# The Standardization Survival Kit: presentation

To support the digital turn of Social Sciences and Humanities research, it is necessary to stabilize knowledge on standards and research good practices. The goal of the **Standardization Survival Kit**, developed with the [PARTHENOS project](http://parthenos-project.eu) project, is to accompany researchers along this way, giving access to standards and best practices in a meaningful way, by the mediation of research scenarios. 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 in an open tool where users are able to publish new scenarios or adapt existing ones. Those 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.

## Why standards after all?

The main issue in defining a policy about standards is to understand what they actually are. 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 comparable (or interoperable) results. This will also foster innovative, cross-disciplinary research paths, and hopefully contribute to bridging the gap between the different cultures that are represented in the broad landscape of Humanities and Cultural Heritage studies.

Nevertheless, standards are not regulations. There is no obligation to follow them 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.

The SSK: a toolkit for Humanities scholars ---------------------------------

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

1. an awareness of the appropriate standards and the advantages to be gained by adopting them,
2. 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 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;
* **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 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 by domain experts, from **real life researcher-oriented use cases (PARTHENOS, 2016)**, divided into different steps, and involving specific tasks.

Those 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 catalog of resources.

## Design principles

TODO

## SSK components

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

### 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 the 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 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.

### How to create a scenario for the SSK

The following instructions help users create a scenario for the SSK. They are conceived as a "scenario", a step-by-step tutorial.

First, contributors should be aware that:

* they can submit their new scenarios directly in TEI 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.
* In TEI, it is possible to translate the prose, by duplicating the elements head, desc or term, and adding xml:lang attribute

image

#### Choose explicit titles starting with a verb or a gerund

The choice of good titles for a scenario and its steps is crucial. It will be the entry point of the users, that need to understand at first reading the scope and the aim of a scenario. To do so, we advice to build titles : <list>

* that starts with a verb or a gerund that describe 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

References:

* Documentation of the TEI element head
* Leahy, Richard. ‘Twenty Titles for the Writer’. College Composition and Communication 43, no. 4 (1992): 516–19. <https://doi.org/10.2307/358644>.

#### 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.
* 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.
* It is possible to point to external links (in TEI, with the following: <ref target="//url here//">text of the link</ref>
* Lists are also available. The TEI elements are <list> and <item>

References:

* Documentation of the TEI element desc
* Universitat Autònoma de Barcelona. ‘Describing a Process’. Coursera. Accessed 29 June 2018. <https://www.coursera.org/lecture/teaching-english/3-1-1-describing-a-process-mjuio>.

#### Associate keywords to the scenario or the step

The SSK vocabularies are

* [Research activities](http://ssk.huma-num.fr/#/glossary/activities), taken from [Tadirah](https://github.com/dhtaxonomy/TaDiRAH)
* [Research techniques](http://ssk.huma-num.fr/#/glossary/techniques), taken from [Tadirah](https://github.com/dhtaxonomy/TaDiRAH)
* [Research objects](http://ssk.huma-num.fr/#/glossary/objects), taken from [Tadirah](https://github.com/dhtaxonomy/TaDiRAH)
* [Standards](http://ssk.huma-num.fr/#/glossary/standards), taken from the SSK Standard Knowledge base (supported by DARIAH-IT)
* [Disciplines](http://ssk.huma-num.fr/#/glossary/standards), taken from [aureHAL](https://aurehal.archives-ouvertes.fr/domain?locale=en)

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

* Disciplines
* Techniques
* Objects
* Standards

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. In general, for each keyword type, we recommend to choose between 1 and 4 keywords.

References:

* Documentation of the TEI element term
* [TaDiRAH - Taxonomy of Digital Research Activities in the Humanities](http://tadirah.dariah.eu/)

#### Choose an illustration for the scenario

* This illustration must closely relates with the purpose of the scenario, i.e. not only with the discipline or the period. Screenshots are accepted
* Landscape orientation image are recommended
* Maximum size : 2 Mo
* Accepted formats : png, jpg
* It must be published under the licence CC-BY or CC-0.

References:

* [Unsplash](https://unsplash.com/), a gallery of free images and photos

#### Identify relevant resources processing

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

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

References:

* [State of the art Wikipedia article](https://en.wikipedia.org/w/index.php?title=State_of_the_art&oldid=845308793)

#### Link the resources to the step

There is 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](https://www.zotero.org/groups/427927/items?)).

* To populate it, a Zotero account is necessary (create it [here](https://www.zotero.org/user/register)) as well as a membership in the SSK group (apply [here](https://www.zotero.org/groups/427927/ssk-parthenos?)). 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](https://www.zotero.org/support/screencast_tutorials).
* After adding a resource in the Zotero Library, it should be linked to the step, with the help of its 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"/ >.
* 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)</item>
* 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

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

he SSK is adaptable by nature and contributors don't have to start from scratch 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: custom.

#### Advanced SSK functions (2) : link scenarios

Link scenarios together, or in other words, include a scenario (entirely or partially) into another is an interesting possibility when a scenario is a pre-condition 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. See param. .. \_reTEI:

# The SSK data model (TEI)

The SSK offers researchers needing standardized methods and resources complete frameworks to carry out their project, in Arts and Humanities and Heritage science. It takes the form of step by step research scenarios where the use of standards is clearly identified. Theses scenarios are divided into different steps, implying specific tasks. Each step contains a set of bibliographical resources.

##### TEI: the underlying data model

The underlying data model of the SSK itself respects a standard, the Text Encoding Initiative, and is publicly available. 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.

## 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 consists 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.

## 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 -->  
 <term mode="add"/>  
 </head>  
 ...  
</event>

* the type attribute is required in most elements. See below for details.

##### The Resources

## Zotero

The resources to be presented in the SSK are preferably stored in a Zotero database, accessible [here](https://www.zotero.org/groups/427927/ssk-parthenos) . To add a resource, an account on Zotero is required. Contact the SSK team to join the group (ssk [at] inria [dot] fr)

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 webpages: website title
  + For blogposts: blog title
  + For journal articles, books, book sections, documents: library catalog

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>

## GitHub

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

* For a GitHub User:
  + name
  + html\_url
  + bio
  + updated\_at
  + type
  + avatar\_url
* For a GitHub repository
  + owner (NB : a GitHub user)
  + full\_name
  + html\_url
  + description

A dedicated GitHub repository has been set for projects supported or maintained by Parthenos. It is available here. Contact the SSK team for more information.

##### 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:
  + sc for scenario
  + an underscore
  + a condensed title of the scenario in camel case
* for steps:
  + the string step
  + an underscore
  + the intials of the step name, with the liaison words in lower case, and the meaningful words in upper case
  + an underscore and the date (optional)

## Scenarios and steps structure

### 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) include 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.

### Steps

A full description of the scenario step.

###### 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 elements of a <event> 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 two linkGrp.

## Content of scenarios and steps

### head

The TEI head element record 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. Create associated documentation

### desc

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

### term

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 NEDIMAH type taxonomy for Information resource (or objects)
* the Dariah-IT Standard Knowledge base
* aureHAL disciplines

###### Functioning

These taxonomies are declared with the attributes type and source. The attributes of term are:

* The type attribute gives an 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 NEMO taxonomy Information Resource Types, 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

###### 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
  + Cleanup
  + 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 :

* 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 common used by Arts and Humanities scholars.

* 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](https://github.com/ParthenosWP4/SSK/tree/master/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\_about\_the \_standard.com</field>  
</doc>

### linkGrp

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, ...
  + `projectResources`: for resources that show examples of real projects using the described standard, protocol, ...
* 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>

### ref

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 URI.

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

### param

See below the advanced features section

##### Advanced features

## 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>

## 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.

### Parameter #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
* NEMO Data types
* 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>

### Parameter #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 an consecutive interval, the steps indexes should be separated by an hyphen: -. To indicate non-consecutive steps, the steps indexes should be separated by a comma: ,. These two behaviours can be mixed (see examples below)

###### A set of steps, sometimes consecutive, sometimes not consecutive

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>

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>

### 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 param 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>

# Technical architecture

The implementation of the SSK is based on a flexible, easy to deploy and maintained architecture, composed of independent entities that communicate together through services (REST / JSON). The most important entity is the core of the SSK, which makes queries to our data repositories (Github, Zotero, etc.) and processes retrieved data. This core also communicates with a search engine (part of the architecture) based on Apache Lucene. 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, which is an entity of our architecture. Below is the architecture of SSK.

image0

### SSK’s input

The SSK processes TEI files stored on Github and divided into two folders, [scenarios](https://github.com/ParthenosWP4/SSK/tree/master/scenarios) and [steps](https://github.com/ParthenosWP4/SSK/tree/master/steps). For more information about the data model, check the dedicated section: reTEI.

### Core of SSK

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](https://spring.io/blog/2017/06/08/spring-boot-1-5-4-available-now)). It contains modules for :

###### Processing SSK data

This means retrieving TEI content from SSK Github repository. Validate retrieved content according to the [RELAX NG schema](https://github.com/ParthenosWP4/SSK/blob/master/spec/TEI_SSK_ODD.rng) defined for SSK files. After validation TEI content are convert into Json format using **XSLT**. After this, some resources are completed such as standards and resources. For standards, a knowledge base of standards is queried to retrieve more informations (standard complete name, multilingual standard description and links). In this same way step’s resources are also completed querying platform like [Zotero](https://www.zotero.org/) (which is a free software for managing bibliographic references) , Github (for project resources). **Website scraping** is also use to complete resources. When all data are completed for a TEI content (scenario or step), the JSON content obtained is then store on ElasticSearch for future easier searches.

###### API serving

The SSK makes its data available via web API. There are API to retrieve scenarios , steps, resources and standards. In fact all SSK content can be retrieve. This API also give the possibilities to specify number of item, fields and type (scenarios , steps, resources or standards) you want to request.

###### User management (for SSK contribution)

TODO

###### Search Engine

TODO

###### SSK Front-End

TODO