



Heritage(s) Digital Ecosystem

Focus on Data platform Fair Data processing chain



Perrine Thuringer
Damien Vurpillot

FAIR Heritage

Digital Methods, Scholarly Editing and
Tools for Cultural and Natural Heritage



1. **Heritage(s) Digital Ecosystem Presentation**
2. **Introduction to Données platform**
3. **Fair Data processing chain**
 - 3.1. **Importance of Metadata**
 - 3.2. **Data management plan definition**
 - 3.3. **Dataset submission**
 - 3.4. **Dataset indexation**
 - 3.5. **Dataset publication on the Web and Web of Data**
 - 3.6. **Metadata Enrichment**
4. **Données platform Demonstration**

- **Heritage(s)** is a project launched in 2017 by Intelligence des Patrimoines research program.
- The main goal is to foster collaboration between researchers producing data from transdisciplinary scientific fields with a focus on cultural and natural heritage and facilitate knowledge dissemination.
- Heritage(s) Digital Ecosystem is :
 - a **suite of digital resources, tools and services**
 - **designed with researchers for researchers** to meet the requirements of research data acquisition, publication and dissemination.
 - **accessible online** from a unique portal

Heritage(s) Digital Ecosystem Presentation

Researchers Requirements

- Provide researchers with **high level scientific equipment to acquire quality data** without financial investment.
- **Guide researchers through the process of submission and publication of research data** on the Web, following a data management plan for Open Data as required by many funding organizations.
- **Enhance visibility and interoperability of heritage data** produced by research projects by making them findable, accessible, interoperable and reusable (FAIR principles) and allow cross-queries between different project datasets and open data flows.
- **Enforce sustainability of research data** by guaranteeing a secured and long-term storage in a certified data warehouse hosted by University of Tours/CNRS and harvesting compliance with institutional open data aggregation platforms.
- **Disseminate new knowledge to general public** through production of digital mediation applications.

Heritage(s) Digital Ecosystem Presentation

Researchers Requirements

- Provide researchers with high level scientific equipment to acquire quality data without financial investment.
- Guide researchers through the process of submission and publication of research data on the Web, following a data management plan for Open Data as required by many funding organizations.
- Enhance visibility and interoperability of heritage data produced by research projects by making them findable, accessible, interoperable and reusable (FAIR principles) and allow cross-queries between different project datasets and open data flows.
- Enforce sustainability of research data by guaranteeing a secured and long-term storage in a certified data warehouse hosted by University of Tours/CNRS and harvesting compliance with institutional open data aggregation platforms.
- Disseminate new knowledge to general public through production of digital mediation applications.

Equipements

Données

Médiation

HeritageS

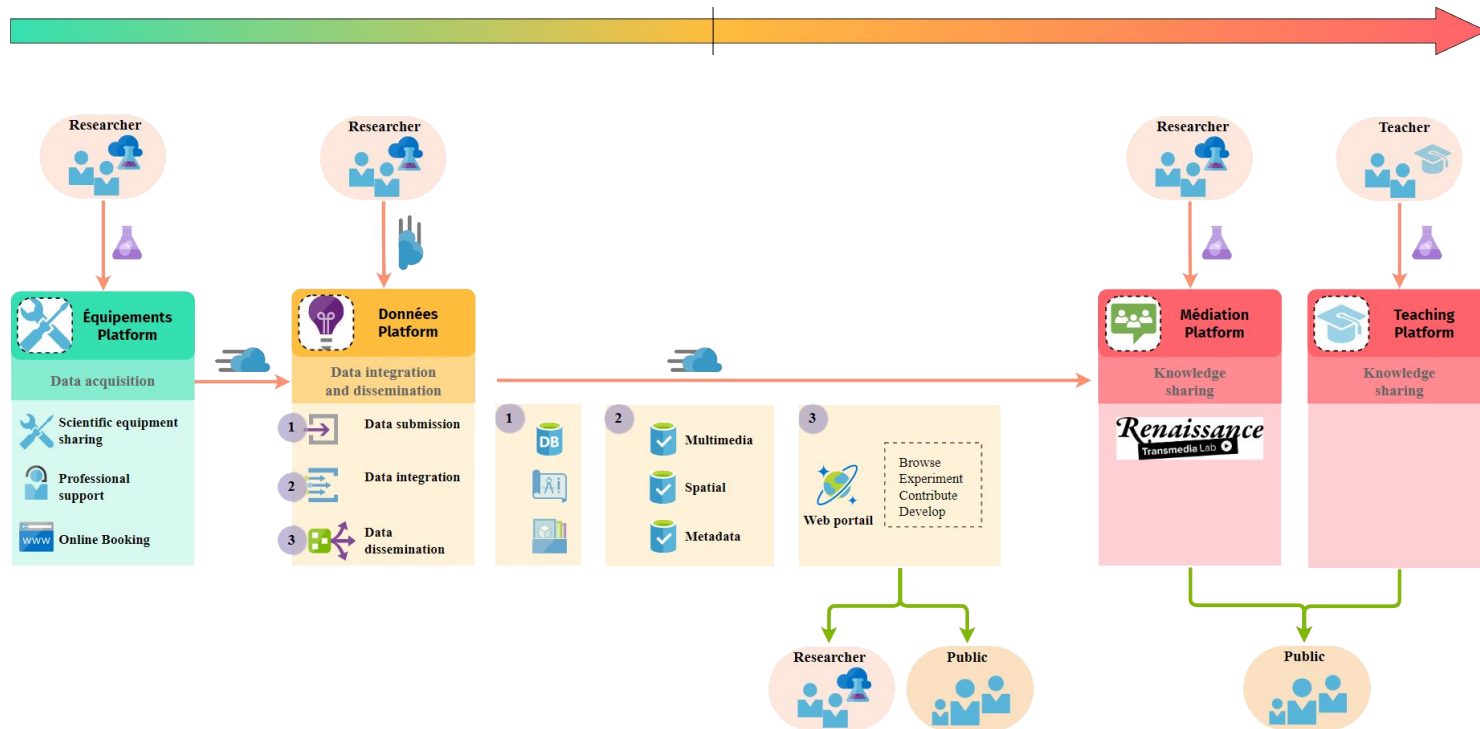
the digital ecosystem

supported by the
Centre-Val de Loire region

Heritage(s) Digital Ecosystem Presentation

Architecture

Heritage(s) Digital Ecosystem



Données platform

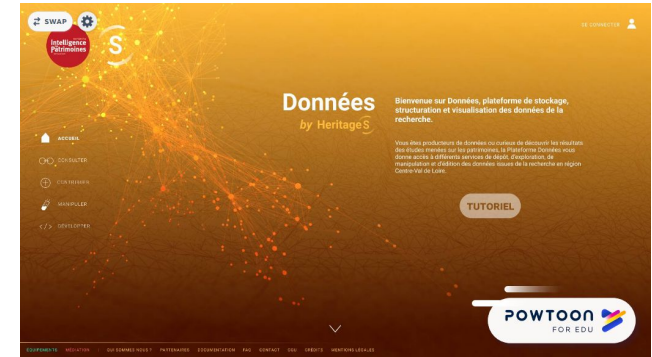
Platform for research data submission, integration and publication of research data

Target :

- Scholars
- Socioeconomic agents
- Interested visitors

Contributors :

- Scholar Data producers



Données platform

Data processing chain is articulated in 5 steps :

1. **Data management plan definition**
2. **Dataset submission**
3. **Dataset indexation**
4. **Dataset publication**
5. **Metadata enrichment**

Importance of Metadata

The data processing chain is based on the descriptive information that accompanies the deposited data sets:
metadata.

Metadata

- allow to **automatically extract information** about the source datasets
- allow to **reconstruct the relationships** between datasets.
- help **making data interoperable**
- make it **easier for indexing** datasets
- **give more visibility** when datasets are published
- help **preserving data intellectual property**

To do this, a common definition and classification of metadata is necessary.

Importance of Metadata

Common definition and classification of metadata for digital heritage research products already exists in standard controlled vocabularies such as :

- **Metadata Schemas** : Dublin Core Metadata Initiative, Data Cite Metadata Initiative, METS, ...
- **Thesaurii** : EuroVoc, PACTOLS, ...
- **Ontologies** : DataCite, CIDOC CRM, CRMgeo, FOAF, BNF-Onto, FRBROO ...

Source Dataset metadata must be mapped to these controlled vocabularies terms

1. Data management plan (DMP) definition

DMP redaction allows data producer to provide :

- **Administrative metadata** on the project and research products : actors, purpose,
- **Legal metadata** about the research products related to intellectual property : rights, data reuse licence...
- **Unique identifiers (DOI)** for the research products datasets in order to make them identifiable and findable on the Web

2. Dataset submission

Once DMP has been defined, submitted and validated by the legal desk of the contributor's institution, a form invites the contributor to :

- **load the datasets** of the research products
- add **generic metadata for indexing** such as : scientific domain, spatial and temporal coverage ...



SE CONNECTER 

Contribuer

Contribuez à la plateforme Données en déposant vos jeux de données. Pour vous accompagner dans votre démarche, un tutoriel est accessible à chaque étape de ce dépôt.

Pour accéder à ce service, vous devez posséder un statut de contributeur et travailler au sein d'un organisme de recherche membre fondateur du programme ARD Intelligence des Patrimoines.

 ACCUEIL
  TUTORIELS
  CONSULTER
  **CONTRIBUER**
 MANIPULER
  DÉVELOPPER

 Tutoriel
  Que deviennent vos données ?

 Rédigez votre Plan de Gestion de Données
  Déposez vos données



[ÉQUIPEMENTS](#)
[MÉDIATION](#)
[QUI SOMMES NOUS ?](#)
[PARTENAIRES](#)
[DOCUMENTATION](#)
[FAQ](#)
[CONTACT](#)
[CGU](#)
[CRÉDITS](#)
[MENTIONS LÉGALES](#)

3. Datasets indexation

When datasets are submitted, an **automated indexation** is run by the system. It comprises extraction and normalization of metadata from the DMP, submission form and datasets contents :

- **Legal and generic** metadata from the Data Management Plan and submission form are mapped to the standard metadata models : Dublin Core, DCAT, Data Cite, FOAF and BNF.
- **Technical metadata** from the datasets multimedia resources are mapped to multimedia formats compatible with the Exiftool library..
- **Spatio-temporal metadata** from both the submission form and datasets multimedia resources are mapped to Geonames and GeoSparql.

Datasets indexation ends with generation of an XML file compiling normalized metadata of the whole project.

```

lessbig.xml  big.xml  <?xml version="1.0" encoding="UTF-8"?>.xml
Users > Perrine > Documents > Développement > Talend_Heritages > big.xml

209759 <frapo:Project>
209760 <dc:title>PRP - Projet de Recherche Patrimonial</dc:title>
209761 <cris:Coordinator>
209762 <foaf:name>MICHELBLANC</foaf:name>
209763 <foaf:mbox>blanc@cnrss.fr</foaf:mbox>
209764 <foaf:Organization/>
209765 </cris:Coordinator>
209766 <ipato:hasParticipant>
209767 <foaf:Person>
209768 <foaf:name/>
209769 <foaf:mbox/>
209770 <foaf:Organization/>
209771 </foaf:Person>
209772 </ipato:hasParticipant>
209773 <frapo:FundingAgency/>
209774 <ipato:scientificDomain/>
209775 <dc:subject/>
209776 <ipato:is_about>Le projet PRP - Projet de Recherche Patrimonial a pour objectif de transformer la zone du Loiret en vaste terrain de d'expérimentation
209777 <dc:description>Produits dérivés des relevés LIDAR sur le Loiret</dc:description>
209778 <dc:Catalog>
209779 <dc:Dataset>
209780 <dc:title>10.23638/haXq1QJlVr</dc:title>
209781 <ipato:dataCite_identifier>768</ipato:dataCite_identifier>
209782 <ipato:dataIpat_identifier>http://eurovoc.europa.eu/1278</ipato:dataIpat_identifier>
209783 <dc:subject>- Type : Collection - Périmètre : Zone A et B du Loiret pour octobre 2020. Il s'agit de relevés topo-bathymétriques. - Echelle : Les dor
209784 <dc:description>fra</dc:description>
209785 <dc:language/>
209786 <dc:format>Université de Tours avec le concours de la Région Centre</dc:format>
209787 <dc:date>Les données sont ouvertes à tous, compte-tenu du volume de données, une notice récapitulative et un système de prévisualisation sur le w
209788 <dc:creator>
209789 <foaf:Person>
209790 <foaf:name/>
209791 </foaf:Person>
209792 </dc:creator>
209793 <dc:publisher>2993024</dc:publisher>
209794 <dc:rights>47.996964</dc:rights>
209795 <dbo:PopulatedPlace>2.732611</dbo:PopulatedPlace>
209796 <dc:spatial/>
209797 <go:lat/>
209798 <go:long/>
209799 </dc:Dataset>
209800 </dc:Catalog>
209801 </frapo:Project>
209802 </Project>
  
```

L 2421, col 22 Espaces : 2 UTF-8 LF XML

4. Dataset publication

Once datasets are indexed, they are automatically published in data catalogs of the Data platform :

- **Catalog of records**
- **Multimedia data catalog**
- **Geographical data catalog**

4. Dataset publication

Once datasets are indexed, they are automatically published in data catalogs of the Data platform :

- **Catalog of records** : provides access to the records of the datasets
 - Each research product is described within a notice which compiles its generic, administrative and legal metadata, accompanied, if the DMP allows it, by a download link or a reference to the web page of the project.
 - If the dataset contains multimedia resources, the notice includes a link to the associated resources in the multimedia data catalog.
 - If the dataset contains geographical informations, the notice includes a link to the associated resources in the geographical data catalog.
- **Multimedia data catalog**
- **Geographical data catalog**

4. Dataset publication

Once datasets are indexed, they are automatically published in data catalogs of the Data platform :

- **Catalog of records**
- **Multimedia data catalog:** describes each multimedia resource (image, audio, video, or 3D model) of a dataset
 - The resource is described in an notice compiling its technical metadata
 - The notice contains an online overview from a standard conversion of the resource (Jpeg, MP3, MP4 or GLTF) and a link to the related research product in the Catalog of records.
 - If the DMP allows it, the notice also contains a download link of the original version of the resource.
- **Geographical data catalog**

4. Dataset publication

Once datasets are indexed, they are automatically published in data catalogs of the Données platform :

- **Catalog of records**
- **Multimedia data catalog**
- **Geographical data catalog:** describes geospatial information from each natively spatialized research product
 - Geospatial information can be displayed on different base maps (dynamic or static).
 - When data have attributes, it is possible to display them in a tooltip.
 - These tooltips allow, when relevant, to make the gateway to the Catalog of records and Multimedia data catalog.



[SE CONNECTER](#) 

Consulter

Retrouvez ici l'ensemble des jeux de données déposés sur la plateforme et publiés dans trois catalogues de données suivants.

 ACCUEIL

 TUTORIELS

 **CONSULTER**

 CONTRIBUER

 MANIPULER

 DÉVELOPPER

| CATALOGUE DE NOTICES



Accédez aux notices génériques des jeux de données comprenant les informations générales, administratives et juridiques du projet de recherche concerné.

| CATALOGUE DE DONNÉES GÉOGRAPHIQUES



Visualisez géographiquement les jeux de données grâce à une carte présentant leur couverture géographique, leur lieu de conservation et/ou d'acquisition, etc.

Un système de points d'intérêt vous permettra ainsi d'accéder à une description du jeu de données concerné et aux liens vers les catalogues de notices et de données multimédia des projets de recherche traités.

| CATALOGUE DE DONNÉES MULTIMÉDIA



Accédez aux ressources multimédia présentes dans certains jeux de données et leurs métadonnées associées. Des liens vers le catalogue de notices du projet de recherche concerné vous seront proposés.



[ÉQUIPEMENTS](#) | [MÉDIATION](#) | [QUI SOMMES NOUS ?](#) | [PARTENAIRES](#) | [DOCUMENTATION](#) | [FAQ](#) | [CONTACT](#) | [CGU](#) | [CRÉDITS](#) | [MENTIONS LÉGALES](#)

4. Dataset publication

The whole normalised metadata is also published on the **Web of Data**.

Thanks to the **generation of RDF triples** from normalized metadata, stored in the Données platform triplestore and whose SPARQL entry point is exposed and accessible online.

5. Metadata enrichment

To complete the automated processing of research products, a complementary step of metadata manual enrichment is proposed to the contributor.

It concerns the schemas of textual resources from the source datasets (XML, CSV, TSV, SQL, Excel etc.) and consists of two activities:

- **alignment of schemas metadata terms to standard thesaurii terms**
- **mapping of textual resources schemas to the ontology of the Données platform**

5. Metadata enrichment

To complete the automated processing of research products, a complementary step of metadata manual enrichment is proposed to the contributor.

It concerns the schemas of textual resources from the source datasets (XML, CSV, TSV, SQL, Excel etc.) and consists of two activities:

- **alignment of schemas metadata terms to standard thesaurii terms:** this activity makes it possible to create keywords which will be added to the metadata of the dataset and thus will give it more visibility when searching or filtering keyword datasets. It also generates semantic annotations in RDF format which will feed the triplestore and give more visibility to the dataset on the Web of Data.
- **mapping of textual resources schemas to the ontology of the Données platform**

5. Metadata enrichment

To complete the automated processing of research products, a complementary step of metadata manual enrichment is proposed to the contributor.

It concerns the schemas of textual resources from the source datasets (XML, CSV, TSV, SQL, Excel etc.) and consists of two activities:

- **alignment of schemas metadata terms to standard thesaurii terms**
- **mapping of textual resources schemas to the ontology of the Données platform:** this activity generates RDF triples describing the textual resources of the datasets that will feed the triplestore of the platform and thus give more precision to the datasets on the Web of Data

SE CONNECTER

S

Manipuler

Manipulez vos données, en interrogeant et enrichissant leurs métadonnées afin d'améliorer leur visibilité.

ACCUEIL

TUTORIELS

CONSULTER

CONTRIBUER

MANIPULER

DÉVELOPPER

INTERROGEZ VOS MÉTADONNÉES

ENRICHISSEZ LES MÉTADONNÉES

SparNatural

Développé par Sparma, cet outil permet de construire intuitivement – sans une connaissance approfondie du langage – vos requêtes SPARQL à partir d'une interface graphique dédiée. Afin de guider vos premiers pas avec SparNatural, un tutoriel et des requêtes utiles vous sont proposés.

OnaGUI version web

Outil d'alignement des termes des schémas vers des termes issus de thésauris standards – il vous permet de créer des mots-clés qui s'ajouteront aux métadonnées du jeu de données pour leur donner plus de visibilité lors d'une recherche ou d'un filtrage des jeux de données. OnaGUI génère également des annotations sémantiques au format RDF qui alimenteront le triplstore et donnera plus de visibilité au jeu de données sur le Web des données.

3M

Outil de mapping des schémas vers l'ontologie de la Plateforme Données – il permet de générer des triplets RDF décrivant les ressources textuelles des jeux de données qui alimenteront le triplstore de la plateforme et donneront ainsi plus de précisions aux jeux de données sur le Web des données.

EQUIPEMENTS MÉDIATION | QUI SOMMES NOUS ? PARTENAIRES DOCUMENTATION FAQ CONTACT CGU CRÉDITS MENTIONS LÉGALES

Données
by HeritageS

Bienvenue sur Données, plateforme de stockage, structuration et visualisation des données de la recherche.

Vous êtes producteurs de données ou curieux de découvrir les résultats des études menées sur les patrimoines, la Plateforme Données vous donne accès à différents services de dépôt, d'exploration, de manipulation et d'édition des données issues de la recherche en région Centre-Val de Loire.

TUTORIEL

POWTOON
FOR EDU

EQUIPEMENTS MEDIATION / QUI SOMMES NOUS ? PARTENAIRES DOCUMENTATION FAQ CONTACT COU CRÉDITS MENTIONS LÉGALES

Thank you for your attention

Questions ?

Benoist PIERRE

Directeur du CESR et du Programme ARD Intelligence des Patrimoines

benoist.pierre@univ-tours.fr - 06 24 39 60 78

Morgane CHAUMIER

Chargée de mission, PhD

morgane.chaumier@univ-tours.fr – 02 47 36 77 99

www.intelligencedespatrimoires.fr

intelligencedespatrimoires@univ-tours.fr

FAIR Heritage

Digital Methods, Scholarly Editing and
Tools for Cultural and Natural Heritage



LE STUDIUM
Loire Valley
Institute for Advanced Studies

