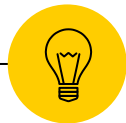


The Standardization Survival Kit (SSK)

Bringing best practices to research communities in the
Humanities

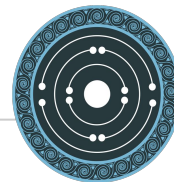


Charles Riondet (Inria - ALMAAnaCH)



Le projet H2020 PARTHENOS

<http://parthenos-project.eu>



PARTHENOS

- 2015-2019
- Construire des synergies à l'échelle des infrastructures de recherche
- Rendre plus fluide l'expérience des utilisateurs, contributeurs de ces infrastructures
 - Services plus intégrés (environnements virtuels de recherche)
 - Recommandations/bonnes pratiques

Concept transversal ⇒ Interopérabilité



Interopérabilité + bonnes pratiques de recherche = Standardization

- La recherche financée par la CE doit être interopérable (FAIR data, DMP, EOSC, ...) ;
- Rendre des données interopérables = les exprimer avec un/des standard(s);
- Dans un processus de recherche en environnement numérique, s'appuyer sur les standards représente un gain objectif

● Standards?

Différents types

- Des formats de données
 - XML :
<https://www.w3.org/TR/xml/>
 - TIFF : ISO 12639:1998
- Des protocoles
 - ISO 11554:2017 : Test methods for laser beam power, energy and temporal characteristics

3 mots clés

- **Consensus** : Représentent un état de l'art dans un domaine précis,
- **Publicité** : *Stricto sensu*, les standards sont des documents normatifs publics,
- **Maintenance** : Sont des objets pérennes, mais capables d'évolution.

Standards?

Rencontrés à tout moment du processus de recherche

- ◉ **Production** : ALTO XML, standard pour la mise en page physique et de la structure logique d'un texte transcrit par OCR ;
- ◉ **Processing** : Annotation linguistique, stockée au format XML TEI (Text Encoding Initiative) ;
- ◉ **Archivage** : Modèle OAIS (Open Archival Information System, ISO 14721:2012) => modèle conceptuel destiné à la gestion, à l'archivage et à la préservation à long terme de documents numériques.



Le SSK dans PARTHENOS



But du *Standardization Survival Kit* dans PARTHENOS :
**Fournir une expertise/appui aux chercheurs sur
l'utilisation des standards**

Quelles communautés?

- Sciences Humaines
- Sciences Sociales
- Sciences du Patrimoine



Le SSK dans PARTHENOS

Avantages d'un projet européen:

- **Diversité** : multitude d'experts de différentes disciplines
- **Synergie** : volonté (au moins formelle) de coopération

⇒ Point de départ pour bâtir des cas d'usages "inspirés des faits réels"



Le SSK dans PARTHENOS

Équipe de développement

Inria ALMAAnaCH

- Laurent Romary (DR, WP leader, superviseur)
- Marie Puren, Charles Riondet (Conduite du projet, data model)
- Dorian Seillier (Design UI/UX)
- Lionel Tadjou, Damien Biabiany (Développement application web)

Itérations, beta tests

PARTHENOS WP4 on Standardization:

- Klaus Illmayer (OEAW),
- Karolien Verbrugge (NIOD),
- Roberta Giacomi (SISMEL),
- Panos Siozos (FORTH),
- etc.



Évolution du concept

1. Fournir une expertise/appui aux chercheurs sur l'utilisation des standards ;
2. Donner du contexte aux standards ;
3. Les rattacher à une activité concrète de recherche ([TaDirah – Taxonomy of Digital Research Activities in the Humanities](#)) ;
4. Lier ces activités entre elles => Décrire un processus de recherche qui s'appuie sur l'utilisation de standards ;



Évolution du concept

Toutes les étapes de recherches ne sont pas exprimables avec des standards:

- Questions légales et éthiques ;
- Évaluation et comparaison de résultats ;
- ...

Mais notion de **Bonnes Pratiques** reste opérante.



Évolution du concept

Problème des standards – protocoles :

- ◉ Texte normatif très formel et difficile à lire ;
- ◉ Un protocole = une suite de tâches utilisant des outils et techniques particulières



Évolution du concept

1. Fournir une expertise/appui aux chercheurs sur l'utilisation des standards ;
2. Du contexte aux standards ;
3. Les rattacher à une activité de recherche (TaDirah – Taxonomy of Digital Research Activities in the Humanities) ;
4. Lier ces activités entre elles => Décrire un processus de recherche qui s'appuie sur l'utilisation de standards ;
5. **Plateforme pour**
 - a. **Documentation des bonnes pratiques de recherche en environnement numérique**
 - b. **Expression "Human readable" de protocoles standardisés**



Le SSK dans PARTHENOS, et au delà.

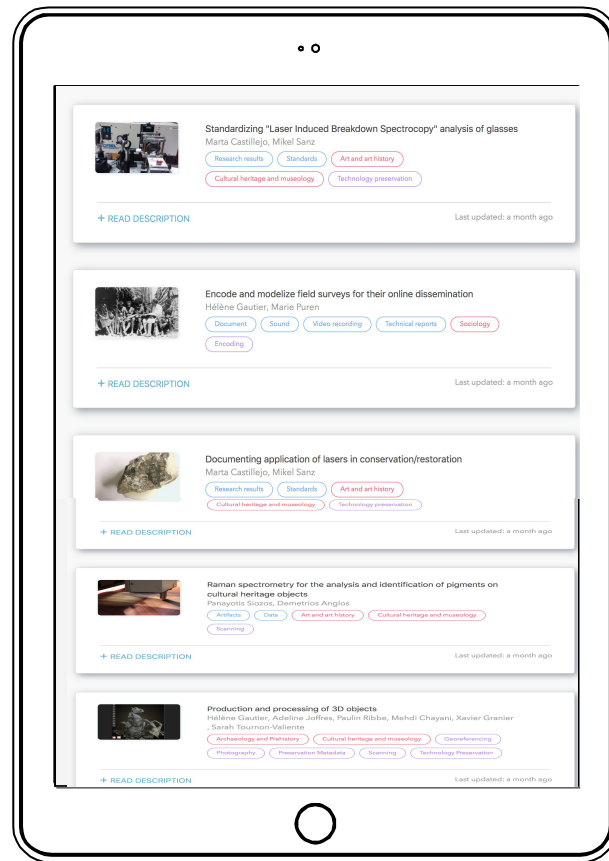
Proposer un service agréable, durable et adaptable

- UI/UX designer
- Méthode "agile": idéation collective, tests utilisateurs
- Modélisation "souple" des données
- Rencontres avec les communautés d'utilisateurs potentiels



Des scénarios

Succession d'étapes qui
proposent des
informations
contextuelles et des
exemples pour utiliser
les standards dans un
projet de recherche.



Structure en 3 niveaux



Scénario

Un cas d'usage complet,
un processus de recherche.



Étape

Une tâche unique à réaliser au
sein d'un scénario, avec l'aide
d'une ou plusieurs ressources.



Ressource

Un outil, service ou document
pouvant guider le lecteur dans
l'accomplissement de la tâche,
avec l'aide de standard(s).



Structure en 3 niveaux

Description scénario (*techniques, disciplines, objets*)

1. Description étape (*activités, standards*)
 - a. ressource
 - b. ressource
2. Description étape (*activités, standards*)
 - a. ressource
 - b. ressource
3. Description étape (*activités, standards*)
 - a. etc.

Ressources sur les standards

Standards documentation
(ISO, TEI)

Official publications and
reports (D4Science, HAL,
Zenodo)

Documentation



Reference libraries
organised by domains and
standards

Maintenance with Zotero

Bibliography



Code snippets (GitHub)

Tools & services
(D4Science)

Technical
resources



Wikis

Blog posts
(Hypotheses.org)

Discussion lists

User
communities





Démo





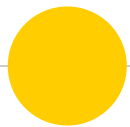
Plateforme intuitive et collaborative

- Consulter et suivre les orientations exprimées dans les scénarios
- Proposer de nouveaux scénarios



Pourquoi utiliser/contribuer ?

- ◉ Aligner ses méthodes sur les bonnes pratiques dans sa discipline,
- ◉ Rendre visible ses choix méthodologiques
- ◉ Partager son projet dans une nouvelles forme, différente de l'article/post de blog.



De la cohérence avec ses principes

Le SSK, une application web 100% standards

Des scénarios en TEI

Les scénarios et étapes sont décrits avec le format TEI (Text Encoding Initiative), pour permettre facilement leur réutilisation/adaptation.

<http://github.com/ParthenosVWP4/SSK/tree/master>

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

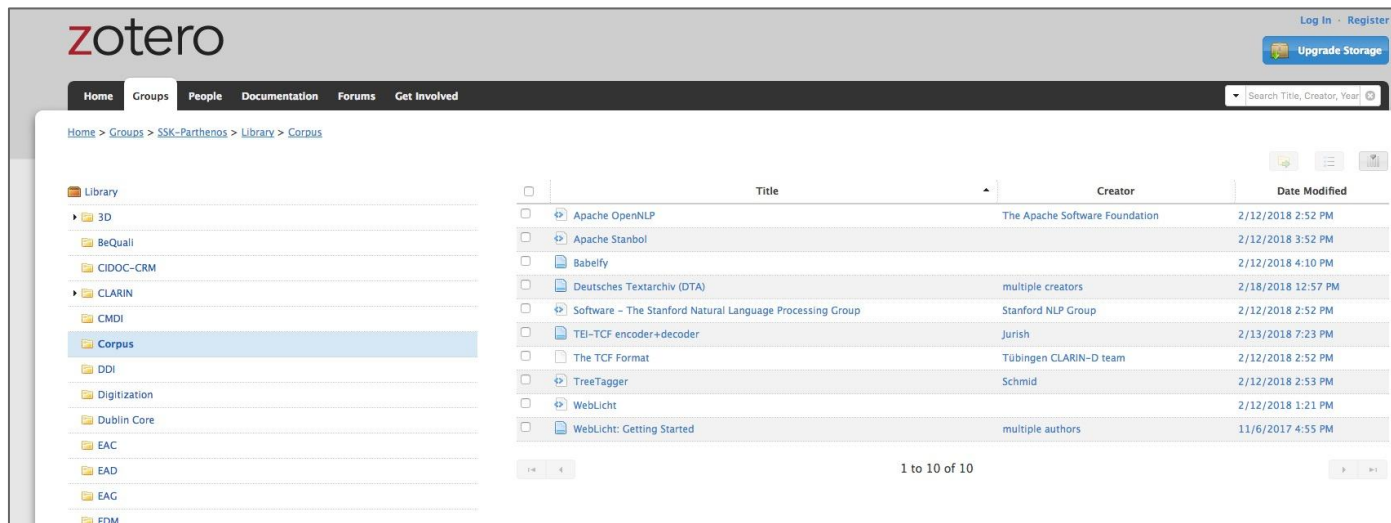
```
<listEvent>
  <event ref="step_OtICoP_171117" type="researchStep" xml:id="s1"/>
  <event ref="step_CaC_171117" type="researchStep" xml:id="s2"/>
  <event ref="step_SaD_171117" type="researchStep" xml:id="s3"/>
  <event ref="step_A_171117" type="researchStep" xml:id="s4"/>
  <event ref="step_CiSF_171117"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

Poster à la conférence TEI 2018 : <https://hal.inria.fr/hal-01902702>

Standards bibliographiques

Toutes les références sont gérées avec le logiciel de gestion de référence bibliographiques **Zotero**

<https://www.zotero.org/groups/427927/ssk-parthenos/>

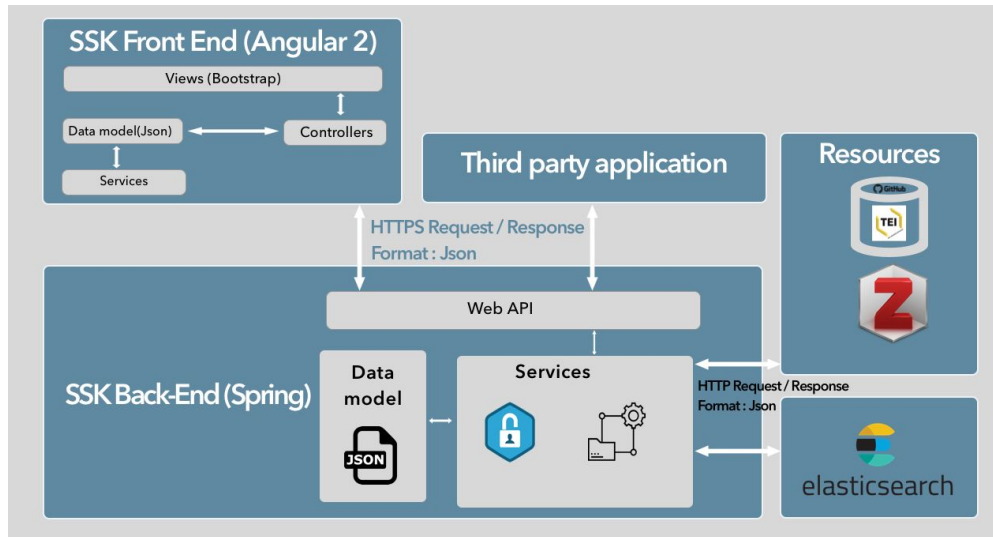


The screenshot displays the Zotero web interface. At the top, the 'zotero' logo is on the left, and 'Log In' and 'Register' links are on the right. Below the logo is a navigation bar with links: Home, Groups, People, Documentation, Forums, and Get Involved. A search bar is located to the right of the navigation bar. The main content area shows a breadcrumb trail: Home > Groups > SSK-Parthenos > Library > Corpus. On the left, a sidebar lists various folders under 'Library', with 'Corpus' selected and highlighted in blue. The main area displays a table of bibliographic references. The table has three columns: Title, Creator, and Date Modified. There are 10 entries in the table, each with a checkbox on the left. The entries are: Apache OpenNLP, Apache Stanbol, Babelify, Deutsches Textarchiv (DTA), Software - The Stanford Natural Language Processing Group, TEI-TCF encoder+decoder, The TCF Format, TreeTagger, WebLicht, and WebLicht: Getting Started. The table is paginated, showing '1 to 10 of 10' items.

	Title	Creator	Date Modified
☐	Apache OpenNLP	The Apache Software Foundation	2/12/2018 2:52 PM
☐	Apache Stanbol		2/12/2018 3:52 PM
☐	Babelify		2/12/2018 4:10 PM
☐	Deutsches Textarchiv (DTA)	multiple creators	2/18/2018 12:57 PM
☐	Software - The Stanford Natural Language Processing Group	Stanford NLP Group	2/12/2018 2:52 PM
☐	TEI-TCF encoder+decoder	Jurish	2/13/2018 7:23 PM
☐	The TCF Format	Tübingen CLARIN-D team	2/12/2018 2:52 PM
☐	TreeTagger	Schmid	2/12/2018 2:53 PM
☐	WebLicht		2/12/2018 1:21 PM
☐	WebLicht: Getting Started	multiple authors	11/6/2017 4:55 PM

Architecture RESTful

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



- Architecture flexible, facile à déployer et à maintenir
- Des blocs indépendants communiquent via services REST

Pérennisation

- Code **Open-source** et licence GPL
- Les données décrites en TEI sont disponibles sur GitHub sous la licence **CC-BY**
- Les ressources bibliographiques sont stockées dans une bibliothèque **Zotero**.
- **Double hébergement:**
 - <http://ssk.parthenos-project.eu>
 - <http://ssk.huma-num.fr>
- Liens avec l'infrastructure **DARIAH** pour la curation des scénarios et ressources

Le futur

- ◉ Accès facilité à la bibliothèque Zotero (capacité de récupérer les données en divers formats)
- ◉ Compte utilisateur riche :
 - gérer des **bookmarks**
 - **Contribuer** au SSK directement via l'interface:
 - création
 - édition
 - adaptation (customization)
- ◉ Interface **multilingue**



merci



<http://ssk.huma-num.fr>

<http://ssk.parthenos-project.eu>



<http://ssk.readthedocs.io>



ssk@inria.fr



Pour créer un scénario

- Documentation (en anglais)
 - https://ssk.readthedocs.io/en/latest/0_userDoc.html#how-to-create-a-scenario-for-the-ssk
 - https://ssk.readthedocs.io/en/latest/1_ssktei.html
- Scénarios et étapes en TEI:
 - <https://github.com/ParthenosWP4/SSK/blob/master/steps/>
 - <https://github.com/ParthenosWP4/SSK/blob/master/scenarios/>
- Vocabulaires: <https://ssk.parthenos-project.eu/ssk/#/glossary>