding characteristics outlined for standards put a strong emphasis on the role of communities of practice and the corresponding bodies that represent them. Ideally, a good standard reflects the work of the relevant community and is maintained by the appropriate body. This is exactly the case of the Text Encoding Initiative (Burnard 2000) with respect to text representation standards and, to a lesser extent, of EAD (Encoded Archival Description), whose maintenance is taken up by the Library of Congress with support of the Society of American Archivists (Library of Congress n.d.).

Because there is no obligation to use a given standard, it is essential to provide potential users with:

1. awareness about the appropriate standards and the interest to adopt them,
2. the cognitive tools to help them identify the optimal use of standards through selection and possibly customization of a reference portfolio.

The experience gained within the various communities and infrastructure represented in PARTHENOS that have been in the need of adopting existing standards, is that there is always an initial phase during which scholars should be made aware of some core standards that are systematically related to the definition of interoperable digital objects. This is basically what has lead us to identify the notion of Standardization Survival Kit

(SSK). In the table below, for instance, we can see a first group of standards without which it is more or less impossible to deal with digital content in a proper way.

ISO 639 series	Codes for the representation of languages and language families
ISO 15924	Codes for the representation of scripts
ISO 3166	Codes for the representation of country names
IETF BCP 47	Standard for encoding linguistic content, combining ISO 639, ISO 15924 and ISO 3166
ISO 10646	Universal encoding of characters (unicode)
ISO 8601	Representation of dates and times
Extensible Markup Language (XML)	Provides the basic technical concept related to XML documents

The concept of SSK goes far beyond these baseline examples and aims at covering reference digital scenarios in the Arts and Humanities: a role of the SSK is to help communities participate in standardization activities where they exist, or at least document and spread the best practices as de facto standardized guidelines. Such a strategy will also contribute to the actual stabilization of existing conceptual and technical knowledge within ongoing projects, and provide a channel for the wider dissemination of the corresponding results.

2.2 The SSK: a toolkit for Humanities scholars

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 a reference for scholars who want to find out more about their role and content. This relates to the specific provision of bibliographic sources, available documentation, specific targeted introductions, as well as providing prototypical examples which can serve as models for similar work, possibly made available through focused Virtual Research Environments within the PARTHENOS infrastructure (Romary et al. 2017);

- **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 potentially cover **all the domains of the Humanities**, from Literature to Heritage science, including History, Social sciences, Linguistics, etc.They have been created and they are added to the SSK by domain experts, from **real life researcher-oriented use cases)**, divided into different steps, and involving specific tasks (Romary et al. 2016).

These scenarios can be seen as a living memory of what should be the best research practices in a given community, made accessible and reusable for other researchers wishing to carry out a similar project but unfamiliar with the recommended tools, formats, methods to use, etc. For that reason, the SSK can be considered as a **complete framework** showing concrete use of standards, rather than simply a catalogue of resources.

2.3 Design principles & features of the SSK

From the very start of the project (Romary et al. 2016), the aim has been to build an **easy-to-use** online and collaborative platform with a **user-friendly** design. The idea of having general, end-to-end scenarios to help researchers carry out their project by following best practices and clear methods in their area of expertise is the most important design principle for the SSK.

The second principle is to add **context**: a "story-telling" approach to the use of standards in the Humanities and Social Sciences. The goal is to avoid providing yet another catalogue of resources, and offer instead contextual, **activity-based information** on how to use standards for researchers who are unfamiliar with them, but could see how they are used and what workflow they help achieve by following a scenario.

With these principles in mind, the SSK should enable the user, to perform two main actions:

1. **Consult and follow the guidelines expressed in the scenarios** you are interested in for your project. Finding the most relevant ones should be easy since the navigation relies on strong taxonomies covering the different aspects of research: the disciplines, the type of techniques and objects involved, the concrete activities carried out, the standards needed.

2. **Propose new scenarios** of your own by following a predefined model, with the possibility of both adding new content (steps as well as resources) and reusing existing content (to avoid duplication if a general step is already available in another scenario).

The first feature is fully operational. It was tested for the first time in April 2018, and iterations with test users have contributed to improve the **information readability** and **attractiveness,** in particular the exploration and search of scenarios (see Figure 1).
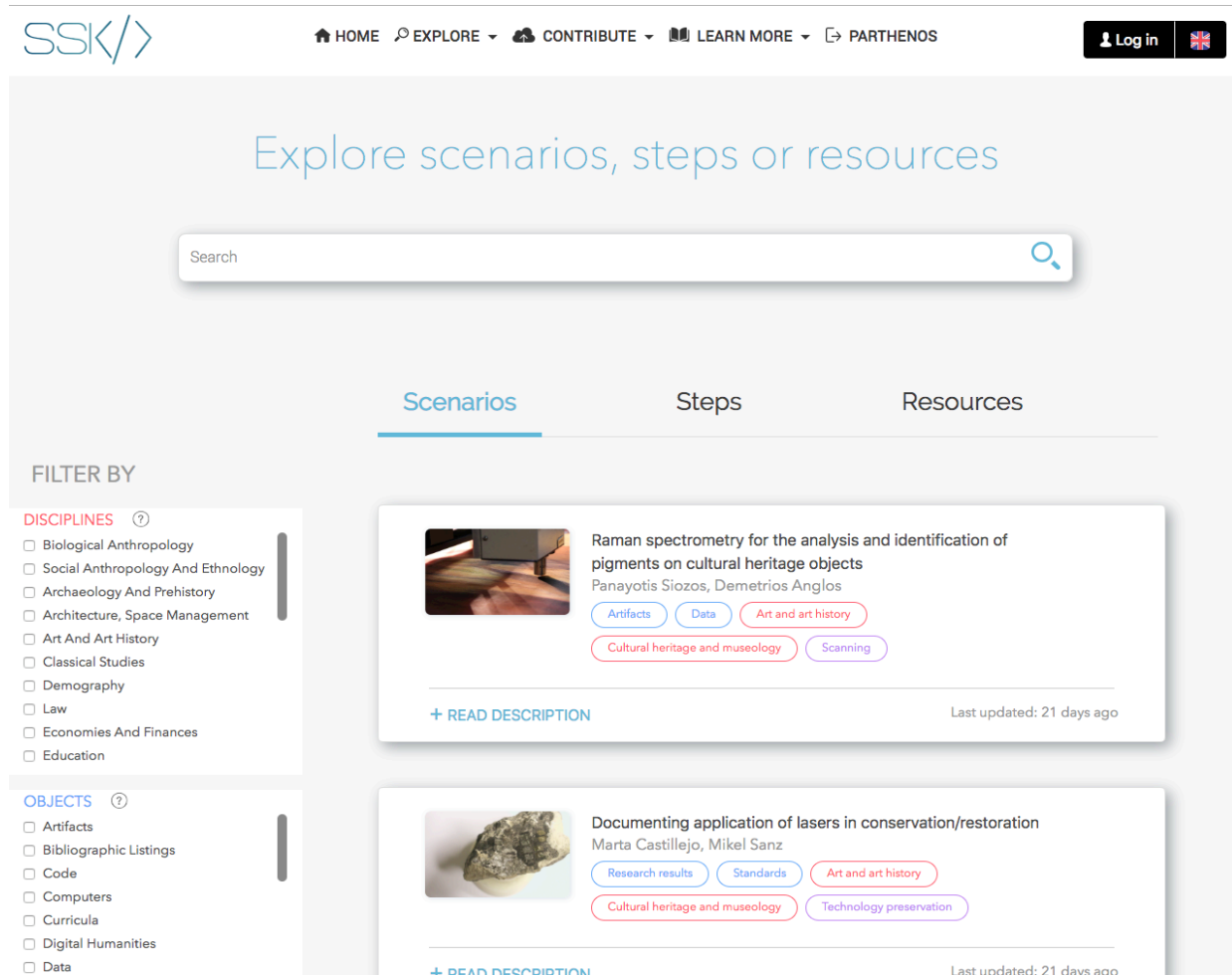
Figure 1: Navigation page for scenarios, steps and resources

The work on the second feature, allowing the user to contribute, is still ongoing. It is possible to create research scenarios with the SSK underlying data model, the Text Encoding Initiative, or TEI (see the dedicated section for more information). However, as we are aware that this solution requires some technical knowledge, a user-friendly interface is currently under development and should be released during the first trimester 2019.

Anyone who has registered and agreed to the "**Terms of use**" clearly stated will be able to contribute. This option has been chosen due to the difficulties of setting up some kind of editorial board in charge of reviewing any scenarios submitted. The **quality check** of the contributions should come from the very strict model that has to be followed in the scenario creation process. In addition, by contributing to the SSK, the user accepts to be

visible and citable as an author, to be responsible for the work that he/she decides to share with others.

Why would you, as a researcher, want to contribute to the SSK? There are three main reasons:

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

2.4 User Testing

At the end of March 2018, a first usability test was organized and carried out by the PARTHENOS Work Package 2 and 4 to assess the user experience of the SSK in its beta version (Seillier 2018).

2.4.1 Why a test?

A usability test is one of the most popular methods to assess the user experience of a system in an iterative way. It consists in putting the user in front of the system in order to watch his or her behaviour, reactions and overall performance while he / she is carrying out predefined tasks on the interface. The data gathered are precious information about the strengths and weaknesses of the assessed system, as well as the user experience. It is an iterative process, so it can be done during several phases of the design: from the early mock-ups to the final product.

The general goals of the test were:

- to determine the degree of usability of the tool in its current state
- to identify the missing elements and bugs
- to evaluate the user experience of a first contact with the tool
- to see if the SSK met the needs of the users.

2.4.2 Settings & methodology

2.4.2.1 The participants

For the purpose of the test, the targeted audience was made essentially of researchers coming from the different "domains" as defined within PARTHENOS: Studies of the Past, Language-related studies, Heritage, applied disciplines and archaeology, Social Sciences.

After sending multiple invitations through different channels, ten people accepted to be participants in live or remote sessions. None of them were related with PARTHENOS in anyway whatsoever; they did not know anything about the SSK before the test. For the purpose of the test, it was important to have the widest variety of users in terms of age, skills, nationality, research topics, etc. Unfortunately, the lack of people willing to participate resulted in a small sample of individuals coming largely from the "Language-related studies" (see Figure 2).
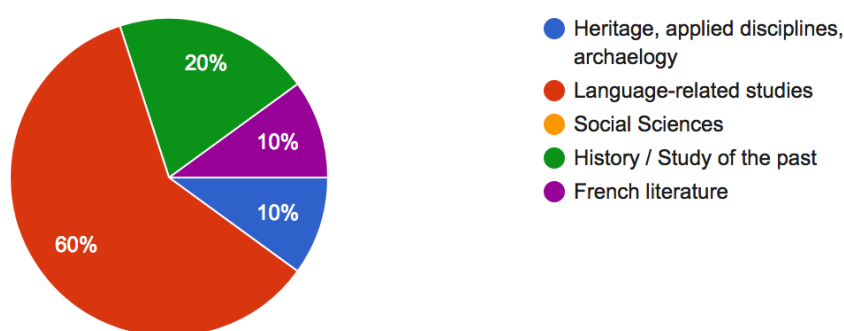
Figure 2: Participant domain percentages

2.4.2.2 Running the test

Since not all participants were able to perform the test face-to-face, participation was also possible remotely. All participants had agreed to sign a consent form allowing the recording of the session, that is to say the screen of the interface as they navigated through, the webcam to record any facial expression and the sound for the comments / remarks made aloud during the test.

The session itself was split into two parts:
- the user tasks, which the participants were asked to perform on the interface while being observed (between 30 and 40 minutes long)

- a sentence completion questionnaire just after the test, which they needed to answer (see it here).

During the test, the participants were free to ask as many questions as necessary, but the moderators would remain as passive as possible in order not to influence them.

2.4.2.3 The user tasks

1. "*Find a research of your academic field or of your interest and follow the steps.*": the goal of this preliminary task was to see how the participant got in touch with the website and the concept of a « research scenario » (on-boarding aspect).
2. "*You want to make a digital edition of your research corpus. First you need some guidelines on how you can digitize your source. Use the SSK to find the way to digitize fragile documents.*": to assess how the user navigate through a scenario and retrieve a specific piece of information.
3. "*You are looking for the definition of « Modelling » in the SSK. Use the glossary to learn more about this concept.*": to make them browse through the glossary and see how they react to the terminologies used (see below for more explanations on the vocabularies used).
4. "*You want to learn more about Dublin Core, and you look after a tutorial in the SSK.*": to see how they handle the tabs and search bar.

2.4.3 User Experience issues

Overall, the SSK was positively received by the participants. It was perceived as a great initiative and its usefulness seemed to be well appreciated by the testers. However, one of the main issues concerned the understanding of the major utility of the tool and some of its features. A lot of testers had trouble figuring out what the SSK was meant for and what they could exactly do (and not do) with it. The fact that the SSK is meant for providing guidelines and best research practices, and not actually a space where you can apply the described methods on your data, was not fully understood.

For less experienced participants in Digital Humanities, handling the SSK was a struggle. They didn't feel at ease because the content was very new to them, as well as the vocabulary and terminology. They seem to need more information and guidance to find their way round with the tool. On the other hand, more experienced participants wanted

more of the SSK than it has now to offer: more details, more resources and more options, like using the offered standards directly within the SSK.

Almost every participant left the homepage very quickly once the test had started without really looking at it and never returned. The main focus of the participants was the page with the scenarios. The homepage seemed to lack attractiveness, the texts were read pretty quickly, even not read at all. As a result, the homepage has been totally redesigned, see figure 3.
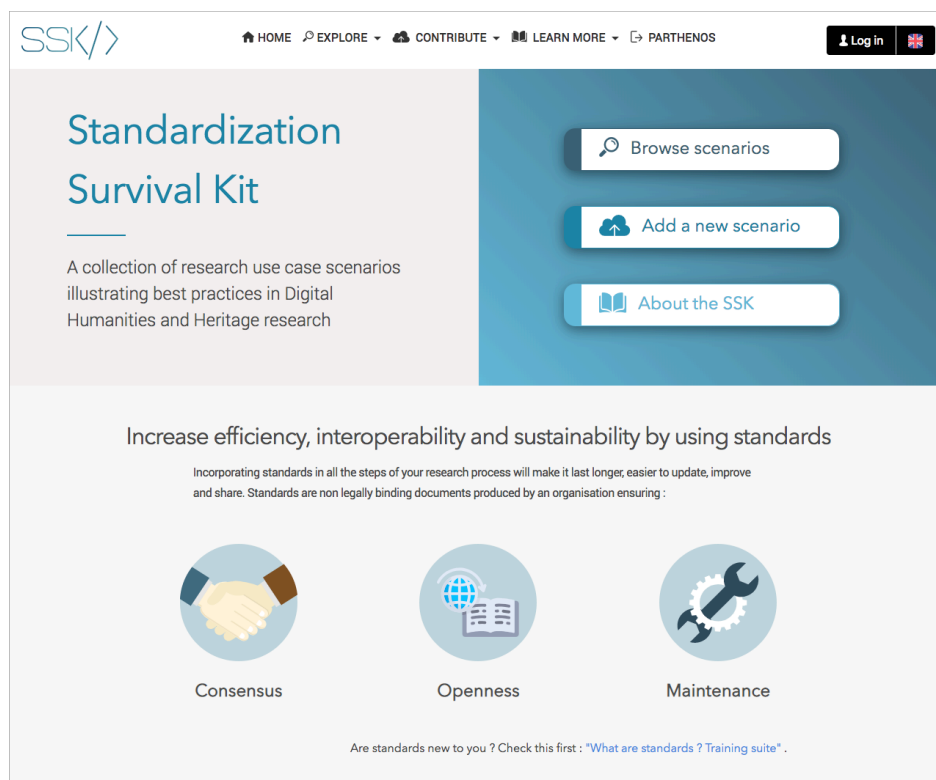
Figure 3: New homepage based on the test's results

The glossary was received very positively, but we needed to clarify where the ontologies come from, why we use them, and also justify our classification. These questions are answered in the sections below, as well as in the SSK online documentation: http://ssk.readthedocs.io

2.5 Upcoming activities

Ensuring the **sustainability** of the SSK beyond the lifetime of the PARTHENOS Project is of primary importance because it is an open tool, meant to be used and maintained by the

research communities at large. It is crucial to create and maintain an **active and motivated users' community**. To do so, the first step is to make the SSK familiar and show its usefulness. In design, we call ***user on-boarding*** (Cascaes Cardoso 2017) the process of getting the user to understand the key principles and features of a tool. It has been proven that it is crucial to ensure the success of a tool in the long run. In our case, we advocate a ***multimodal on-boarding***, which includes two parts:

1) face-to-face meetings on the form of use case development workshops, very important to engage more deeply the users with the SSK

We plan to organize three workshops, in three different locations - Vienna (Austria), Tours (France), Liege (Belgium), with three different focuses (Linguistics, Archaeology and Medieval studies), with the following objectives:

- Promoting the use of standards and best practices in the Arts and Humanities;
- Raising awareness of the SSK, within the PARTHENOS and DARIAH communities and beyond;
- Demonstrating how this tool can be useful for both learning and teaching digital best practices;
- Encouraging researchers to enhance existing research scenarios and create new ones.

2) the creation of digital educational materials: a presentation cartoon**.**

The **usability testing** of the SSK, run between March and April 2018, revealed that the main purpose of the tool and its features were not fully understood. The homepage was not appealing enough to get the user to read it in detail which led eventually to confusion and misinterpretation of what the SSK had to offer. To avoid that confusion, and to foster the use of the tool, we want to include a short cartoon on the homepage that would show its usefulness in a simple and funny way. It will be part of a more global onboarding experience, with tutorials and highlights of the user interface. The resulting cartoon (inspired by a comic, produced in the context of PARTHENOS, that successfully addressed the issue of standards in general, read it here) should be both short and instructive. The creation and production of a cartoon requires several steps, some of them

would be assumed by an external production company. The partners will elaborate the story and the message we want to convey, and write the script. A production firm would then take care of graphic design, storyboard drawings, production, editing and mixing of the cartoon.

2.6 Sustainability planning

The SSK code is open-source and under a GPL licence. Moreover, the entire underlying data of the scenarios and steps, are described in TEI and hosted on GitHub under the licence CC-BY (reuse and modifications allowed, citation of authors) and the bibliographical resources are part of a Zotero open library. Therefore, the sustainability of the application is made easier. However, ensuring the availability of the SSK on the long run requires two main strands of actions: technical and intellectual.

On the technical side, the SSK has to be available to the public for the long term. To ensure this long lifetime, we put in place a double hosting on the PARTHENOS D4Science platform (http://https://ssk-application.parthenos.d4science.org/ssk/ redirecting to http://ssk.parthenos-project.eu) and, anticipating the fact that the PARTHENOS project draws to a close in 2019, the SSK is also on the Huma-Num platform (http://ssk.huma-num.fr), the French national infrastructure for Digital Humanities, and National coordinator for DARIAH in France.

Most importantly, the DARIAH communities are willing to take over SSK's everyday intellectual maintenance, both on an institutional and content level. The DARIAH Working Group Guidelines and Standards, chaired by Klaus Illmayer (OEAW, CLARIAH-AT), Marie Puren (LARHRA, DARIAH-FR) and Charles Riondet (Inria, DARIAH-FR), aims at being a place where the SSK can be administered. It seems also very natural that the SSK becomes a place where methodological resources and best practices created or gathered by the various DARIAH Working Groups can be presented and archived. Doing so, we also pave the way for the SSK to be integrated into the DARIAH Marketplace and used by Arts, Humanities and Social Sciences researchers in the context of the European Open Science Cloud (EOSC).

3. SSK components

The SSK is a web platform built on three main layers nested within each other following a specific order: Research scenarios, steps and resources.

3.1 Scenarios < Steps < Resources

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, which was specially created for this project. This choice was made to ease the resource selection process and to allow for a collaborative watch and curation of relevant information. When the resource is available online, a direct link is provided; otherwise, the user is given all the necessary metadata.
- The SSK also offers the possibility to point to more **technical resources**, such as stylesheets, code samples, software or services.
- **Training materials** such as tutorials.

3.2 How to create a scenario for the SSK

The following instructions help users create a scenario for the SSK. These instructions are themselves conceived as a "scenario", a step-by-step tutorial. First, contributors should be aware that:

- they can submit their new scenarios directly in TEI — see the dedicated section — and upload them on the SSK GitHub (http://github.com/ParthenosWP4/SSK), or by using the dedicated SSK contribution workspace (still work in progress).
- scenarios and steps follow the same data model. The difference is that a scenario points to a set of steps whereas a step points to a set of external resources.

- It is possible to create a scenario in any language but ideally, we would advocate for at least bilingual scenarios, with an additional version in English, in order to make it more visible. The forthcoming user interface will implement such behaviour and in TEI, it is possible to translate the prose, by duplicating the elements `<head>`, `<desc>` or `<term>`, and adding `@xml:lang` attributes.

3.2.1 Choose explicit titles starting with a verb or a gerund

The choice of good titles for a scenario and its steps is crucial (Leahy 1992). They will be the entry point for the readers, who need to understand at first reading the scope and the aim of a scenario. To ensure that the message will be driven home, we advise to build titles:

- that start with a verb or a gerund, describing the process (for a scenario) or the action (for a step) that the user will read about;
- with a number of characters comprised between 10 and 100.

3.2.2 Describing a scenario or a step

The description of the scenario and its steps is the longer text that the contributor has to provide.

- For a **scenario**, it should explain the scientific problematic and describe the solution put in place.
- For a **step**, It should describe the purpose of the action, how it relates with the previous ones and give an overall presentation of the different kind of methods and tools the resources would point to (Universitat Autònoma de Barcelona n.d.).
- For both, it is important to extend the acronyms cited and to briefly present the projects mentioned.

The form of this text should respect the following constraints:

- It shouldn't exceed 1500 characters (but should not be too short either).
- It is possible to point to external links. In TEI, use the following code:
  `<ref target="//url here//">text of the link</ref>`
- Lists are also available. The TEI elements are <list> and <item>

3.2.3 Associate keywords to the scenario or the step

In order to enhance discoverability and search relevance, the SSK resources are described with a set of controlled vocabularies, particularly created for describing Humanities research. They are:

- Research activities, taken from Tadirah;
- Research techniques, taken from Tadirah;
- Research objects, taken from Tadirah;
- Standards, taken from the SSK Standard Knowledge base (supported by DARIAH-IT);
- Disciplines, taken from aureHAL (Humanities and Social Sciences subcategory)

When editing the description of a scenario, the available keywords are:

- Research techniques
- Research objects
- Standards
- Disciplines

For the steps, the most important keyword is the **Activity**, that should be unique for each step. It also possible to pick some **techniques**, **objects** and **standards**. For each keyword type, we recommend to choose between **1 and 4** terms.

3.2.4 Choose an illustration for the scenario

The illustration must closely relate with the purpose of the scenario, i.e. not only with the discipline or the period studied.

- Screenshots are accepted.
- Landscape orientation image are recommended
- Maximum size: 2 MB
- Accepted formats: png or jpg

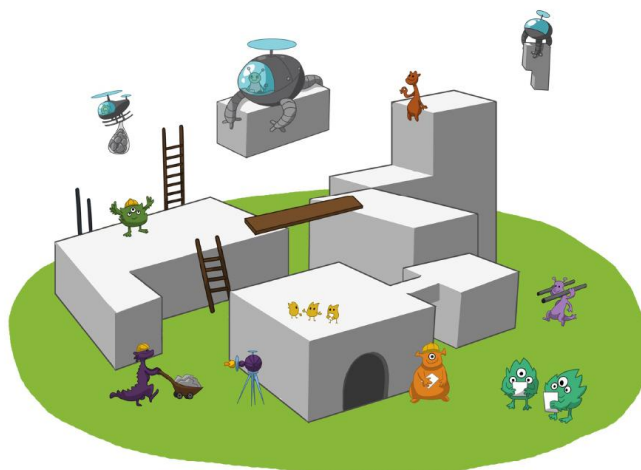It must be published under the licence CC-BY or CC-0. See for instance, Unsplash, a gallery of free images and photos.

Figure 4: Illustrate a scenario with a relevant image

3.2.5 Identify relevant resources

Identifying state of the art references is a prerequisite before actually adding the resources to the steps. When we are talking about resources, we mean a standardized tool, service or document helpful for the task completion.

These take the form of a digital object: a webpage, a journal article referenced in an online catalogue or an archive, a code repository, a blog, etc.

The resources should be separated into several groups:

- First, general resources like standard specifications
- Second, project-related resources, i.e. how the standards are used in a real research project.

3.2.6 Link the resources to the step

There are different ways to link resources to a step (TEI: <ref> element). The one we favour is the recording of the resource metadata in the dedicated SSK Zotero Library (see here).

1. To populate it, a Zotero account is necessary (create it here) as well as a membership in the SSK group (apply here). The SSK library is organized in collections and sub-collections, by domains or standards. To learn more about how to use Zotero, many tutorial and learning resources are available here.

2. Each group of resources (general and project specific) should be gathered: one group for the general resources and one group for each project. In the TEI, these groups are represented by <linkGrp> elements.

3. The references added to the Zotero Library are linked to the step with the help of their Zotero key, i.e. the last part of the URL of the resource record on the Zotero

website. For instance, in the following example, the key is 4B62GJ5I:

https://www.zotero.org/groups/427927/ssk-parthenos/items/itemKey/4B62GJ5I

In TEI, the Zotero key should be used like this:

```
<ref type="zotero" key="4B62GJ5I"/>
```

4. It is possible to put directly the URL of a GitHub repository or a document stored in HAL, and skip the Zotero part (the metadata would be fetched directly via the APIs)

5. It is possible and recommended to add a description of the resource, in addition of the Zotero metadata. This description should make the link between the resource and the SSK step that references it. In TEI, the element to use is <desc>, inside a <ref> element.

3.2.7 Advanced SSK functions (1): customize a step or a scenario

The SSK is adaptable by nature and contributors don't have to start from scratch with their scenario. It is possible to create a scenario with existing steps as basis. But if the content of the step doesn't exactly fit, it is also possible to modify it, by updating the initial step (but with care), or, more safely, directly in the new scenario.

In TEI, the update of element is made with the help of the attribute @mode. See more in the section dedicated to the TEI model.

3.2.8 Advanced SSK functions (2): link scenarios

Linking scenarios together, or in other words, include a scenario (entirely or partially) into another is an interesting possibility when a scenario is a precondition or the continuation of another one. For instance, a scenario related to the preservation of 3D models can be preceded by a scenario explaining how to create such models.

The most common use cases are the following:

- Add a prerequisite scenario (as a first step)

- Associate a scenario that can be the follow-up of the current (as a last step)

- Insert a scenario (totally or partially) inside the current scenario, with the use of parameters that allows the user to choose which step of the external scenario should be included.

4   The SSK data model

4.1 Contribute to the SSK on GitHub with the TEI

Users willing to create scenarios in TEI should follow the following instructions:

- Download or fork the SSK data repository in GitHub It is necessary to have an account on GitHub: https://github.com/ParthenosWP4/SSK/tree/master/ (NB: to fork a repository, a GitHub user account is necessary);

- Create your files with your favourite XML editor. Don't forget to validate them against the SSK schema (see above);

- To publish scenarios on the SSK, the TEI files need to be in the *scenarios* and *steps* folders;

- Users with a GitHub account can make a pull request to ask for the update of the repository. Users without an account can contact the SSK team at ssk [at] inria [dot] fr.

4.2 Main elements

The main elements of the SSK TEI are:

- `<listEvent>` (list of events) contains a list of descriptions, each of which provides information about an identifiable event.

- `<event>` contains data relating to any kind of significant event associated with a person, place, or organization.

- `<linkGrp>` (link group) defines a collection of associations or hypertextual links.

- `<ref>` (reference) defines a reference to another location, possibly modified by additional text or comment.

A scenario is a list of events (`<listEvent>`), each scenario step is an event (`<event>`).

The `<event>` element is the core of SSK scenarios. It contains the full description of scenario step:

- a `<label>` that contains any label or heading used to identify part of a text, typically but not exclusively in a list or glossary.

- a `<desc>` that contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

- some descriptive terms following controlled vocabularies: `<term>` elements that contains a single-word, multi-word, or symbolic designation which is regarded as a technical term.
- bibliographical references: `<ref>` (reference) defines a reference to another location, possibly modified by additional text or comment.

Events are stored in external files, allowing them to be used in different scenarios. Scenario files gather `<event>` in a `<listEvent>` element, by referencing them with a `@ref` attribute. It is however possible to modify the content of the event called in a scenario, using the attribute `@mode` (see below)

The resources consist of bibliographical references. They are gathered in `<linkGrp>` elements. They are of two types: general resources and project specific. `<linkGrp>` elements can be repeated (one per project). At each level, the elements are enriched with keywords that can be term or XML attributes.

4.3 Common attributes

The attributes used by all the elements are:

- @xml:lang. This attribute is mandatory in all the content elements, namely <label>, <desc>, <term>. The authoritative list of registered language subtags is maintained by IANA and is available at http://www.iana.org/assignments/language-subtag-registry:

```
<label xml:lang="en">Create associated documentation</label>
<label xml:lang="fr">Création de la documentation associée</label>
```

- @mode, available for all content elements. the mode is used in scenario files to allow for modification of the imported steps:

```
<event xml:id="s1" type="step" ref="step_EaXswO_290517">
 <head mode="change">
  <!-- The new <Label> replace the initial step label, in the
   declared language-->
  <label mode="change" xml:lang="jp">...</label>
  <!-- A new term is added to the step -->
```

22

```
  <term mode="add"/>
 </head>
 ...
</event>
```

The @type attribute is required in most elements. See below for details.

4.4 The resources

4.4.1 Zotero

The resources to be presented in the SSK are preferably stored in a Zotero database, accessible here. The Zotero database fields required by the SSK are:

- *Item type*: The item type is most of the time identified by Zotero but it's important to check it. The most used item types are:
    - webpage
    - blogpost
    - journal article
    - book section
    - book
    - presentation
- *Title:* The title of the resource
- *Author*: The author of the resource
- *Date*: The date of the resource
- *Url*: The URL of the resource
- *Language*: The language of the resource
- *Source*:
    - For web pages: website title
    - For blog posts: blog title
    - For journal articles, books, book sections, documents: library catalogue

NB: A short description of the resource should be provided when possible. In Zotero, the appropriate field is abstract, but it is also possible to add this description in the TEI, with a <desc> element, as shown in the following example:

```
<linkGrp type="generalResources">
<ref source="zotero" subtype="book" target="PM5P3JDB" type="tutorial">
 <desc xml:lang="en" type="resourceDesc">This booklet is
  intended as an introductory textbook for students and
  end-users interested in knowing more about the exciting
  developments in this high-tech area of conservation and
  conservation science. Their teachers are invited to use the
  texts and photographic materials for educational purposes,
  while the conservation scientist might appreciate the short
  reviews of applications and of the science underlying the
  described processes.</desc>
<term key="Laser cleaning" type="tutorial"/>
</ref>
<linkGrp/>
```

4.4.2 GitHub

It is also possible to point to a GitHub user or repository. In this case, the information that the SSK shows (via the API) are:

- For a GitHub User:
  - name
  - html_url (i.e. the URL of the profile page)
  - bio
  - updated_at
  - type
  - avatar_url
- For a GitHub repository
  - owner (NB: a GitHub user)
  - full_name
  - html_url
  - description

NB: A dedicated GitHub repository has been set for projects supported or maintained by Parthenos (https://github.com/ParthenosWP4/standardsLibrary).

4.5 Detailed structure

Scenarios and steps are represented in different files. This choice has been made to facilitate the use of a step in several scenarios, with or without modifications.

Files naming conventions are the following:

For scenarios:

1. `sc` for scenario
2. an underscore: `_`
3. a condensed title of the scenario in camel case: `myScenarioTitle`

example: `sc_myScenarioTitle.xml`

For steps:

1. the string `step`
2. an underscore: `_`
3. the initials of the step name, with the liaison words in lower case, and the meaningful words in upper case. For example, if a step title is: `Searching for a relevant step title`, it would give: `SfaRST`.
4. an underscore and the date (optional)

example: `step_SfaRST_10092018.xml`

4.5.1 Scenarios and steps structure

4.5.1.1 Scenarios

The scenario is represented by the element `<listEvent>`, containing a set of event elements that reference external TEI files.

Header

The structure of the Scenario header is as follows:

```
<TEI type="scenario" xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
  <fileDesc>
   <titleStmt>
    <title>
    <!-- Title of the tei document, not title of the scenario -->
    </title>
```

```
    <author>
     <persName>...</persName>
     <affiliation>...</affiliation>
    </author>
    <sponsor>PARTHENOS</sponsor>
   </titleStmt>
   <publicationStmt>
    <authority>...</authority>
    <availability>
     <licence target="http://creativecommons.org/licenses/by/4.0/">
      <p>The Creative Commons Attribution 4.0 Unported
      (CC BY 4.0) Licence applies to this document.</p>
     </licence>
    </availability>
   </publicationStmt>
   <sourceDesc>
    <p>Created from scratch</p>
   </sourceDesc>
  </fileDesc>
  <revisionDesc>
   <change>
   <!-- Only for major changes: addition of an author, of a step, etc. -->
   </change>
  </revisionDesc>
 </teiHeader>
 ...
</TEI>
```

The scenario header includes the following data elements:

- the title of the document (which is not the title of the scenario)
- the authors of the scenarios
- the major modifications

Structure

In a scenario file, `<event>` elements are used as pointers to link to full event elements stored in external files.

```
<listEvent>
<event xml:id="s1" type="step" ref="step_EaXswO_290517"/>
<event xml:id="s2" type="step" ref="step_Eprimrf_300517"/>
<event xml:id="s3" type="step" ref="step_Cad_300517"/>
<event xml:id="s4" type="step" ref="step_Tdats_300517"/>
<event xml:id="s5" type="step" ref="step_Sapditnf_300517"/>
</listEvent>
```

It is also possible to refer to another scenario, that will be entirely (or partially by using parameters - see below) included in the described scenario.

```
<listEvent>
<event type="scenario" ref="SSK_digitization.xml"/>
<event xml:id="s1" type="step" ref="step_KedKep_170717"/>
...
</listEvent>
```

It is possible to modify the content of an existing step directly in the scenario file. See the advanced features for more information.

4.5.1.2 Steps

Step files record the full description of the scenario step. Several elements have the same meaning and behaviour than those in scenario files. The main difference is the content of the `<event>` element.

Header

The structure of the step header is as follows:

```
<TEI type="step" xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
<fileDesc>
<titleStmt>
<title>
<!-- title of the file, not title of the step -->
</title>
<author>
<persName>Charles Riondet</persName>
<affiliation>Inria</affiliation>
```

```
    </author>
  </titleStmt>
  <publicationStmt>
   <authority>Parthenos</authority>
   <availability>
    <licence target="http://creativecommons.org/licenses/by/4.0/">
     <p>The Creative Commons Attribution 4.0 Unported
        (CC BY 4.0) Licence applies to this document.</p>
    </licence>
   </availability>
  </publicationStmt>
  <sourceDesc>
   <p>Created from scratch</p>
  </sourceDesc>
 </fileDesc>
 <revisionDesc>
  <change/>
 </revisionDesc>
</teiHeader> ...
</TEI>
```

The step header includes the following data elements:

- the title of the document
- the author of the step
- the major modification

Structure

The main components of a `<event>` element are the description of the event, and the resources related to it. The description is recorded in the elements `head` (see below) and desc and the resources are contained by one or several `<linkGrp>`.

4.5.2 Content of scenarios and steps

5.5.2.1 <head> element

The TEI `<head>` element records the title of a scenario or a step.

The attribute `@xml:lang` is mandatory. The element `<head>` can be repeated to give as many translated versions as possible.

```
<label xml:lang="en">Create associated documentation</label>
<label xml:lang="fr">Création de la documentation associée</label>
```

4.5.2.2 <desc> element

The element `<desc>` is used in two ways for the description of the scenarios and the steps. The distinction is made with the attribute `@type`.

- When the value of `@type` is definition, the content of `<desc>` is a short text describing the scenario or the step
- When the value of `@type` is term, the content of `<desc>` is a set of term elements

4.5.2.3 <term> element

`<term>` elements are used to tag the scenarios, the steps and the resources, according to the SSK taxonomies, that are:

- Tadirah activities, objects and techniques
- the DARIAH-IT Standard Knowledge base
- aureHAL disciplines.

4.5.2.3.1 Functioning

These taxonomies are declared with the attributes `@type` and `@source`. The attributes of `<term>` elements are:

- The `@type` attribute gives information about the kind of term used. Its values are:
  - *standard*: the `@key` gives the id of a standard referenced in the **SSK standard Knowledge base;**
  - activity: the value of `@key` is taken from the **Tadirah** ontology, research activities section;
  - *object*: the value of `@key` is taken from the **Tadirah** ontology, research objects section;
  - *technique*: the value of `@key` is taken from the **Tadirah** ontology, research techniques section;
  - discipline, taken from the **aureHAL** taxonomy.
- The `@source` attribute sets a reference link for the taxonomy.

The `@key` attribute gives either an URI when the label of the term can be taken from or directly a label.

4.5.3.1.2 Taxonomies

Tadirah activities

The activities must be chosen in the following list (only pick between the second level values):

- Capture
  - Conversion
  - Data Recognition
  - Discovering
  - Gathering
  - Imaging
  - Recording
  - Transcription
- Creation
  - Designing
  - Programming
  - Translation
  - Web development
  - Writing
- Enrichment
  - Annotating
  - Clean up
  - Editing
- Analysis
  - Content Analysis
  - Network Analysis
  - Relational Analysis
  - Spatial Analysis
  - Structural Analysis
  - Stylistic Analysis
  - Visualization

- Interpretation
    - Contextualizing
    - Modeling
    - Theorizing
- Storage
    - Archiving
    - Identifying
    - Organizing
    - Preservation
- Dissemination
    - Collaboration
    - Commenting
    - Communicating
    - Crowdsourcing
    - Publishing
    - Sharing
- Meta-Activities
    - Assessing
    - Community Building
    - Give Overview
    - Project Management
    - Teaching / Learning

Tadirah techniques

The Tadirah techniques are the following (note that this list is open, new terms can be added to it):

- Bit Stream Preservation
- Brainstorming
- Browsing
- Cluster Analysis
- Collocation Analysis
- Commenting

- Concordancing

- Debugging

- Distance Measures

- Durable Persistent Media

- Emulation

- Encoding

- Gamification

- Georeferencing

- Information Retrieval

- Linked Open Data

- Machine Learning

- Mapping

- Migration

- Named Entity Recognition

- Open Archival Information Systems

- Pattern Recognition

- Photography

- POS-Tagging

- Preservation Metadata

- Principal Component Analysis

- Replication

- Scanning

- Searching

- Sentiment Analysis

- Sequence Alignment

- Technology Preservation

- Topic Modeling

- Versioning

- Web Crawling

- Text Mining

TaDIRAH Objects

The TaDIRAH objects vocabulary contains 36 types of research objects, including the most commonly used by Arts and Humanities scholars (note that this list is open, new terms can be added to it).

- Artifacts
- Bibliographic Listings
- Code
- Computers
- Curricula
- Digital Humanities
- Data
- File
- Images
- Images (3D)
- Infrastructure
- Interaction
- Language
- Link
- Literature
- Manuscript
- Map
- Metadata
- Methods
- Multimedia
- Multimodal
- Named Entities
- Persons
- Projects
- Research
- Research Process
- Research Results
- Sheet Music

- Software

- Sound

- Standards

- Text

- Text Bearing Objects

- Tools

- Video

- VREs

AureHAL disciplines

The disciplines must be chosen in the following list:

- Biological anthropology

- Social Anthropology and ethnology

- Archaeology and Prehistory

- Architecture, space management

- Art and art history

- Classical studies

- Demography

- Law

- Economies and finances

- Education

- Environmental studies

- Gender studies

- Geography

- Management

- History, Philosophy and Sociology of Sciences

- History

- Communication sciences

- Linguistics

- Literature

- Cultural heritage and museology

- Musicology and performing arts

- Philosophy

- Psychology

- Religions

- Political science

- Sociology

- Methods and statistics

Standards knowledge base

The list of the standards already described in the Standards Knowledge base can be found here. If you don't find the standard you want, you can create a description using this sample file and upload it here to the GitHub folder standardsDesc.

Note that the value to indicate in the key is the value of the field "standard_abbr_name". See below the sample file.

```
<doc>
<field name="id">33 (must be incremented by 1 for each new standard)</field>
<field name="standard_abbr_name">Standard abbreviated Name This information will
be used in the TEI file to refer to this description</field>
<field name="standard_complete_name">Standard Complete name</field>
<field name="standard_type">Two values: 'standard' OR 'method'.
 When describing a format, use 'standard',
 when describing a protocol or a set of techniques, use 'method'</field>
<field name="standard_desc_eng">English Description</field>
<field name="standard_desc_fr">French Description</field>
<field name="standard_desc_deu">German Description</field>
<field name="standard_desc_esp">Spanish Description</field>
<field name="standard_data_type">select from: Horizontal->e.g. XML, CSV and
vertical ->e.g. EDM</field>
<field name="standard_link">http://link to standard official page.com</field>
<field name="standard_tags">Tag1: example-> Classification</field>
<field name="standard_tags">Tag2: example-> Human-history</field>
<field name="standard_tags">Tag3: example-> Research Activities-
Organizing</field>
<field name="standard_tags">Tag4: example-> Research Objects - Digital
Humanities</field>
<field name="standard_resources">http://link to resource</field>
</doc>
```

4.5.2.4 <linkGrp> element

`<linkGrp>` is the container for the resources associated to a given step. It can have three attributes:

- The attribute `@type` is required and can have two values:

  - *generalResources*: for resources that give general input about a standard, a protocol, etc.

  - *projectResources*: for resources that show examples of real projects using the described standard, protocol, etc.

- When `@type` has *projectResources* for value, two more attributes are required:

- `@source` for the name of the project mentioned

- `@corresp` for a url pointing to or identifying the project

```
<linkGrp type="generalResources">
    <ref type="Report" source="zotero" target="ZQVB6CIP"/>
  </linkGrp>
  <linkGrp type="projectResources" source="CODATA"
    corresp="http://www.codata.org/">
    <ref type="Report" source="zotero" target="G4UPDPG3"/>
  </linkGrp>
```

4.5.2.5 <ref> element

The attributes for `<ref>` are `@type, @subtype, @source` and `@target`.

- The attribute @type is required. Its values are taken from the Zotero item types, plus SSK specific values. Possible values are:

  - spec: the specification, of a standard for instance.

  - report: technical reports

  - blog: blog posts

  - tutorial: tutorials or guidelines

  - script: Scripts and code samples

  - paper: Scholarly papers

  - library: Computing libraries

  - tool: a link to a service or a software useful for a given step.

  - database

  - method

- bibliography
- schema
- the @source attribute in <ref> is used by the SSK to record where the full information about the resource is stored, and that the SSK queries. The values are a semi-closed list. The source attribute has for possible values:
  - *zotero*: The Parthenos WP4 Zotero library: WP4 Zotero Library
  - *github*: resources hosted in a GitHub repository, preferably the Parthenos WP4 repository, but not exclusively
  - *isidore*: resources described in the platform of search Isidore dedicated to Humanities and Social Sciences.
- The @target attribute specifies the destination of the reference with an URL.

```
<ref type="spec" subtype="standard"
 target="http://zotero.org/groups/427927/items/BEVAWMPX"/>
```

4.5.2.6 <param> element

See below the advanced features section.

4.5.3 Advanced features

4.5.3.1 Customize a step or a scenario

It is possible to modify the content of a step directly in the scenario file, for instance, modifying the label to contextualize it, or adding a very specific resource. To do so, event and its children can be specified with the mode attribute; with the possible following values:

- change
- add

```
<event type="step" ref="step_EaXswO_290517">
<head mode="add" xml:lang="jp">...</head>
<desc type="definition" mode="change">...</desc>
...
</event>
```

4.5.3.2 The parameters

When pointing to a step inside a scenario, it is possible to use parameters to refine the behaviour of this step. This parametrization uses the element <param> in <event>. Two different uses are possible for the moment, to refine the resources selection in a given step, or to include some steps of a scenario in another scenario.

4.5.3.2.1 Refine the resources

It is possible to select the resources to be displayed in a scenario. The criteria are based on the taxonomies used by the SSK model:

- Tadirah Activities
- Tadirah techniques
- aureHAL disciplines
- Standards

The element <param> contains an attribute @name, that contains a formal name to identify on which taxonomy the parameter is applied. The possible values are:

- activity
- technique
- datatype
- discipline
- standard

Another attribute @value contains the term used to select the wanted resources. In other words, in the example below, the resources displayed would only be the <ref> that contains one or more <term> elements with values "XML", "conversion" and "Text Bearing Objects".

```
<event type="researchStep" ref="referencedStep">
<label mode="replace">New label</label>
<desc mode="replace">new description</desc>
<!-- resources -->
<param name="standard" value="XML"/>
<param name="activity" value="conversion"/>
<param name="technique" value="Text Bearing Objects"/> ...
</event>
```

In this situation, all the following resources would be selected.

```
<ref type="code" target="// URL //">
<term type="activity" source="tadirah" key="conversion"/>
<term type="standard" key="XML"/>
</ref>


<ref type="code" target="// URL //">
<term type="technique" source="tadirah" key="Text Bearing Objects"/>
<term type="standard" key="XML"/>
</ref>


<ref type="code" target="// URL //">
<term type="technique" source="tadirah" key="Text Bearing Objects"/>
<term type="activity" source="tadirah" key="conversion"/>
</ref>


<ref type="code" target="URL">
<term type="standard" key="XML"/>
</ref>
```

4.5.3.2.2 Include partially a scenario into another

This mechanism records:

- The reference to a scenario
- A set of steps, not necessarily consecutive.

In this case, the attributes of `<param>` are also @name and @value, but they have a different behaviour. The @name value is *range*. The attribute @value records the interval of the steps (i.e. their order number) in the scenario to include. To indicate a consecutive interval, the steps indexes should be separated by a hyphen (-). To indicate non-consecutive steps, the steps indexes should be separated by a comma (,). These two behaviours can be mixed (see examples below).

Ex1: This parameter would select steps 1, 2 and 3.

```
<event xml:id="jjjj" type="researchScenario" ref="scenario_to_Be_Included">
<param name="range" value="1-3"/>
</event>
```

Ex2: This parameter would select steps 1 and 3.

```
<event xml:id="jjjj" type="researchScenario" ref="scenario_to_Be_Included">
 <param name="range" value="1,3"/>
</event>
```

This parameter would select steps 1, 3, 5, 6 and 7.

```
<event xml:id="jjjj" type="researchScenario" ref="scenario_to_Be_Included">
 <param name="range" value="1,3,5-7"/>
</event>
```

4.5.3.2.3 Mixing parameters

This example shows the inclusion of a scenario into another and a filter based on a keyword for a particular step in this subset. If the @corresp is not used, the parameter will be applied to all the included steps.

```
<event xml:id="jjjj" type="researchScenario" ref="scenario_to_Be_Included">
 <param name="range" value="1,2,4-6"/>
 <!-- filter resources of the step nr 2 -->
 <param name="standards" value="XML" corresp="#2"/>
</event>
```

5   Technical infrastructure

The implementation of the SSK is based on a flexible, easy to deploy and maintain architecture. It is composed of independent entities that communicate together through services (REST / JSON):

- The main entity, the core of the SSK, is the back-end, which makes queries to our data repositories (GitHub, Zotero, etc.) and processes retrieved data.
- The data processed from the core part and from the search engine are all delivered via an API to third-party applications like the SSK interface (the front-end), which is an entity of our architecture.
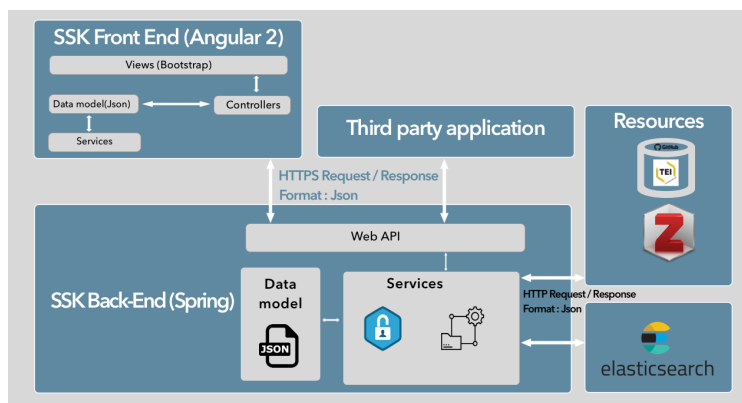
- This core/back-end communicates with a search engine, part of the architecture, based on Apache Lucene: ElasticSearch.

Figure 5: Architecture of the SSK application

5.1 SSK input

The SSK processes TEI files stored on GitHub and divided into two folders, scenarios and steps. For more information about the data model, check the dedicated section.

5.2 SSK Back-end

This part is the main component of the SSK, it has been built using **Spring Boot version 1.5.4.RELEASE**, a Java based framework (more details here). It contains modules for:

5.2.1 Processing SSK data

This means retrieving TEI content from SSK GitHub repository. A very important step is to validate the retrieved content according to the RELAX NG schema defined for SSK files. If the file is validated, TEI content is converted in JSON format using the built-in scripts provided by the TEI consortium:

https://github.com/ParthenosWP4/SSK/blob/dev/SSK-Server/src/main/resources/lib/bin/teitojson.

The references contained in TEI files are resolved to complete the data:

- For keywords, in particular the `standards`, the data retrieved in the TEI files is used to get extra information about the terms. For standards, a knowledge base of standards is queried to retrieve more information (standard complete name, multilingual standard description and links).
- For resources, queries are made on platforms such as Zotero and **GitHub** (for project resources). **Website scraping** is also used to make resources more consistent.

Once the data is completed, it is then pushed into Elasticsearch for indexing and search.

Note that each scenario and its steps are also pushed on Elasticsearch. Each step is linked to the scenarios it is part of, by a parent attribute, directly in Elasticsearch. Resources and descriptive metadata have been also targeted with their parent identifier in the same way.

5.2.2 API serving

The SSK Back-End makes its data available via a REST API, built with Spring boot. This API allows third party applications to retrieve scenarios, steps, resources and their descriptive metadata.

The API V1 is accessible via the URL: http://ssk.huma-num.fr/ssk_services-0.0.1/ssk

Example queries:

- Get all the standards mentioned in scenarios and steps: http://ssk.huma-num.fr/ssk_services-0.0.1/standard/all

- To get the descriptive metadata of a scenario, the query is composed of the keyword scenario, the Id of the scenario, and, as parameters, the metadata the user wants to get. For instance, the following query serves the title, description, image and descriptive terms for the scenario whose ID is SSK_sc_DisseminationFieldSurveys:

  http://ssk.huma-num.fr/ssk_services-0.0.1/scenario/SSK_sc_DisseminationFieldSurveys?fields=title,desc,image,scenario_metadata,author&fromSSK=true

A V2 is planned in order to serve more easy-to-handle content, for instance giving all the scenarios with a given author or related to a given institution, discipline or standard, etc.

5.2.3 User management

This part is still work in progress. We plan to deliver it by the **first trimester 2019**.

Creating an account will allow the user to:

- bookmark scenarios and steps in order to facilitate future navigations or stay in touch with some specific research fields.

- Create their own scenarios based on existing ones or by starting from scratch.

5.2.4 Creation and update of scenarios

We mentioned that it will be soon possible to create or update scenarios from forms in the Front-End part of SSK. Meanwhile, it is also possible to do so on our GitHub repository, via directly updating the TEI files. To take into account these modifications or updates in our Elasticsearch search engine, we use GitHub webhooks which allows us to make POST requests to the Back-End. Thus an end-point in this module of SSK (Back-End) receives the GitHub request and its data, processes these data so that it can be pushed into Elasticsearch.

5.2.5 Search Engine

The search engine module has been built in order to allow refined information retrieval. It relies on Elasticsearch, version 6.2.4, a full-text search and analytics engine, that allows us to easily propose multi-criteria and full-text queries to the users, but also to autocomplete suggestions.

The indexed data is not only the information contained in the scenarios and steps descriptions stored in TEI files, but also the data hosted on the Zotero database and the SSK Standards Knowledge Base.

5.3 SSK Front-end

The SSK Front-End is the client part of the SSK, where users can see the SSK data (TEI files + Zotero references).

It is built with Angular 5.2.11, a framework using Typescript 2.9.2 for building web applications. Angular proposes to set a hierarchy of components (or classes), associated with HTML templates. Components use services (or functions) to communicate with the server (to fetch the data for example) and to link components between them.

The image below (taken from https://angular.io/guide/architecture) shows the architecture of an Angular application.
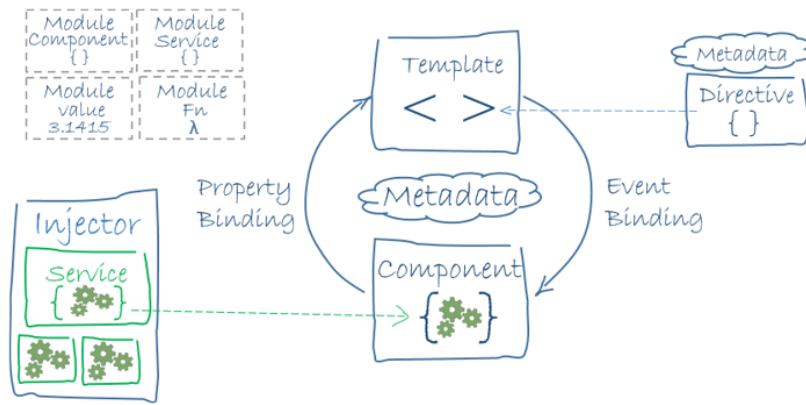
Figure 6: Overview of an Angular application

To display SSK data on the web interface, we created several components, services and templates.

Components, combined with templates, are used to represent the different SSK layers: scenarios, steps, resources.

Services are used to share data between these layers, but they also allowed us to design functions that queries data from main modules of the SSK (Core SSK or Back-End) via a REST API.

5.4 Deployment

As the SSK is based on three main parts, each of its modules (Elasticsearch, Front-End, Back-End) need to be deployed independently. The communication between the modules is made with dedicated endpoints. The Elasticsearch endpoint serve the data to the Back-end, and in the same way, the Back-End also offers an endpoint to the Front-end so that it can get SSK's data for display.

5.4.1 Install ElasticSearch

The binary packages of Elasticsearch have only one dependency: Java. The oldest supported version is Java 8. To download and install Elasticsearch, use the commands that work with your system (deb for Debian/Ubuntu, rpm for Redhat/Centos/Fedora, mac for OS X, and win for Windows). See this link for more details.

5.4.2 Deploy Back-end (Spring boot application)

The Back-End is composed of two main elements: **Spring Boot** and **Tomcat** (Java Servlet Container).

Spring Boot is a *convention over configuration* framework that allows us to set up a production-ready setup of a Spring project. By default, Spring Boot builds a standalone

Java application that can run as a desktop application or be configured as a system service. For the SSK, we use it as a service.

Opposite to standalone applications, Tomcat is also installed as a service that can manage multiple applications within the same application process, avoiding the need for a specific setup for each application. The SSK spring boot application use Gradle as build automation system.

To build a Tomcat-deployable WAR application:
1. execute the Gradle `build` command.
2. The WAR file is generated at `target/ssk_services.war` (assuming the Gradle `artifactId` is `ssk_services`).
3.

To have our WAR file deployed and running in Tomcat, we need to complete the following steps:
1. and unpackage it into a tomcat folder
2. Copy our WAR file from `target/ssk_services.war` to the `tomcat/webapps/` folder
3. From a terminal navigate to `tomcat/bin` folder and execute `catalina.bat run` (on Windows) and `catalina.sh run` (on Unix-based systems)
4. Go to http://localhost:8080/ssk_services/ssk

This is how the SSK Back-End has been deployed on the D4Science infrastructure, although Elasticsearch and the Tomcat server have been configured by the platform engineers.
Source: Deploy a Spring Boot WAR

5.4.3 Front-end Deployment (Angular application)

Build and deploy the Front-end of the SSK which is an Angular based application, requires Angular CLI to be installed on your computer.
The steps to follow are:
1. To build angular applications, execute the `ng build` command. This will generate files in the `dist` folder located at the root of the application folder.

2. Copy everything within the output folder (dist/ by default) to a folder on the server.

3. Configure the server to redirect requests for missing files to index.html

Source: Angular deployment.

6   Conclusions

The Standardization Survival Kit, though released, is a living product that will evolve in the following months, until the end of the PARTHENOS project and even after. The user tests have shown its usefulness, but that it needs careful 'on-boarding', at least 30 scenarios from across all the Cultural Heritage domains and should be multilingual. These demands will be taken in account in the SSK V2.0, to be released by the end of April 2019. This version will also offer to users the possibility to create an account, save favourite scenarios and create new ones directly via a form-based interface.

The development of the contribution feature will be accompanied by another user testing phase, following the organisation used for the first tests. The SSK development team will also engage with the users by organising workshops around the tool, in order to gather feedback and encourage the creation of new scenarios.

References

Burnard, Lou 2000, Text Encoding for Interchange: A New Consortium. Ariadne(24). http://www.ariadne.ac.uk/issue24/tei/, accessed April 24, 2011.

Cascaes Cardoso, Marina 2017, The Onboarding Effect: Leveraging User Engagement and Retention in Crowdsourcing Platforms. *In* Pp. 263–267. ACM. http://dl.acm.org/citation.cfm?id=3027063.3027128, accessed October 29, 2018.

Leahy, Richard 1992, Twenty Titles for the Writer. College Composition and Communication 43(4): 516–519.

Library of Congress n.d., EAD: Encoded Archival Description (EAD Official Site, Library of Congress). http://www.loc.gov/ead/, accessed November 24, 2015.

Riondet, Charles, 2018, The SSK Data Model (TEI) — SSK Documentation. https://ssk.readthedocs.io/en/latest/1_ssktei.html, accessed October 24, 2018.

Romary, Laurent, Piotr Banski, Jack Bowers, et al. 2017, Report on Standardization (Draft). https://hal.inria.fr/hal-01560563.

Romary, Laurent, Emiliano Degl'Innocenti, Charles Riondet, et al. 2016, Standardization Survival Kit (Draft). https://hal.inria.fr/hal-01513531.

Seillier, Dorian 2018, A Brief Report on the Usability Test for the Standardization Survival Kit (SSK). Other. Inria Paris. https://hal.archives-ouvertes.fr/hal-01901057, accessed October 24, 2018.

Universitat Autònoma de Barcelona N.d., Describing a Process. Coursera. https://www.coursera.org/lecture/teaching-english/3-1-1-describing-a-process-mjuio, accessed June 29, 2018.

References