

# PARTHENOS

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

## REPORT ON STANDARDISATION (FINAL)

Laurent ROMARY  
Charles RIONDET  
Sophia SOTIROPOULOU  
and all WP4 partners

19 March 2019



PARTHENOS is a Horizon 2020 project funded by the European Commission. The views and opinions expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.



HORIZON 2020 - INFRADEV-4-2014/2015:

Grant Agreement No. 654119

PARTHENOS

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

REPORT ON STANDARDISATION (FINAL)

Deliverable Number	D4.4
Dissemination Level	Public
Delivery date	06 March 2019
Status	Final
Author(s)	Laurent Romary (Inria)
	Charles Riondet (Inria)
	Lionel Tadjou (Inria)
	Sophia Sotiropoulou (FORTH)
	Dorian Seillier (Inria)

NB: the authors of the scenarios are  
named under each scenario title.



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	4.4
Document title	REPORT ON STANDARDISATION (FINAL)
Author(s)	Laurent Romary (Inria), Charles Riondet (Inria), Lionel Tadjou (Inria), Sophia Sotiropoulou (FORTH), Dorian Seillier (Inria). NB: the authors of the scenarios are named under each scenario title.
Dissemination level/distribution	public

#### Document History

Version/date	Changes/approval	Author/Approved by
Final 11/03/2019	draft Final review	S. Bassett, PIN.
Final 15/03/2019	draft Final amendments after review	S. Sotiropoulou, FORTH
Final 19/03/2019	check Final version.	S. Bassett, PIN.

## Table of contents

1	Executive Summary.....	1
2	Introduction .....	2
3	Research scenarios .....	4
3.1	Linguistic annotation of corpora .....	5
3.2	Digitizing textual material.....	12
3.3	Production and processing of 3D objects .....	17
3.4	Preserving 3D digital objects .....	24
3.5	Tracking the dissemination of a specific work.....	30
3.6	Create a dictionary in TEI.....	37
3.7	Research standards for scholars .....	41
3.8	Creating Metadata.....	46
3.9	Encode and modelize field surveys for their online dissemination.....	53
3.10	Perform Statistical Analysis on Historical and Contemporary Occupations.....	60
3.11	Prosopography research.....	65
3.12	Creating interoperable TEI text resources with the DTA 'Base Format' (DTABf).....	71
3.13	Extract textual content from images .....	79
3.14	Collaborative Digital Edition of a Musical .....	87
3.15	Aggregate different datasets to obtain a single reference resource .....	94
3.16	Aggregate different datasets using RUBRICA .....	98
3.17	Creation of a TEI-based corpus.....	104
3.18	Project-centered EAD customization .....	111
3.19	Raman spectrometry for the analysis and identification of pigments on cultural heritage objects	118
3.20	In situ X-Ray fluorescence analysis for the characterisation of materials on cultural heritage objects	129
3.21	In situ FTIR Analysis for the characterisation of materials on archaeological or cultural heritage objects.....	137
3.22	Macro-X-ray fluorescence mapping of materials on cultural heritage objects, .....	146
3.23	Multispectral imaging for the in-situ characterization of painted works of art.....	153
3.24	Documenting application of lasers in conservation/restoration.....	159
3.25	Standardizing "Laser Induced Breakdown Spectroscopy" analysis of glasses .....	163
4	Related Heritage Science standardization activities.....	168
5	Annex 1. Heritage Science specific activities .....	173



6	Annex 2. Creation of a standards and methods Knowledge base .....	199
---	---	-----

# 1 Executive Summary

The present report reflects the final stage of the definition of the Standardization Survival Kit (SSK) within Work Package 4 of the PARTHENOS project. Based on the draft report on standardization (Romary et al. 2017) which presented a review of the activities that have to be carried out to provide support to researchers in using specific standards, this final report presents these activities in the context of complete research scenarios, created by domain experts and designed to be displayed within the Standardization Survival Kit web application, visible at <http://ssk.huma-num.fr>.

To date, there are twenty five complete scenarios documented in this report covering the following disciplines:

- Social Anthropology and Ethnology (1)
- Archaeology and Prehistory (2)
- Art and Art History (7)
- History (16)
- Linguistics (4)
- Literature (2)
- Cultural Heritage and Museology (10)
- Musicology and Performing Arts (1)
- Philosophy (1)
- Sociology (3)
- Methods and Statistics (1)

Each scenario is described in Section 3 according to a pre-defined format outlined in the Introduction (Section 2). The Heritage Sciences scenarios, due to their specific nature, are presented at the end of this section, preceded by a contextual introduction. Section 4 presents further standardization activities related to this specific sector and also the Knowledge base of standards created during the development of the SSK.



## 2 Introduction

The present report reflects the final stage of the definition of the Standardization Survival Kit (SSK) within Work Package 4 of the PARTHENOS project. Based on the draft D4.2 Report on Standardization (Romary et al. 2017), which presented a review of the activities that had to be carried out to provide support to researchers in using specific standards, this final report presents these activities in the context of complete research scenarios, created by domain experts and designed to be displayed within the Standardization Survival Kit web application, visible at <http://ssk.huma-num.fr>.

The draft deliverable (D4.2) was organized into three sections reflecting three domains of standardization that have a specific role in the research process - namely Community standards, references resources, protocols and procedure for the Cultural Heritage domains. This final report presents the produced scenarios in alphabetical order. This choice reflects the fact that one research scenario, representing complete research processes, may refer to the three domains mentioned previously. Note that many of the resources within the SSK tool are accompanied by a description or abstract. For the purposes of brevity, these descriptions are not included in this report since they are available at the given URL for each resource (which may appear in several scenarios).

These scenarios follow the SSK data model and are presented in three main parts:

- Scenario metadata
- Title
- Authors
- Link towards the scenario online
- Scenario description
- Paragraphs describing the scientific problem and describe the solution put in place
- Controlled vocabularies: Disciplines, Research objects and research techniques.  
The vocabularies are described in the Deliverable 4.3 (Standardization Kit Final) which documents the release of the Standardization Survival Kit web application (Riondet et al. 2018)
- An illustration



- Steps, metadata and description

NB: The same Steps may be present in more than one scenario, which means that the same information is repeated several times in the report, for readability reasons.

- Title
- Authors
- Paragraphs describing the purpose of the step, 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.
- Resources, i.e. a list of bibliographical references giving access to standardized tools, services or documents helpful for the task completion.



### 3 Research scenarios

The following scenarios represent the status of the SSK content at the time of the report (i.e. end of February 2019). They are automatically exported from the SSK GitHub repository so they may not be exactly equivalent with the scenarios visible in the SSK application at <http://ssk.huma-num.fr>.

Important caveat: The Heritage Sciences scenarios, due to their specific nature, are presented at the end of this section, preceded by a contextual introduction.

### 3.1 Linguistic annotation of corpora

Matej Ďurčo (OEAW-ACDH), Tom Gheldof (KU Leuven, DARIAH-BE), Hannes Pirker (OEAW-ACDH), and Gerlinde Schneider (ZIM-ACDH)

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_linguisticAnnotationOfCorpora](http://ssk.huma-num.fr/#/scenarios/SSK_sc_linguisticAnnotationOfCorpora)

This scenario explains the steps to take to annotate a corpus in order to conduct linguistic and statistical analysis based upon it. This scenario provides general information for people starting with linguistic annotation. The aim is to provide a generic scenario with no specific tool in mind: tools may be referred to but without specifying how to use them. There are various tools and frameworks to perform the steps in this scenario that can be used depending on the language/s being working with or the programming environment. A number of tool-boxes for Natural Language Processing (NLP) exist, which are able to perform several of the annotation steps in an integrated way. These resources are listed below under "Using an existing NLP pipeline". The next step, after performing the procedures described in this scenario, usually is to put the annotated corpus into a corpus query engine to query and analyse it based on its annotations. Again, some popular query engines already provide a built-in pipeline that is able to perform the basic processing steps in one go, taking most of the burden off the user.

#### Scenario descriptive terms

##### Discipline

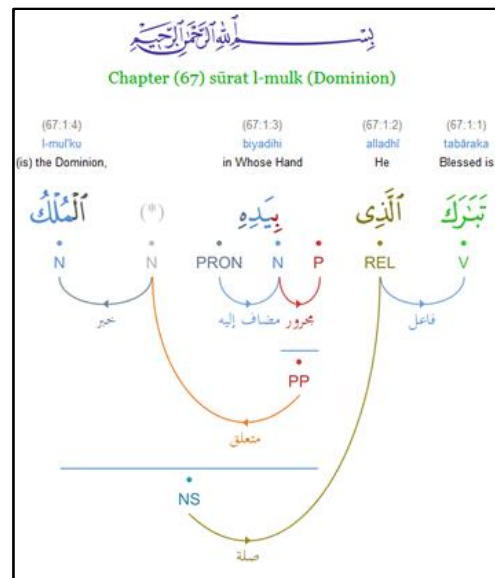
- Linguistics

##### Research Objects

- Text

##### Research techniques

- Cluster analysis
- Collocation analysis
- Concordancing
- Encoding
- Named entity recognition
- Pattern recognition
- Sequence Alignment
- POS-tagging





## **1- Tokenizing**

Tokenizing refers to the task of splitting a text into identifiable units – the tokens. Or to state it in less formal terms: to split the text into words and punctuation marks. Tokenization is a prerequisite for basically all other forms of linguistic analysis, as it provides the basic units on which to operate. Though some "higher level" NLP tools allow the use of plain text as input i.e. deal with the necessary tokenization themselves, many tools still rely on tokenized text. Tokenizers are typically quite simple programmes: the basic heuristic is to split the text at white spaces and punctuation characters. An important factor in many western languages is to properly distinguish punctuation characters which are actually part of an abbreviation as in "e.g.". The quality of many tokenizers thus typically relies on the application of language specific abbreviation lists and regular expressions for correctly identifying sequences of characters which form a token. E.g. web and e-mail addresses, date and time etc.

### ***Step descriptive terms***

Research activity

- annotating

Standards

- XML
- TEI
- TCF
- LAF

### ***Resources***

- Unitok – Corpus tools

Link: <http://corpus.tools/wiki/Unitok>

## **2-Part-of-speech tagging**

Part-of-speech tagging (POS or PoS tagging) refers to assigning grammatical information to the units of the text. There are defined tagsets commonly used for individual languages or language groups. Particularly noteworthy is the Universal Dependencies (UD) initiative which aims at mapping the idiosyncratic tagsets to one harmonized vocabulary/tagset. Most current tools used for PoS tagging rely on machine learning technologies.

### ***Step descriptive terms***

**Research activity**

- annotating

**Standards**

- XML
- TEI
- TCF
- LAF
- tagset

**Resources**

- German Tagsets | Institut für Maschinelle Sprachverarbeitung | Universität Stuttgart  
Link: <http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/GermanTagsets.html>
- Universal POS tags  
Link: <http://universaldependencies.github.io/docs/u/pos/all.html>
- OLiA ontologies  
Link: <http://www.acoli.informatik.uni-frankfurt.de/resources/olia/#olia>
- TreeTagger  
Author(s): Helmut Schmid  
Link: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

**3- Lemmatizing**

Lemmatizing refers to the task which determines the lemma (also known as canonical form, dictionary form, or citation form) for each word form present in the text. It relies on the information extracted in the tokenization and POS tagging steps, and is often integrated into the POS tagging tool. The relevance of the lemma information becomes obvious for morphologically richer languages, where the possibility to simply refer to the canonical form in order to identify all its inflected forms is a real asset. Generally speaking, most lemmatizers rely on some sort of dictionary or word list in the background where mapping between word forms and its lemmas are defined. Because these dictionaries are not able to cover all words of a language, missed cases in the lemmatization are to be expected.



### ***Step descriptive terms***

Research activity

- annotating

Standards

- XML
- TEI
- TCF
- LAF

### ***Resources***

- TreeTagger

Link: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

- Stanford NLP Group

Link: <https://nlp.stanford.edu/software/>

## ***4- Stemming***

Stemming refers to the process of reducing inflected or sometimes derived word forms to their word stem, base or root form. The stem does not need to be identical to the morphological root of the word, nor has it to be a proper word form at all. Stemming can be viewed as a simplified version of lemmatization. In some respects, stemming can be seen as a poor man's version of lemmatizing, c.f. simple affix stripping as the most traditional approach used. However, stemming can relate different lemmata to each other via a shared root, something lemmatization typically does not provide for. Stemming thus is still widely used in the context of information retrieval and by search engines because it allows for generalizations on related words.

### ***Step descriptive terms***

Research activity

- Annotating

**Standards**

- XML
- TEI
- TCF
- LAF

**Resources**

- TreeTagger  
Link: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Stanford NLP Group  
Link: <https://nlp.stanford.edu/software/>

**5- Named-entity recognition**

Named-entity recognition (NER) identifies named entities/proper names in a text and classifies them into predefined classes (e.g. person names, place names, date and time etc.).

**Step descriptive terms****Research activity**

- annotating

**Standards**

- XML
- TEI
- TCF
- LAF

**Resources**

- Stanford NER  
Link: <https://nlp.stanford.edu/software/CRF-NER.shtml>

**6- Manual annotation**

Sometimes, a research question requires a very specific annotation that cannot be covered by the automatic annotation provided by NLP tools and frameworks. In this case, annotation has to be performed semi-automatically or entirely manually. This is often done in



collaborative work and often requires several correction cycles. Fortunately, there are tools that support this work and also create formats that can be easily processed later on.

### ***Step descriptive terms***

Research activity

- annotation

Standards

- XML
- TEI
- TCF
- LAF

### ***Resources***

- CLARIN-D WebAnno

Author(s): Computer Science Department of the Technische Universität Darmstadt

Link: <https://webanno.sfs.uni-tuebingen.de/>

- TokenEditor

Link: <https://clarin.oeaw.ac.at/tokenEditor/>

## ***7- Using an existing NLP pipeline***

In practice, the above steps are combined to pipelines implemented in various NLP frameworks.

### ***Step descriptive terms***

Research activity

- annotation

Standards

- XML
- TEI
- TCF
- LAF



## **Resources**

- WebLicht  
Link: [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page)
- WebLicht: Getting Started, 2017  
Author(s): multiple authors  
Link: [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting\\_Started](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting_Started)
- NLTK book  
Link: <http://www.nltk.org/book/>
- OpenNLP  
Link: <http://opennlp.apache.org/index.html>
- The Stanford NLP  
Link: <https://nlp.stanford.edu/software/>

## 3.2 Digitizing textual material

*Frank Uiterwaal (NIOD-KNAW) and Reinier de Valk (DANS-KNAW)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_digitization](http://ssk.huma-num.fr/#/scenarios/SSK_sc_digitization)

The digitization process of original objects lies at the core of many activities in the digital humanities field. The way in which a source is digitised, determines to a large extent what kind of information can be gathered from this virtual replica. While there are many complimentary processes which can enrich the object (adding metadata, augmented layers etc.), these data will build upon the digital facsimile.

### **Scenario descriptive terms**

#### Disciplines

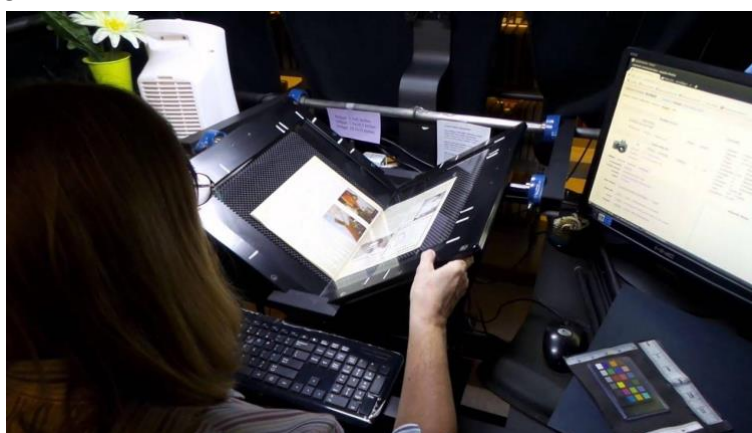
- Social Anthropology and Ethnology
- History
- Philosophy and Sociology of Sciences
- Library and Information Sciences

#### Research objects

- Data
- Image
- Manuscript
- Sheet Music
- Text Bearing Objects

#### Research techniques

- Scanning



### **1- Assess legal issues**

Before starting work with material, looking into its legal status is a strict requirement. By doing this, you find out what the originator of the material, be it a picture, a document or an object, allows you to do with the object in terms of (re-)use.

### **Step descriptive terms**

#### Research activity

- Assessing

**Resources**

- Open Data in a Big Data World. An international accord, 2015  
Link: [http://www.science-international.org/sites/default/files/reports/open-data-in-big-data-world\\_long\\_en.pdf](http://www.science-international.org/sites/default/files/reports/open-data-in-big-data-world_long_en.pdf)
- The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations, 2015  
Link: <https://zenodo.org/record/33830#.WnmipoliEdW>

**2- Determine scanning method**

A thorough examination of the object provides insight in which scanning method is the most appropriate one given the type of material, its dimensions and fragility.

**Step descriptive terms**

Research activity

- Imaging

**Resources**

- Digitizing Entire Collections: Chapter 4, For Scanning Technicians, 2015  
Author(s): Barbara Aikens  
Link: <https://www.aaa.si.edu/documentation/digitizing-entire-collections-chapter-4-for-scanning-technicians>

**3- Define imaging specifications**

Setting requirements for the digital facsimile helps to make sure that it is suitable for intended purposes. This includes defining image file format(s), resolution, metadata, naming protocol, colour schema, etc.

**Step descriptive terms**

Research activity

- Imaging



## **Resources**

- Metamorfoze Preservation Imaging Guidelines, 2012

Link:

[https://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie\\_documenten/Metamorfoze\\_Preservation\\_Imaging\\_Guidelines\\_1.0.pdf](https://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie_documenten/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf)

- Recommended Formats Statement - Resources

Author(s): Library of Congress

Link: <https://www.loc.gov/preservation/resources/rfs/>

- Recommendations for digitisation projects

Author(s): Succeed project

Link: <https://www.digitisation.eu/training/recommendations-for-digitisation-projects/>

## **4- Image the documents**

Now that the scanning method and the requirements are set, the imaging procedure can start.

### **Step descriptive terms**

Research activity

- Imaging

## **Resources**

- Digitisation at The National Archives, 2015

Link: <http://www.nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>

## **5- Ensure quality control**

After the imaging has been completed, it is best practice to verify whether the actual digital version aligns with the requirements set earlier. This allows for timely adjustment or re-doing (parts of the) the procedure before the object returns to its place in the archive.

### **Step descriptive terms**

Research activity

- Imaging

**Resources**

- Best Practices in Digitization. Quality Control and Image Processing  
Author(s): ETH-Bibliothek  
Link: <http://www.library.ethz.ch/en/ms/DigiCenter/Best-Practices-in-Digitization/Digitization-Process/Quality-Control-and-Image-Processing>

**6- Store master files**

Storing the master files in a trusted location is a good way to prevent having to rescan an image when requirements change over time (due to new research needs or improved technology which sets higher expectations) or when better post processing methods become available.

**Step descriptive terms**

Research activity

- Archiving

**Resources**

- Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report, 2002  
Link: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
- Preservation and Archiving Special Interest Group (PASIG), Change (Y)our World  
Author(s): PASIG group  
Link: <https://preservationandarchivingsig.org/>

**7- Post process files**

Software can be used to enhance the quality of the scanned image. This could be done by, for instance, reducing noise or by de-skewing the picture, making it easier to read (both by man and by machine).

**Step descriptive terms**

Research activity

- Clean up



## **Resources**

- Post-processing of Scanned 3D Surface Data, 2004  
Author(s): Tim Weyrich, Mark Pauly, Richard Keiser, Simon Heinzle, Sascha Scandella, et al.

## **8- Store edited files**

After first storing the master file, the next step to store the edited file as well.

## **Step descriptive terms**

Research activity

- Archiving

## **Resources**

- Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report, 2002  
Link:  
<https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>

### 3.3 Production and processing of 3D objects

*Hélène Gautier (Huma-Num/CNRS), Adeline Joffres (Huma-Num/CNRS), Paulin Ribbe (Huma-Num/CNRS), Mehdi Chayani (3D-SHS Huma-Num's consortium), Xavier Granier (3D-SHS Huma-Num's consortium and Sarah Tournon-Valiente (3D-SHS Huma-Num's consortium).*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_production3DObjectsScenario](http://ssk.huma-num.fr/#/scenarios/SSK_sc_production3DObjectsScenario)

The project aim is to digitize objects at different scale from an artefact to cultural heritage site to help scientific community to visualize, analyze, simulate and synthesize the result of their research. It allows for better comprehension of the case studies. The 3D object is also a powerful means to bring out new hypotheses of research. In some cases, producing an "accurate" virtual version of cultural heritage site is a means of preserving it. Those best practices and methods are detailed in the 3D-SHS consortium white paper published in 2017, and in the PARTHENOS White Paper "Digital 3D Objects in Art and Humanities: challenges of creation, interoperability and preservation"

#### **Scenario descriptive terms**

##### Research techniques

- Georeferencing
- Photography
- Preservation Metadata
- Scanning
- Technology Preservation

##### Discipline

- Archaeology And Prehistory
- Cultural Heritage And Museology

##### Objects

- Images (3D)



#### **1- Acquisition of 3D object (by scanning)**

The choice of an appropriate scanning system depends on the object to digitize. The main parameter is, of course, the size of the object. The method will be different for a small artefact than that for an excavation site. The main methods are:



- Triangulation based artefact scanners
- Terrestrial time of flight laser scanners
- Airborne laser scanning
- Mobile Mapping systems

For any system, the methodological framework should be the following:

- Data acquisition
- Register scan
- Georeference scan
- Creation polygonal mesh
- Optimization/Decimation mesh

### ***Step descriptive terms***

Research activity

- Imaging

Standards

- STL (file format)
- FLS
- FWS

### ***Resources***

- Digital 3D Objects in Art and Humanities: challenges of creation, interoperability and preservation. White paper, 2017  
Author(s): Pierre Alliez, Laurent Bergerot, Jean-François Bernard et al.  
Link: <https://hal.inria.fr/hal-01526713/document>
- APPLIED DIGITAL DOCUMENTATION IN THE HISTORIC ENVIRONMENT, 2018  
Link: <https://pub-prod-sdk.azurewebsites.net/api/file/ba5c52a1-a5f8-429c-8e88-a8a8011a6ac7>

## ***2- Acquisition of a 3D object (by photogrammetry)***

The photogrammetry method relies on a combination of photographs of a building or a site.

Therefore, the first step is to choose the appropriate device:

- Digital SLR camera
- Bridge camera
- Compact Camera



- Mobile phone
- UAV photogrammetry

As for 3D scanning, the photogrammetry process should follow a precise methodological framework:

- Data acquisition (overlapping photographs)
- Calibration
- Align photo
- Point cloud creation
- Mesh generation
- Texture generation

### ***Step descriptive terms***

Standards

- OBJ
- PLY
- .3ds
- VRML
- AutoCAD DXF
- JPEG
- TIFF
- PNG
- BMP

Research activity

- Imaging

### ***Resources***

- COLLADA – Digital Asset Schema Release 1.5.0 Documentation, 2008  
Link: [https://www.khronos.org/files/collada\\_spec\\_1\\_5.pdf](https://www.khronos.org/files/collada_spec_1_5.pdf)
- COLLADA – Digital Asset Schema Release 1.5.0 Schema, 2008  
Link: [https://www.khronos.org/files/collada\\_schema\\_1\\_5](https://www.khronos.org/files/collada_schema_1_5)
- OpenCOLLADA, 2018  
Link: <https://github.com/KhronosGroup/OpenCOLLADA>



- Digital 3D Objects in Art and Humanities: challenges of creation, interoperability and preservation. White paper, 2017  
Author(s): Pierre Alliez, Laurent Bergerot, Jean-François Bernard et al.  
Link: <https://hal.inria.fr/hal-01526713/document>
- Agisoft PhotoScan  
Link: <http://www.agisoft.com>
- ContextCapture tutorial, 2013  
Link: <https://www.acute3d.com/tutorials/>
- Micmac, photogrammetry software, 2019  
Link: <https://github.com/micmacIGN/micmac>  
Link: <http://micmac.ensg.eu/index.php/Accueil>
- Micmac, documentation, 2019  
Link: <https://github.com/micmacIGN/Documentation/raw/master/DocMicMac.pdf>
- Capturing Reality  
Link: <https://www.capturingreality.com>
- Capturing Reality, tutorials  
Link: <https://support.capturingreality.com/hc/en-us/categories/115000560989-Tutorials>
- VisualSFM : A Visual Structure from Motion System  
Author(s): Wu Changchang  
Link: <http://ccwu.me/vsfm/>
- 3DF Zephyr - photogrammetry software - 3d models from photos  
Link: <https://www.3dflow.net/3df-zephyr-pro-3d-models-from-photos/>
- Pix4D  
Link: <https://pix4d.com/>
- VisualSFM : A Visual Structure from Motion System / Documentation  
Author(s): Wu Changchang  
Link: <http://ccwu.me/vsfm/doc.html>
- Tinkercad | Create 3D digital designs with online CAD  
Link: <https://www.tinkercad.com/>
- X3D & VRML, The Most Widely Used 3D Formats | Web3D Consortium  
Link: <http://www.web3d.org/x3d-vrml-most-widely-used-3d-formats>

### **3- Migration to target format**

Due to the diversity of digitizing devices and techniques, a wide range of data format can be produced. It is, therefore, crucial to check the content and the structure of the produced files. This verification step can be automatized or carried out by humans.

#### **Step descriptive terms**

##### Standards

- OBJ
- PLY
- .3ds
- VRML
- AutoCAD DXF
- JPEG
- TIFF
- PNG
- BMP

##### Research activity

- Modelling

#### **Resources**

- X3D & VRML, The Most Widely Used 3D Formats | Web3D Consortium  
Link: <http://www.web3d.org/x3d-vrml-most-widely-used-3d-formats>
- COLLADA - 3D Asset Exchange Schema, 2011  
Link: <https://www.khronos.org/collada/>
- COLLADA – Digital Asset Schema Release 1.5.0 Documentation, 2008  
Link: [https://www.khronos.org/files/collada\\_spec\\_1\\_5.pdf](https://www.khronos.org/files/collada_spec_1_5.pdf)
- COLLADA – Digital Asset Schema Release 1.5.0 Schema, 2008  
Link: [https://www.khronos.org/files/collada\\_schema\\_1\\_5](https://www.khronos.org/files/collada_schema_1_5)
- OpenCOLLADA, 2018  
Link: <https://github.com/KhronosGroup/OpenCOLLADA>
- 3ds Max | 3D Modelling, Animation & Rendering Software  
Link: <https://www.autodesk.com/products/3ds-max/overview>
- 3ds Max Tutorials: Introduction  
Link: <http://docs.autodesk.com/3DSMAX/16/ENU/3ds-Max-Tutorials/>



- SketchUp  
Link: <https://www.sketchup.com/home>
- Blender Documentation  
Author(s): Blender Foundation  
Link: <https://docs.blender.org/>
- Blender - Free and Open 3D Creation Software  
Author(s): Blender Foundation  
Link: <https://www.blender.org/>
- Maya Learning Channel  
Link: [https://www.youtube.com/channel/UCHmAXsicpLK2EHMZo5\\_BtDA](https://www.youtube.com/channel/UCHmAXsicpLK2EHMZo5_BtDA)
- Maya | Computer Animation & Modelling Software | Autodesk  
Link: <https://www.autodesk.com/products/maya/overview>

#### **4- Visualisation**

Visualize 3D models requires specific (web-based or not) software or libraries. Most of them propose, in addition of the visualisation itself, to interact with the data (measurement systems, orientation, movement, etc.)

#### **Step descriptive terms**

Research activity

- Publishing

#### **Resources**

- OpenSpace3D  
Link: <http://www.openspace3d.com/lang/en/softwarelogiciel>
- Digital 3D Objects in Art and Humanities: challenges of creation, interoperability and preservation. White paper, 2017  
Author(s): Pierre Alliez, Laurent Bergerot, Jean-François Bernard et al.  
Link: <https://hal.inria.fr/hal-01526713/document>
- A360 | Project Collaboration in the Cloud  
Link: <https://a360.autodesk.com/>
- OpenAQ visualization, 2017  
Link: <https://github.com/openaq/cc-viewer>
- CC Viewer - User Manual

Author(s): Vela

Link: <http://www.vela.com/ProductDocumentation/CC%20Viewer.pdf>

- Potree

Link: <http://www.potree.org/>

- Potree, 2018

Link: <https://github.com/potree/potree>

- Sketchfab - Your 3D content on web, mobile, AR, and VR.

Link: <https://sketchfab.com>

- 3DHOP

Link: <http://vcg.isti.cnr.it/3dhop/>

Link: <https://github.com/cnr-isti-vclab/3DHOP>

- 3DHOP - How To

Link: <http://vcg.isti.cnr.it/3dhop/howto.php>

- CRYENGINE

Link: [www.cryengine.com/](http://www.cryengine.com/)

- CRYENGINE V Training

Link: [www.cryengine.com/tutorials](http://www.cryengine.com/tutorials)

- CRYENGINE V Manual- Documentation

Link: <http://docs.cryengine.com/display/CEMANUAL/CRYENGINE+V+Manual>

- Unity Learn Tutorials

Link: <https://unity3d.com/fr/learn/tutorials>

- Unity - Manual: Unity User Manual (2018.1)

Author(s): Unity Technologies

Link: <https://docs.unity3d.com/Manual/index.html>

- Unity

Link: <https://unity3d.com>

### 3.4 Preserving 3D digital objects

*Sarah Tournon-Valiente, Mehdi Chayani, Valentin Grimaud, Xavier Granier (3D-SHS Huma-Num's consortium), H  l  ne Gautier, Adeline Joffres and Paulin Ribbe (Huma-Num/CNRS)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_production3DObjectsScenario](http://ssk.huma-num.fr/#/scenarios/SSK_sc_production3DObjectsScenario)

The long-term preservation of 3D digital objects raises some critical issues. Indeed, firstly, different kinds of data can be archived: the original files depending on the method used to digitize; the final version, depending on the goal (simple acquisition or modelisation) and any relevant information (text, images, PDF, scenarios...) added for a better comprehension of the 3D object. Secondly, it is important that metadata accompany the archive to ensure that the 3D object will be human-readable in the long-term. These metadata have to be completed in distinct files. The 3D consortium of Huma-Num developed a tool, aLTAG3D to easily produce metadata file for the full 3D archive. Lastly, a place to preserve 3D data in the long term has to be found. Many solutions exist, depending on the countries. In France, there are solutions such as the 3D conservatory for 3D data research for mid-term preservation, and the CINES in the continuity for long-term preservation.

#### **Scenario descriptive terms**

##### Research techniques

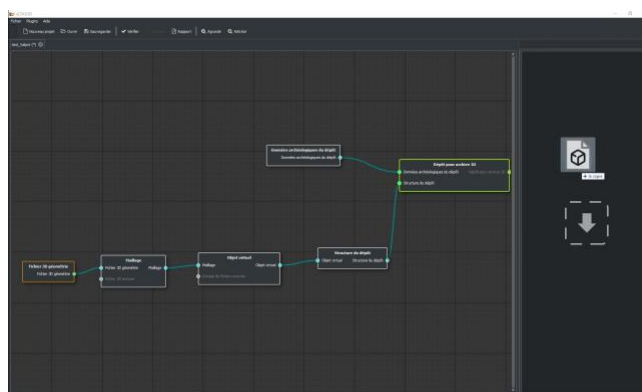
- Georeferencing
- Photography
- Preservation Metadata
- Scanning
- Technology Preservation

##### Research Object

- Images (3D)

##### Discipline

- Archaeology and Prehistory



**1- Convert to 3D long-term format**

Choose an open file format with acceptable losses for archiving 3D objects.

**Step descriptive terms**

Research activity

- preservation

Standards

- PLY
- COLLADA

**Resources**

- PLY (file format), 2018

Link: [https://en.wikipedia.org/w/index.php?title=PLY\\_\(file\\_format\)&oldid=842574812](https://en.wikipedia.org/w/index.php?title=PLY_(file_format)&oldid=842574812)

- COLLADA – Digital Asset Schema Release 1.5.0, 2008

Link: [https://www.khronos.org/files/collada\\_spec\\_1\\_5.pdf](https://www.khronos.org/files/collada_spec_1_5.pdf)

- COLLADA\_schema\_1\_5, 2008

Link: [https://www.khronos.org/files/collada\\_schema\\_1\\_5](https://www.khronos.org/files/collada_schema_1_5)

- COLLADA - 3D Asset Exchange Schema, 2011

Link: <https://www.khronos.org/collada/>

- OpenCOLLADA - Github, 2018

Link: <https://github.com/KhronosGroup/OpenCOLLADA>

- Meshlab - Github

Author(s): ISTI - CNR

Link: <https://github.com/cnr-isti-vclab/meshlab>

- MeshLab

Author(s): ISTI-CNR

Link: <http://www.meshlab.net/>

- MeshLab - Tutorials Youtube

Link: <https://www.youtube.com/user/MrPMeshLabTutorials>

- Blender

Link: <https://www.blender.org/>

- Blender Tutorials

Link: <https://www.blender.org/support/tutorials/>

- Blender Documentation



Author(s): Blender Foundation

Link: <https://docs.blender.org/>

- CloudCompare - 3D point cloud and mesh processing software Open Source Project

Link: <http://www.cloudcompare.org/>

- CloudCompare - Tutorials Video

Author(s): Eugene Liscio

Link: <http://www.danielgm.net/cc/tutorials.html>

- CloudCompare - Tutorials Youtube

Link:

<https://www.youtube.com/playlist?list=PLBNUxsUA00UAT63O0d95pByrCjtqIXN4>

- CloudCompare - Documentation

Link: <http://www.danielgm.net/cc/documentation.html>

- CloudCompare - Github

Link: <https://github.com/CloudCompare/CloudCompare>

- OpenFlipper

Link: <https://www.openflipper.org/>

- OpenFlipper Developer Documentation

Link: <http://openflipper.org/Documentation/latest/index.html>

- OpenFlipper-Free

Link:

<https://www.graphics.rwth-aachen.de:9000/OpenFlipper-Free/OpenFlipper-Free>

## ***2- Describe and annotate***

Enrich the 3D object with annotation and/or description based on existing Europeana format for metadata.

### ***Step descriptive terms***

Research activity

- Annotating



## Standards

- EDM
- CIDOC-CRM
- LIDO
- METS

## Resources

- CARARE Metadata Schema, 2013  
Link: <http://pro.carare.eu/doku.php?id=support:metadata-schema>
- Definition of the Europeana Data Model v5.2.8, 2017  
Link: [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation//EDM\\_Definition\\_v5.2.8\\_102017.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf)
- ARCO: An Architecture for Digitization, Management and Presentation of Virtual Exhibitions, 2004  
Author(s): Martin White, Nicholaos Mourkoussis, Joe Darcy, Panos Petridis et al.  
Link: <https://ieeexplore.ieee.org/document/1309277/>
- CIDOC-CRM  
Link: <http://www.cidoc-crm.org/>
- Conceptual Reference Model, 2011  
Link: [http://new.cidoc-crm.org/sites/default/files/cidoc\\_crm\\_version\\_5.0.4.pdf](http://new.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf)
- Definition of the CRMdig, 2016  
Link: [http://www.ics.forth.gr/isl/CRMext/CRMdig/docs/CRMdig\\_v3.2.1.pdf](http://www.ics.forth.gr/isl/CRMext/CRMdig/docs/CRMdig_v3.2.1.pdf)
- LIDO - Lightweight Information Describing Objects Version 1.0, 2010  
Link: <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>
- LIDO v1.0 XML Schema Definition  
Author(s): ICOM-CIDOC Working Group Data Harvesting and Interchange  
Link: <http://www.lido-schema.org/schema/v1.0/lido-v1.0.xsd>
- METS Implementation Registry: Metadata Encoding and Transmission Standard (METS) Official WebSite.  
Link: <http://www.loc.gov/standards/mets/mets-registry.html>
- METS : Schema and documentation  
Link: <http://www.loc.gov/standards/mets/mets-schemadocs.html>  
Link: <http://www.loc.gov/standards/mets/mets.xsd>



- METS Example Documents: Metadata Encoding and Transmission Standard (METS) Official WebSite

Link: <http://www.loc.gov/standards/mets/mets-examples.html>

- Digital 3D Objects in Art and Humanities, 2017

Link: <https://hal.inria.fr/hal-01526713/document>

### ***3- Create an archiving packet with 3D data and metadata***

Bundle every needed file in an archive file and add related metadata. Use an application such as aLTAG3D in order to make the addition of information easier, to reduce duplication and to generate an archive compliant with the long-term storage service.

#### ***Step descriptive terms***

Research activity

- Archiving

#### ***Resources***

- Schéma XML pour les métadonnées métier de l'archivage de modèles 3D en archéologie, 2017

Author(s): Sarah Tournon-Valiente

Link: <http://altag3d.huma-num.fr/data/mdacst3d.xsd>

- aLTAG3D - a Long Term Archive Generator for 3D

Author(s): consortium 3D

Link: <http://altag3d.huma-num.fr/>

### ***4- Upload the archive in an official reliable repository***

Submit the packet to a long-term storage service, preferably an official academic public permanent structure (like CINES in France, for instance).

#### ***Step descriptive terms***

Research activity

- Archiving

## ***Resources***

- CINES | Centre Informatique National de l'Enseignement Supérieur  
Link: <https://www.cines.fr/>

### 3.5 Tracking the dissemination of a specific work

*Maurizio Sanesi (SISMEL), Roberta Giacomini (SISMEL), Emiliano Degl'Innocenti (CNR-OVI), Claus Spiecker (FHP), and Karolien Verbrugge (NIOD)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_trackingTheDisseminationV2](http://ssk.huma-num.fr/#/scenarios/SSK_sc_trackingTheDisseminationV2)

In medieval studies, the dissemination of a work can provide important information on the life and movements of the author; plus, it can influence local history and everyday activities: i.e. some medical practices were performed just in places where the work carrying information about them was spread. Moreover, it can be useful also to understand if the work was disseminated alone or together with other works, and why, for instance, in medieval universities the essential works for the study of a discipline were often gathered in the same manuscript. Therefore, scholars may have interest in the itinerary made by a given work, and may want to track it.

#### **Scenario descriptive terms**

Research objects

- Metadata

Discipline

- History



#### **1- Select the work on which the survey is conducted**

Select the work on which you want to conduct the survey

#### **Step descriptive terms**

Research activity

- Discovering

Standards

- XML
- TEI

**Resources**

- MIRABILE, Digital Archives for Medieval Culture  
Author(s): SISMELE, Fondazione Ezio Franceschini  
Link: <http://www.mirabileweb.it/>
- TEXT-inc  
Author(s): 15cBOOKTRADE Project  
Link: <http://textinc.bodleian.ox.ac.uk/>
- BASE - Bielefeld Academic Search Engine  
Author(s): Bielefeld University Library  
Link: <https://www.base-search.net/about/en/>

**2- Define the geographic and the chronological boundaries**

Definition of chronological and geographic boundaries is needed in order to delimit the research field.

**Step descriptive terms**

Research activity

- Theorizing

**Resources**

- Getty Thesaurus of Geographic Names  
Author(s): Getty Research Institute  
Link: <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>

**3- Select data which are to be retained for curation**

Gather and select data that need to be curated and select the type of curation needed: it is important to preserve, maintain and eventually update data in order to maintain validity over time.

**Step descriptive terms**

Research activity

- Gathering

Standards

- EAC
- CIDOC-CRM



## **Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

## **4- Choosing a sustainable archive, repository, data centre or custodian for data storage**

Data needs to be stored in a sustainable archive or repository: therefore, the most suitable custodian has to be selected.

### **Step descriptive terms**

Research activity

- Identifying

## **Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

## **5- Adhere to guidance, policies and legal requirements**

Search for policies and guidelines suitable for data selected for the research.

### **Step descriptive terms**

Research activity

- Organizing

## **Resources**

- PARTHENOS Policies & Guidelines

Author(s): PARTHENOS Project

Link: [http://www.parthenos-project.eu/portal/policies\\_guidelines](http://www.parthenos-project.eu/portal/policies_guidelines)

## **6- Collect all existent editions of the selected work**

After the selection of the work, it is necessary to collect and compare all the editions available.

**Step descriptive terms**

Research activity

- Gathering

**Resources**

- CERL Portal  
Link: [https://www.cerl.org/resources/cerl\\_portal](https://www.cerl.org/resources/cerl_portal)
- KIT-Bibliothek | Karlsruher Virtueller Katalog KVK  
Author(s): Karlsruhe Institute of Technology  
Link: <https://kvk.bibliothek.kit.edu/>
- Incunabula Short Title Catalogue  
Author(s): The British Library  
Link: <http://www.bl.uk/catalogues/istc/>
- MIRABILE, Digital Archives for Medieval Culture  
Author(s): SISMELE, Fondazione Ezio Franceschini  
Link: <http://www.mirabileweb.it/>
- TEXT-inc  
Author(s): 15cBOOKTRADE Project  
Link: <http://textinc.bodleian.ox.ac.uk/>
- BASE - Bielefeld Academic Search Engine  
Author(s): Bielefeld University Library  
Link: <https://www.base-search.net/about/en/>

**7- Assess the use of the editions in the selected period.**

It is needed to assess and verify how much the modern editions and manuscripts were known and used in a determined timespan. Moreover, it is needed to assess how many manuscripts of this work survive today in our libraries.

**Step descriptive terms**

Research activity

- Theorizing



## **Resources**

- WorldCat

Author(s): OCLC

Link: <https://www.worldcat.org/>

- Material Evidence in Incunabula (MEI)

Author(s): CERL

Link: [https://data.cerl.org/mei/\\_search](https://data.cerl.org/mei/_search)

- Manus online, Census of Manuscripts held in Italian libraries

Author(s): Istituto Centrale per il Catalogo Unico

Link: <https://manus.iccu.sbn.it/>

## **8- Understand the popularity and circulation of this work**

Discover who were the users of the surviving copies and how well it was known.

### **Step descriptive terms**

Research activity

- Modelling

## **Resources**

- Biblissima Resources

Author(s): Equipex Biblissima

Link: <https://projet.biblissima.fr/en/resources/biblissima-resources>

- Medieval Libraries of Great Britain

Author(s): Richard Sharpe, James Willoughby

Link: <http://mlgb3.bodleian.ox.ac.uk/>

## **9- Ensure inclusiveness of data**

Ensure that the CERL Thesaurus is running at the back of the above listed tools.

### **Step descriptive terms**

Research activity

- Structural Analysis

Standards

- OPAC



## **Resources**

- CERL Thesaurus  
Author(s): CERL  
Link: [https://data.cerl.org/thesaurus/\\_search](https://data.cerl.org/thesaurus/_search)

## **10- Bundle collected data**

The collected data needs to be packaged in one location for better use.

## **Step descriptive terms**

Research activity

- storage

Standards

- XML
- CIDOC-CRM
- TEI

## **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments  
Author(s): CNR-ISTI  
Link: <https://services.d4science.org/explore>

## **11- Interpret data and offer explanations on statistical information**

With the support of services and tools, it is possible to analyze and gather information on the corpus from different points of view. It is possible to interpret these data and offer explanations on statistical information, metrics, or other methods are used to analyse the corpus

## **Step descriptive terms**

Research activity

- Publishing

## **Resources**

- NERD: Named Entity Recognition and Disambiguation  
Author(s): Eurecom



Link: <http://nerd.eurecom.fr/>

- Dandelion API | Semantic Text Analytics as a service

Author(s): SpazioDati

Link: <http://www.dandelion.eu>

- Pleiades: A community-built gazetteer and graph of ancient places, 2006

Author(s): Roger Bagnall, Richard J. A. Talbert, Sarah Bond et al.

Link: <http://pleiades.stoa.org>

- Getty Thesaurus of Geographic Names

Author(s): Getty Research Institute

Link: <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>

## ***12- Sharing data through articles and publications***

Publication of a OAI-PMH repository according to specific application profiles.

### ***Step descriptive terms***

Research activity

- Sharing

### ***Resources***

- The open archive HAL.

Link: <https://hal.archives-ouvertes.fr>

## ***13- Ensuring that the required legal permissions are available***

In order to publish and share the results of the research, it is necessary to check and ensure that all the required legal permissions are available.

### ***Step descriptive terms***

Research activity

- Publishing

### ***Resources***

- PARTHENOS Policies & Guidelines

Author(s): PARTHENOS Project

Link: [http://www.parthenos-project.eu/portal/policies\\_guidelines](http://www.parthenos-project.eu/portal/policies_guidelines)

### 3.6 Create a dictionary in TEI

*Karlheinz Mörth (OEAW-ACDH), Charles Riondet (Inria) and Lionel Tadjou (Inria)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_dictionaryInTei](http://ssk.huma-num.fr/#/scenarios/SSK_sc_dictionaryInTei)

This scenario sets out the best practices for creating a born-digital dictionary, especially with the TEI (Text Encoding Initiative). However, building a standardized lexicographical dataset is not only a data format problem, it is also an intellectual and technical process where one has to choose how to model the data, and with which tools to operate in order to create an easy-to-use and sustainable resource.

#### **Scenario descriptive terms**

Discipline

- Linguistics

Techniques

- Encoding

Standards

- TEI



#### **1- Build the model of the dictionary**

The first step is to identify the structural components of the dictionary and test if they can be represented in TEI. Some steps to consider may be:

- Identify the appropriate language code (ideally BCP 47) for the linguistic varieties involved in the dictionary project
- Check the TEI guidelines (section 9, Dictionaries)
- For existing dictionaries, try to find one example for all possible variants of the entry.
- Experiment with an XML editor (like Oxygen) and a TEI schema.

#### **Step descriptive terms**

Research activity

- Modelling



## **Resources**

- 9 Dictionaries - The TEI Guidelines  
Link: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>
- BARTOC.org | Basel Register of Thesauri, Ontologies & Classifications  
Link: <https://bartoc.org/>
- CLARIN Concept Registry | CLARIN ERIC  
Link: <https://www.clarin.eu/ccr>
- BCP 47 - Tags for Identifying Languages  
Author(s): IETF Network Working Group  
Link: <https://tools.ietf.org/html/bcp47>
- Language tags in HTML and XML  
Link: <https://www.w3.org/International/articles/language-tags/>

## **2- Create a corpus of useful resources for the dictionary.**

These resources can be corpora as well as other dictionaries, used as references or direct sources to build the dictionary. Depending on the cases, this step may require digitizing printed text (see the related scenario on digitization), and/or build a normalized and lemmatized corpus (see here the related scenario).

## **Step descriptive terms**

Research activity

- Modelling

## **Resources**

- Learning attention for historical text normalization by learning to pronounce, 2017  
Author(s): Marcel Bollmann, Joachim Bingel, Anders Søgaard  
Link: <http://aclweb.org/anthology/P17-1031>
- KonText - Corpora query  
Author(s): Czech National Corpus  
Link: <https://kontext.korpus.cz/>
- TreeTagger  
Author(s): Helmut Schmid  
Link: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

### **3- Setting up the editing environment**

Choose a tool for editing the dictionary, and also consider the back-end. For the latter one, we can discern two prototypical approaches: systems based on relational databases (MySQL, PostgreSQL etc.) and native XML databases (eXist, BaseX).

#### **Step descriptive terms**

Research activity

- Modelling

#### **Resources**

- VLE - Viennese Lexicographic Editor  
Author(s): Karlheinz Mörrth, Omar Siam  
Link: <https://www.oeaw.ac.at/en/acdh/tools/vle/>
- Dictionary Writing Systems & Corpus Query Systems  
Author(s): Simon Krek, Andrea Abel, Carole Tiberius, et al.  
Link: [http://www.elexicography.eu/wp-content/uploads/2015/04/ENeL\\_WG3\\_Vienna\\_DWS\\_CQS\\_final\\_web.pdf](http://www.elexicography.eu/wp-content/uploads/2015/04/ENeL_WG3_Vienna_DWS_CQS_final_web.pdf)
- BaseX Documentation  
Link: [http://docs.basex.org/wiki/Main\\_Page](http://docs.basex.org/wiki/Main_Page)
- Introducing Lexonomy: an open-source dictionary writing and publishing system  
Author(s): Michal Měchura  
Link: <https://www.lexonomy.eu/docs/elex2017.pdf>

### **4- Setting up the publishing environment**

For web publication, it may be necessary to set up a rich environment with complex querying functionalities. Tools like Elasticsearch can help provide such a framework, but require development skills. For printed output, you can choose between XSL-FO, CSS for Paged Media, LaTeX or InDesign Markup Language with post-processing.

#### **Step descriptive terms**

Research activity

- Modelling

Standards

- XSL



- CSS Paged Media
- XSL-FO
- Apache FOP
- InDesign Mark-up Language

### **Resources**

- The Extensible Stylesheet Language Family (XSL)  
Author(s): W3C XSLT Working Group  
Link: <https://www.w3.org/Style/XSL/>
- Elastic Search  
Link: <https://github.com/elastic/elasticsearch>
- Introducing Lexonomy: an open-source dictionary writing and publishing system  
Author(s): Michal Měchura  
Link: <https://www.lexonomy.eu/docs/elex2017.pdf>

### **5 - Provide for long-term availability**

Archive the dictionary in a repository providing for long-term and free accessibility, such as Zenodo. Archive the dictionary, the documentation (in ODD) and the stylesheets used for the post-processing and the publication.

### **Step descriptive terms**

Research activity

- Modelling

### **Resources**

- TEI: Getting Started with ODDs, 2013  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/Guidelines/Customization/odds.xml>
- Zenodo - Research. Shared  
Link: <https://zenodo.org/>

### 3.7 Research standards for scholars

*Maurizio Sanesi (SISMEL), Roberta Giacomi (SISMEL), Emiliano Degl'Innocenti (CNR-OVI), Claus Spiecker (FHP) and Karolien Verbrugge (NIOD)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_scholarsStandardsV2](http://ssk.huma-num.fr/#/scenarios/SSK_sc_scholarsStandardsV2)

In medieval studies, it is often needed to define, in an unambiguous way, information such as person names (i.e.: authors), titles of works and documents identifiers (i.e.: shelf marks). In fact, information can frequently be ambiguous: there are several cases of homonymy; an author can have more than one known appellation; the title of a work can have different forms, and also shelf marks may have different structures. The quantity of material involved in the research and the time employed to perform it may vary depending on the subject of the research itself.

#### **Scenario descriptive terms**

##### Discipline

- History

##### Research objects

- Metadata
- Bibliographic Listings
- Persons



#### **1 - Define the research target**

Define the research target (persons, works, shelf marks), and thematic, geographic and chronological boundaries.

#### **Step descriptive terms**

##### Research activity

- Modelling

##### Standards

- XML
- TEI



## **Resources**

- Getty Thesaurus of Geographic Names  
Author(s): Getty Research Institute  
Link: <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>
- MIRABILE, Digital Archives for Medieval Culture  
Authors: SISMELE, Fondazione Ezio Franceschini  
Link: <http://www.mirabileweb.it/>

## **2- Select data which are to be retained for curation**

Gather and select data that need to be curated and select the type of curation needed: it is important to preserve, maintain and eventually update data in order to maintain validity over time.

### **Step descriptive terms**

Research activity

- Gathering

Standards

- EAC
- CIDOC-CRM

## **Resources**

- The open archive HAL  
Link: <https://hal.archives-ouvertes.fr>

## **3- Choosing a sustainable archive, repository, data centre or custodian for data storage**

Data needs to be stored in a sustainable archive or repository: therefore, the most suitable custodian has to be selected.

### **Step descriptive terms**

Research activity

- Organizing



## Standards

- XML
- TEI

## **Resources**

- The open archive HAL  
Link: <https://hal.archives-ouvertes.fr>

## **4 - Adhere to guidance, policies and legal requirements**

Search for policies and guidelines suitable for the data selected for the research.

### **Step descriptive terms**

#### Research activity

- Identifying

#### Standards

- XML
- TEI

## **Resources**

- PARTHENOS Policies & Guidelines  
Author(s): PARTHENOS Project  
Link: [http://www.parthenos-project.eu/portal/policies\\_guidelines](http://www.parthenos-project.eu/portal/policies_guidelines)

## **5 - Collect standards about the selected resources**

Analyse the selected resources and choose and apply suitable standards.

### **Step descriptive terms**

#### Research activity

- Gathering

#### Standards

- EAC
- FOAF
- GND
- VIAF



## **6 - Assessment of Standards**

The standards selected for the resources have to be assessed and checked, in order to see if they are working.

### ***Step descriptive terms***

Research activity

- Gathering

Standards

- XML
- TEI

## **7 - Synthesize the results**

Gather, merge and synthesize the results of the research.

### ***Step descriptive terms***

Research activity

- Theorizing

Standards

- XML
- TEI
- CIDOC-CRM

## **8 - Interpret these data**

Analyse the data from the available (digital) resources; combine/interpret data (sources and literature).

### ***Step descriptive terms***

Research activity

- Theorizing

Standards

- XML
- TEI

## **9 - Sharing data through articles and publications**

Publication of an OAI-PMH repository according to specific application profiles.

### **Step descriptive terms**

Research activity

- Sharing

Standards

- TEI
- CIDOC-CRM

### **Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

## **10 - Ensuring that the required legal permissions are available**

In order to publish and share the results of the research, it is necessary to check and ensure that all the required legal permissions are available.

### **Step descriptive terms**

Research activity

- Publishing

Standards

- XML
- CIDOC-CRM
- OAI-PMH

### **Resources**

- PARTHENOS Policies & Guidelines

Author(s): PARTHENOS Project

Link: [http://www.parthenos-project.eu/portal/policies\\_guidelines](http://www.parthenos-project.eu/portal/policies_guidelines)

## 3.8 Creating Metadata

Antonio Davide Madonna (MIBACT-ICCU)

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_creatingMetadata](http://ssk.huma-num.fr/#/scenarios/SSK_sc_creatingMetadata)

Scenario consisting of a step-by-step guide to make metadata in a specific application profile from data structured in different manners.

### Scenario descriptive terms

Research objects

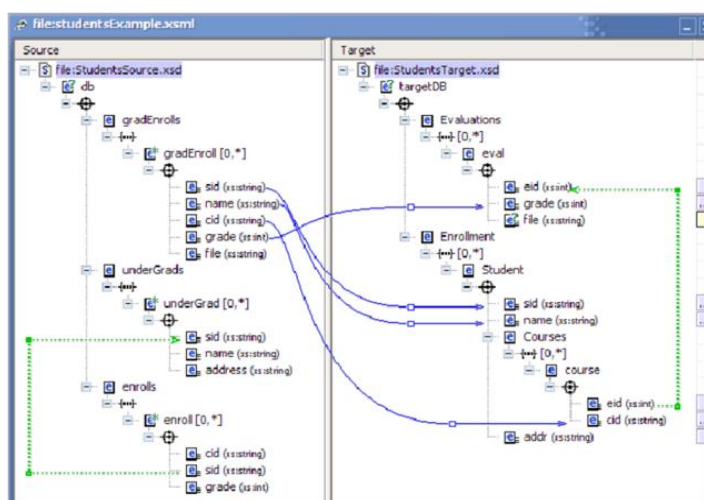
- Metadata
- Standards

Research techniques

- Mapping

Research activities

- Creation
- Conversion



### 1- Select Data (or existing Metadata) to be transformed in Metadata

This step describes the data (or metadata) selection process, based on predefined criteria: format, homogeneity, implemented fields, etc.

### Step descriptive terms

Research activity

- Analysis

### Resources

- Describing LR with Metadata: Towards Flexibility and Interoperability in the Documentation of LR, 2012

Link:

<http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf>

- Reusing CMDI components for a textCorpusProfile - towards a generic textCorpus Profile, 2014

Author(s): Lene Offersgaard, Dorte Haltrup Hansen

Link: [http://www.clarin.eu/sites/default/files/cac2014\\_submission\\_18\\_0.pdf](http://www.clarin.eu/sites/default/files/cac2014_submission_18_0.pdf)

- World Digital Library - Content Workflow

Author(s): WDL Content Selection Committee

Link: <http://project.wdl.org/content/>

## ***2- Defining responsibilities and policies for metadata maintenance***

This step describes in which way the people responsible for metadata maintenance must be identified and the related policies for metadata preservation in short, medium and long period.

### ***Step descriptive terms***

Research activity

- Preservation

### ***Resources***

- Data Service Infrastructure for the Social Sciences and Humanities, Metadata Quality Improvement, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH-D5.2\\_AB\\_final\\_\\_25nov-R.PDF](http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final__25nov-R.PDF)

## ***3- Adopting a standard for metadata***

This step describes the reasons for choosing a specific standard to create metadata.

### ***Step descriptive terms***

Research objects

- Standards

### ***Resources***

- Data Service Infrastructure for the Social Sciences and Humanities, Metadata Quality Improvement, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH-D5.2\\_AB\\_final\\_\\_25nov-R.PDF](http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final__25nov-R.PDF)
- ARIADNE, D3.2 REPORT ON PROJECT STANDARDS, 2013



Link: <http://www.ariadne-infrastructure.eu/content/download/1782/9961/version/2/file/D3.2+Report+on+project+standards.pdf>

- ATHENA EU projects, report on standards, 2009  
Link: <http://www.athenaeurope.org/getFile.php?id=435>

#### ***4- Referring to shared controlled vocabularies or ontologies for metadata***

This step describes the metadata enrichment process via controlled vocabularies or ontologies.

##### ***Step descriptive terms***

Research activity

- Enrichment

##### ***Resources***

- Data Service Infrastructure for the Social Sciences and Humanities, Metadata Quality Improvement, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH-D5.2\\_AB\\_final\\_\\_25nov-R.PDF](http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final__25nov-R.PDF)
- Documentation and User Manual of the META-SHARE Metadata Model, 2012  
Link: [http://www.meta-net.eu/public\\_documents/t4me/META-NET-D7.2.4-Final.pdf](http://www.meta-net.eu/public_documents/t4me/META-NET-D7.2.4-Final.pdf)  
World Digital Library - Content Workflow
- Author(s): WDL Content Selection Committee  
Link: <http://project.wdl.org/content/>  
ARIADNE, D3.2 REPORT ON PROJECT STANDARDS, 2013  
Link: <http://www.ariadne-infrastructure.eu/content/download/1782/9961/version/2/file/D3.2+Report+on+project+standards.pdf>

#### ***5- Mapping the data***

This step describes tools and good practice to map data in a specific metadata format without loss of useful information.

**Step descriptive terms**

Research activity

- Mapping

**Resources**

- Data Service Infrastructure for the Social Sciences and Humanities, Metadata Quality Improvement, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH-D5.2\\_AB\\_final\\_\\_25nov-R.PDF](http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final__25nov-R.PDF)
- Reusing CMDI components for a textCorpusProfile - towards a generic textCorpusProfile, 2014  
Author(s): Lene Offersgaard, Dorte Haltrup Hansen  
Link: [http://www.clarin.eu/sites/default/files/cac2014\\_submission\\_18\\_0.pdf](http://www.clarin.eu/sites/default/files/cac2014_submission_18_0.pdf)
- Semantic metadata mapping in practice: the Virtual Language Observatory, 2012  
Author(s): Dieter Van Uytvanck, Herman Stehouwer, Lari Lampen  
Link:  
[http://pubman.mpg.de/pubman/item/escidoc:1454694:11/component/escidoc:1478393/VanUytvanck\\_LREC\\_2012.pdf](http://pubman.mpg.de/pubman/item/escidoc:1454694:11/component/escidoc:1478393/VanUytvanck_LREC_2012.pdf)
- Using TEI, CMDI and ISOcat in CLARIN-DK, 2014  
Author(s): Dorte Haltrup Hansen, Lene Offersgaard, Sussi Olsen  
Link: [http://www.lrec-conf.org/proceedings/lrec2014/pdf/325\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/325_Paper.pdf)
- World Digital Library - Content Workflow  
Author(s): WDL Content Selection Committee  
Link: <http://project.wdl.org/content/>

**6- Creating or converting metadata into a standard format**

This step describes tools and good practice to create or convert data and metadata into a specific metadata format without loss of useful information.

**Step descriptive terms**

Research activity

- Conversion



## **Resources**

- Data Service Infrastructure for the Social Sciences and Humanities, Metadata Quality Improvement, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH-D5.2\\_AB\\_final\\_25nov-R.PDF](http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final_25nov-R.PDF)
- Reusing CMDI components for a textCorpusProfile - towards a generic textCorpusProfile, 2014  
Author(s): Lene Offersgaard, Dorte Haltrup Hansen  
Link: [http://www.clarin.eu/sites/default/files/cac2014\\_submission\\_18\\_0.pdf](http://www.clarin.eu/sites/default/files/cac2014_submission_18_0.pdf)
- Semantic metadata mapping in practice: the Virtual Language Observatory, 2012  
Author(s): Dieter Van Uytvanck, Herman Stehouwer, Lari Lampen  
Link: [http://pubman.mpg.de/pubman/item/escidoc:1454694:11/component/escidoc:1478393/VanUytvanck\\_LREC\\_2012.pdf](http://pubman.mpg.de/pubman/item/escidoc:1454694:11/component/escidoc:1478393/VanUytvanck_LREC_2012.pdf)
- Using TEI, CMDI and ISOcat in CLARIN-DK, 2014  
Author(s): Dorte Haltrup Hansen, Lene Offersgaard, Sussi Olsen  
Link: [http://www.lrec-conf.org/proceedings/lrec2014/pdf/325\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/325_Paper.pdf)
- World Digital Library - Content Workflow  
Author(s): WDL Content Selection Committee  
Link: <http://project.wdl.org/content/>

## **7- Quality checking**

This step describes the workflow to verify the quality of metadata produced and their correspondence with the application profile adopted.

### **Step descriptive terms**

Research activity

- Assessing

## **Resources**

- Data Service Infrastructure for the Social Sciences and Humanities, Metadata Quality Improvement, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH-D5.2\\_AB\\_final\\_25nov-R.PDF](http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final_25nov-R.PDF)



- World Digital Library - Content Workflow  
Author(s): WDL Content Selection Committee  
Link: <http://project.wdl.org/content/>
- Describing LR's with Metadata: Towards Flexibility and Interoperability in the Documentation of LR, 2012  
Link: <http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf>

## **8- Applying a licence**

This step describes good practices to identify the right licence for metadata and related digital objects.

### **Step descriptive terms**

Research activity

- Publishing

### **Resources**

- Data Service Infrastructure for the Social Sciences and Humanities, Handbook on legal and ethical issues for SSH data in Europe, Part II, 2014  
Link: [http://dasish.eu/publications/projectreports/DASISH\\_D6.5\\_februar\\_2015.pdf](http://dasish.eu/publications/projectreports/DASISH_D6.5_februar_2015.pdf)
- Intellectual Property Guidelines Version 1.0 , 2008  
Author(s): MinervaEC Working Group  
Link: [http://www.minervaeurope.org/publications/MINERVAeC%20IPR%20Guide\\_final1.pdf](http://www.minervaeurope.org/publications/MINERVAeC%20IPR%20Guide_final1.pdf)
- IPR Guidebook , 2013  
Author(s): EUROPEANAPHOTOGRAPHY  
Link: <http://www.europeana-photography.eu/getFile.php?id=298>
- META-SHARE: Licenses, Legal, IPR and Licensing issues, 2011  
Author(s): Khalid Choukri, Stelios Piperidis, Prodromos Tsiavos, et al.  
Link: [http://www.meta-net.eu/public\\_documents/t4me/META-NET-D6.1.1-Final.pdf](http://www.meta-net.eu/public_documents/t4me/META-NET-D6.1.1-Final.pdf)
- D5.1 Minimum requirements for Europeana Cloud, 2013  
Author(s): Maarten Zeinstra, Lisette Kalshoven, Alastair Dunning, et al.



Link:

[http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_Cloud/Deliverables/D5.1%20Minimum%20requirements%20for%20the%20cloud.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D5.1%20Minimum%20requirements%20for%20the%20cloud.pdf)

- Research in the Humanities and Social Sciences, in Studies on Subject-Specific Requirements for Open Access Infrastructure, 2011

Link: <http://pub.uni-bielefeld.de/publication/2458719>

- Studies on Subject-Specific Requirements for Open Access Infrastructure, 2013

Author(s): Daniel Schmidutz, Lorna Ryan, Anje Müller Gjesdal, Koenraad De Smedt

Link: [http://dasish.eu/publications/projectreports/D6.1\\_final.pdf](http://dasish.eu/publications/projectreports/D6.1_final.pdf)

- ARIADNE, D3.3 REPORT ON DATA SHARING POLICIES, 2014

Link: <http://www.ariadne->

[infrastructure.eu/content/download/2106/11888/version/2/file/D3.3+Report+on+data+sharing+policies\\_final.pdf](http://www.ariadne-infrastructure.eu/content/download/2106/11888/version/2/file/D3.3+Report+on+data+sharing+policies_final.pdf)

- Surf Share, What researchers want, 2011

Link:

[http://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2011/What\\_researchers\\_want.pdf](http://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2011/What_researchers_want.pdf)

### 3.9 Encode and modelize field surveys for their online dissemination

*Hélène Gautier (Huma-Num/CNRS) and Marie Puren (INRIA)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_DisseminationFieldSurveys](http://ssk.huma-num.fr/#/scenarios/SSK_sc_DisseminationFieldSurveys)

A researcher wants to preserve and to give access to field surveys and their related materials. Indeed, to provide a consistent context and make the reuse of raw data easier, the documentation has to include not only field materials but also documents about the research process steps. That is to say the preparation of the survey (research project, funding file, etc.), the field collection (interviews, field notebooks, etc.) and the data analysis (summary sheets, drafts, etc.). This corpus can thus cover a great variety of media (paper, digital, picture, audio or video records...) and is organised according to a classification plan. It is supposed that people interested have agreed that data will be made publicly available (potentially with some restrictions).

#### **Scenario descriptive terms**

Discipline

- Sociology

Objects

- Sound
- Video

Research Results

- Techniques
- Encoding



#### **1- Obtain the informed consent of participants**

Seek consent of the participants by advising them about risks and benefits of research participation and data sharing. Inform them how the research data will be stored, preserved and used in the long-term, and how confidentiality can be protected when needed.

#### **Step descriptive terms**

Research activity

- Gathering



## **Resources**

- Legal and ethical issues

Link: <https://www.ukdataservice.ac.uk/manage-data/legal-ethical>

- Questions éthiques et droit en SHS

Link: <http://ethiquedroit.hypotheses.org/>

## **2- Collect and classify**

Gather the different materials related to a survey. Classify them according to a classification plan divided into three sections (preparation of the survey, field collection, data analysis). Do an analytical inventory of all the documents collected (provide for each document the file name, the author, the date, the version and its position into the classification plan). Give a unique identifier to each document.

## **Step descriptive terms**

Research activity

- Gathering

Standards

- EAD

## **Resources**

- EAD: Encoded Archival Description (EAD Official Site, Library of Congress)

Author(s): Library of Congress

Link: <http://www.loc.gov/ead/>

- Encoded Archival Description Tag Library - Version 2002 (EAD Official Site, Library of Congress)

Link: <http://www.loc.gov/ead/tglib/index.html>

## **3- Select and digitize**

Prepare paper documents for digitization: removal of metal paper clips and staples, pagination, repairs, possibly light dusting and erasing in order to improve the contrast when digitizing. Digitize. Check the quality of digitized documents.

**Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001

**Resources**

- BnF - Numérisation à la BnF : contrôle qualité  
Author(s): Bibliothèque nationale de France  
Link: [http://www.bnf.fr/fr/professionnels/numerisation\\_boite\\_outils/a.num\\_controle\\_qualite.html](http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.num_controle_qualite.html)
- Jpylyzer  
Link: <http://jpylyzer.openpreservation.org/>
- FACILE - Service de validation de formats  
Link: <https://facile.cines.fr/>
- ISO 2859-1:1999 - Sampling procedures for inspection by attributes -- Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection  
Link: <https://www.iso.org/standard/1141.html>

**4- Anonymize**

If the document contains sensitive or personal data, it may be necessary to make them available to the public and, therefore, anonymize them. The point is to do that without reducing the usefulness of data (i.e. need to think about the scale/scope of the anonymization). It should be noted that pre-planning and agreeing with participants before collecting data or during the consent process, on what may and may not be recorded or transcribed, can be a much more effective way of handling sensitive data and is easier than spending time later removing the data from a recording or transcript. Perform the anonymization on file copies: use repositionable tapes for paper documents and XML markers for digital documents.

**Step descriptive terms**

Research activity

- Assessing



## **Resources**

- Anonymising Research Data, 2006  
Author(s): Andrew Clark
- UK Data Service  
Link:  
[www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative](http://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative)
- Data Management Guidelines - Anonymisation and Personal Data | Data Archive  
Author(s): FSD  
Link: <http://www.fsd.uta.fi/aineistonhallinta/en/anonymisation-and-identifiers.html>
- Research ethical guidelines and anonymity, 2005  
Author(s): Geoffrey Walford  
Link: <http://dx.doi.org/10.1080/01406720500036786>
- Techniques d'anonymisation, 2015  
Author(s): Benjamin Nguyen  
Link: [http://journal-sfds.fr/index.php/stat\\_soc/article/view/398](http://journal-sfds.fr/index.php/stat_soc/article/view/398)
- The Anonymity of Research Participants: Assumptions, Ethics and Practicalities  
Author(s): Anne Grinyer  
Link: <http://sru.soc.surrey.ac.uk/SRU36.pdf>
- Timescapes Anonymisation Guidelines  
Author(s): Timescapes Archive ESRC  
Link: <http://www.timescapes.leeds.ac.uk/assets/files/archive/Timescapes-Anonymisation-Guidelines-18Aug08.doc>

## **5- Convert into sustainable formats**

Choose a sustainable (open, interoperable and maintained) format according to the nature of the data. Consult an up-to-date list of maintained formats.

### **Step descriptive terms**

Research activity

- Conversion

Standards

- ODT
- PDF
- XML

- JPEG
- ISO/TR 21254-4:2011

### **Resources**

- FACILE - Service de validation de formats

Link: <https://facile.cines.fr/>

### **6- Transcribe the interviews**

Provide a digital transcript of the interviews in XML-TEI, with the help of its module "Transcriptions of speech", that gives general guidelines and a set of tags for the encoding of spoken material. Do not forget to mark up the data that needs to be anonymized.

### **Step descriptive terms**

Research activity

- Transcription

Standards

- TEI

### **Resources**

- 8 Transcriptions of Speech - The TEI Guidelines

Link: <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/TS.html>

### **7- Add metadata**

Add metadata according to the appropriate schema (for example keywords in DDI, classification plan in EAD).

### **Step descriptive terms**

Research activity

- Annotating

Standards

- DDI
- EAD
- METS
- DC



## **Resources**

- Explore Documentation | Data Documentation Initiative  
Link: <https://www.ddialliance.org/explore-documentation>
- EAD: Encoded Archival Description (EAD Official Site, Library of Congress)  
Author(s): Library of Congress  
Link: <http://www.loc.gov/ead/>
- Encoded Archival Description Tag Library - Version 2002 (EAD Official Site, Library of Congress)  
Link: <http://www.loc.gov/ead/tglib/index.html>
- METS Example Documents: Metadata Encoding and Transmission Standard (METS) Official Web Site  
Link: <http://www.loc.gov/standards/mets/mets-examples.html>
- METS schema (version 1.1)  
Link: <http://www.loc.gov/standards/mets/mets.xsd>
- METS : Schema and documentation  
Link: <http://www.loc.gov/standards/mets/mets-schemadocs.html>
- METS Implementation Registry: Metadata Encoding and Transmission Standard (METS) Official Web Site.  
Link: <http://www.loc.gov/standards/mets/mets-registry.html>
- DCMI Documents  
Link: <http://dublincore.org/documents/>
- DCMI Encoding Guidelines  
Link: <http://dublincore.org/resources/expressions/>
- DCMI Metadata Terms  
Link: <http://dublincore.org/documents/dcmi-terms/>
- DCMI Schemas  
Link: <http://dublincore.org/schemas/>
- Using Dublin Core  
Link: <http://dublincore.org/documents/usageguide/>

## **8- Contextualize the research**

Contextualizing the research is valuable to make the reuse of the materials easier. Based on the survey materials, the publications from the research and interviews of the researcher(s), a report completes the survey. It is composed of different sections giving



information about the genesis of the survey, the theoretical framework, the realisation of the field survey, the corpus established, the data analysis and the offspring of the survey.

### ***Step descriptive terms***

Research activity

- Contextualizing

## ***9- Disseminate and archive***

Repackage the paper documents in permanent paper folders and retain them in archive boxes. Possibly send them to the relevant archive service. Use the digital version for dissemination and submit it to a long-term archiving repository/service. Provide the option to download the data in formats compatible with analysis tools.

### ***Step descriptive terms***

Research activity

- Archiving
- Dissemination

### ***Resources***

- beQuali, Qualitative social science survey bank

Author(s): Centre de données socio-politiques

Link: <http://bequali.fr/en/>

## 3.10 Perform Statistical Analysis on Historical and Contemporary Occupations

*Emilie Kraikamp (DANS)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_statisticalAnalysisOccupations](http://ssk.huma-num.fr/#/scenarios/SSK_sc_statisticalAnalysisOccupations)

Given a large dataset with several raw variables, a social science researcher needs to (re)code some of the data in order to properly conduct statistical analysis and modelling techniques. To allow comparison across studies, researchers often develop and use certain standards for recoding some aspects of individuals, such as recoding contemporary occupations into the ISCO. After (re)coding variables, scales are often constructed by using standard methods such as factor and reliability analysis. Several software packages, such as SPSS (Statistical Package for Social Sciences), exist that are widely used for these purposes by social science researchers. Statistical analyses and modelling also have certain software standards. Finally, reporting the results is done within a consistent style, such as APA Style. Some of these coding, software and reporting standards are covered in this use case.

### **Scenario descriptive terms**

Research techniques

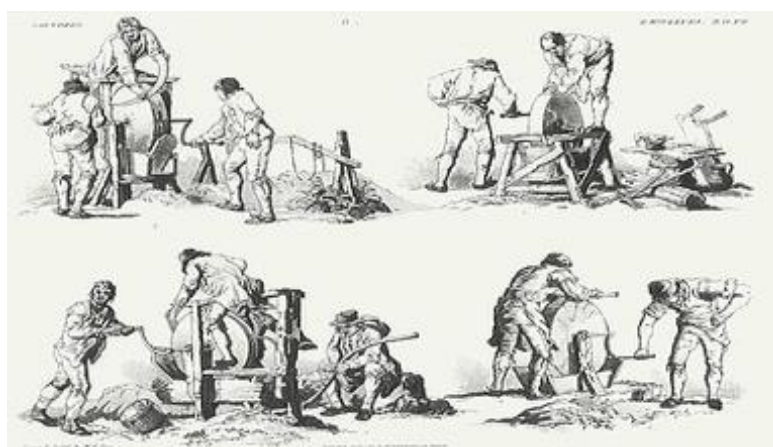
- Encoding

Research objects

- Data

Discipline

- History
- Sociology
- Methods and Statistics



### **1- Recode variables**

Recode occupation variable into a class and status score for occupations in the 19th and 20th century, and for contemporary occupations by using the HISCO occupational coding scheme.

**Step descriptive terms**

Research activity

- Contextualizing

**Standards**

- HISCO
- ISCO

**Resources**

- HISCO: Historical International Standard Classification of Occupation, 2002  
Author(s): Marco H.D. van Leeuwen, Sören Edvinsson, Ineke Maas, Andrew Miles  
Date: "2002"
- Home - Collaboratory  
Link: <https://collab.iisg.nl/web/hisco/>
- ISCO - International Standard Classification of Occupations  
Link: <http://www.ilo.org/public/english/bureau/stat/isco/isco08/>
- History Of Work Information System  
Author(s): HISCO  
Link: <https://historyofwork.iisg.nl/>
- HISCAM  
Author(s): Ineke Maas, Marco van Leeuwen  
Link: <http://www.hisma.org/HISMA/HISCAM.html>

**2- Combine variables into scales**

Perform Exploratory Factor Analysis (EFA).

**Step descriptive terms**

Research activity

- Contextualizing

**Resources**

- The R Manuals  
Link: <https://cran.r-project.org/manuals.html>
- Books related to R  
Link: <https://www.r-project.org/doc/bib/R-books.html>



- The R Journal  
Link: <https://journal.r-project.org/>
- JASP, 2019  
Link: <https://github.com/jasp-stats/jasp-desktop>
- PSPP manual  
Link: <https://www.gnu.org/software/pspp/manual/>
- STATA webpage  
Link: <http://www.stata.com/>
- SPSS homepage  
Link: <https://www.ibm.com/analytics/de/de/technology/spss/>

### ***3- Test for reliability of the scales***

Perform Reliability Analysis.

#### ***Step descriptive terms***

Research activity

- Assessing

#### ***Resources***

- The R Manuals  
Link: <https://cran.r-project.org/manuals.html>
- Books related to R  
Link: <https://www.r-project.org/doc/bib/R-books.html>
- The R Journal  
Link: <https://journal.r-project.org/>
- JASP, 2019  
Link: <https://github.com/jasp-stats/jasp-desktop>
- PSPP manual  
Link: <https://www.gnu.org/software/pspp/manual/>
- STATA webpage  
Link: <http://www.stata.com/>
- SPSS homepage  
Link: <https://www.ibm.com/analytics/de/de/technology/spss/>

#### **4- Model the relationship between variables**

Perform Structural Equation Modelling (SEM).

##### **Step descriptive terms**

Research activity

- Structural-analysis

##### **Resources**

- The R Manuals  
Link: <https://cran.r-project.org/manuals.html>
- Books related to R  
Link: <https://www.r-project.org/doc/bib/R-books.html>
- The R Journal  
Link: <https://journal.r-project.org/>
- JASP, 2019  
Link: <https://github.com/jasp-stats/jasp-desktop>
- PSPP manual  
Link: <https://www.gnu.org/software/pspp/manual/>
- STATA webpage  
Link: <http://www.stata.com/>
- SPSS AMOS homepage  
Link: <https://www.ibm.com/fr-en/marketplace/structural-equation-modeling-sem>

#### **5- Analyse**

Calculate p-values for significance, by using Analysis of Variance (ANOVA) or regression analyses.

##### **Step descriptive terms**

Research activity

- Structural-analysis



## **Resources**

- The R Manuals  
Link: <https://cran.r-project.org/manuals.html>
- Books related to R  
Link: <https://www.r-project.org/doc/bib/R-books.html>
- The R Journal  
Link: <https://journal.r-project.org/>
- PSPP manual  
Link: <https://www.gnu.org/software/pspp/manual/>
- STATA webpage  
Link: <http://www.stata.com/>
- SPSS homepage  
Link: <https://www.ibm.com/analytics/de/de/technology/spss/>

### 3.11 Prosopography research

*Maurizio Sanesi (SISMEL), Roberta Giacomini (SISMEL) and Emiliano Degl'Innocenti (CNR-OVI)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_ProsopographyV2](http://ssk.huma-num.fr/#/scenarios/SSK_sc_ProsopographyV2)

Prosopography is the investigation of the common background characteristics of a group of actors in history, making a collective study of their lives. Prosopography is mostly used by historians to address two main research questions: 1. roots of political action, e.g. the interests behind the rhetoric of politics or the social and economic affiliations of political groupings; 2. social structure and social mobility, e.g. the role in society, the degree of social mobility and the correlation of intellectual or religious movements with other factors.

#### **Scenario descriptive terms**

Discipline

- History

Objects

- Metadata
- Persons



#### **1- Formulate specific working hypotheses and specific historical questions**

Formulate specific working hypotheses and specific historical questions concerning the target group. Defining strict criteria for limiting the target population and the geographic, chronological and thematic boundaries.

#### **Step descriptive terms**

Research activity

- Theorizing

Standards

- XML
- TEI



## ***2- Select data which are to be retained for curation***

Gather and select data that need to be curated and select the type of curation needed: it is important to preserve, maintain and eventually update data in order to maintain validity over time.

### ***Step descriptive terms***

Research activity

- Gathering

Standards

- EAC
- CIDOC-CRM

### ***Resources***

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

## ***3- Choosing a sustainable archive, repository, data centre or custodian for data storage***

Data need to be stored in a sustainable archive or repository: therefore, the most suitable custodian has to be selected.

### ***Step descriptive terms***

Research activity

- Identifying

### ***Resources***

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

## ***4- Adhere to guidance, policies and legal requirements***

Search for policies and guidelines suitable for data selected for research.



### ***Step descriptive terms***

Research activity

- Organizing

### ***Resources***

- PARTHENOS Policies & Guidelines

Author(s): PARTHENOS Project

Link: [http://www.parthenos-project.eu/portal/policies\\_guidelines](http://www.parthenos-project.eu/portal/policies_guidelines)

## ***5- Select appropriate reference resource***

Appropriate reference resources have to be selected according to the objectives of the researcher and the data chosen, in order to get useful results.

### ***Step descriptive terms***

Research activity

- Gathering

Standards

- EAC
- FOAF
- GND
- VIAF

## ***6- Select / Extrapolate only useful data***

According to the question that needs to be answered in order to obtain results, and checking the selection of reference resources chosen, the researcher can extrapolate useful data.

### ***Step descriptive terms***

Research activity

- Gathering

## ***7-Merge different databases in one Unique Reference Resource***

The researcher can now combine the different databases he retains useful for his research in a Unique Reference Resource.



### ***Step descriptive terms***

Research activity

- Modelling

### ***8- Disambiguate data***

According to the Unique Reference Resource, the researcher is now able to disambiguate the selected data.

### ***Step descriptive terms***

Research activity

- Organizing

### ***9- Synthesize the results***

Gather, merge and synthesize the results of the research.

### ***Step descriptive terms***

Research activity

- Theorizing

### ***10- Interpret the data***

Analyse the data from the available (digital) resources; Combine/interpret data (sources and literature).

### ***Step descriptive terms***

Research activity

- Theorizing

### ***11- Sharing data through articles and publications***

Publication of an OAI-PMH repository according to specific application profiles.

### ***Step descriptive terms***

Research activity

- Sharing

**Resources**

- The open archive HAL  
Link: <https://hal.archives-ouvertes.fr>

**12- Share the Unique Reference Resource**

Share/publish the Unique Reference Resource as result of the research.

**Step descriptive terms**

Research activity

- Sharing

Standards

- XML
- OAI-PMH

**13- Ensuring that the required legal permissions are available**

In order to publish and share the results of the research, it is necessary to check and ensure that all the required legal permissions are available.

**Step descriptive terms**

Research activity

- Publishing

Standards

- XML
- OAI-PMH
- CIDOC-CRM

**Resources**

- PARTHENOS Policies & Guidelines  
Author(s): PARTHENOS Project  
Link: [http://www.parthenos-project.eu/portal/policies\\_guidelines](http://www.parthenos-project.eu/portal/policies_guidelines)

**14- Provide tools that allow collaboration in the use and reuse of data**

Data curation in time has to allow data reuse: the researcher has to provide appropriate tools and support in order to facilitate the use and reuse of shared data.



### ***Step descriptive terms***

Research activity

- Collaboration

Standards

- XML
- OAI-PMH
- CIDOC-CRM

### ***15- Apply access controls and authentication procedures***

The researcher can apply access controls or authentication procedures to data.

### ***Step descriptive terms***

Research activity

- Preservation

Resources

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## 3.12 Creating interoperable TEI text resources with the DTA 'Base Format' (DTABf)

Susanne Haaf (Berlin-Brandenburg Academy of Sciences and Humanities)

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_DTABf](http://ssk.huma-num.fr/#/scenarios/SSK_sc_DTABf)

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, especially if working with popular works, may find a considerable amount of resources already digitized and ready to use. However, with these resources originating from various data providers (individual scholars, singular research projects, large infrastructures, ...) scholars usually face a great variety of digitization guidelines and formats. Gathering resources from different sources will hence almost always require their harmonization on different levels of processing. The CLARIN-D centre at the Berlin-Brandenburg Academy of Sciences and Humanities provides an infrastructure for this task based on the DTA 'Base Format' (DTABf). The DTABf is a TEI-P5 format for homogeneous text annotation including set of extensive guidelines for text transcription. This scenario describes the steps to take in order to create a homogeneous, DTABf-based text corpus.

### Scenario descriptive terms

#### Discipline

- History
- Linguistics
- Literature

#### Objects

- Metadata
- Data
- Standards
- Text
- Techniques

#### Techniques

- Encoding
- Preservation Metadata





## ***1- Selection of textual works relevant for the research question***

This is a qualitative task: as a first step, identify suitable texts for your research question and decide on the text selection which the planned work should be based on. This step can already be accompanied by checks for digital collections that might be suitable and available for the planned research (cf. step\_textSelection).

### ***Step descriptive terms***

Research activity

- Designing
- Discovering

## ***2- Selection of texts to gather in a research corpus***

Based on the project's bibliography it is advised to search for existing digital resources which might be reused in the current research context. Browsing for digital resources might be performed on different platforms, e.g. VLO, Europeana, German Digital Library (for German resources). The primary goal here would be to find digitized full-texts of reliable versions of the works in consideration. However, if digitized full-texts are not available, text recognition has to be considered. For that, the current step may also include a search for digital images of the written records (Europeana, VD, KVK) to be used as a basis for text recognition.

### ***Step descriptive terms***

Research activity

- Discovering
- Gathering

### ***Resources***

- CLARIN Virtual Language Observatory  
Author(s): multiple creators  
Link: <https://vlo.clarin.eu>
- Europeana Collections  
Author(s): multiple creators  
Link: <https://www.europeana.eu/portal/en>
- Deutsche Digitale Bibliothek  
Author(s): multiple creators

Link: <https://www.deutsche-digitale-bibliothek.de/>

### ***3- Determination of digitization and annotation guidelines for the research corpus***

Corpus research should be based on interoperable data, i.e. data that was harmonized with regard to transcription and encoding guidelines. Before the harmonization, it is necessary to pick or define proper guidelines. For the encoding of structural phenomena of text, it has become best practice in text processing contexts to apply the tagset provided by the Text Encoding Initiative (TEI-P5). However, the TEI tagset as a whole is broad offering tagging solutions for the most various domains. Hence, in order to achieve unambiguous encoding, it has to be reduced and contoured for its intended application. There are TEI formats published for re-use, e.g. those maintained by the TEI consortium, which can be evaluated for possible re-use rather than creating a new format from scratch. With the DTABf, a TEI format has been proposed, that was customized for the encoding of originally written texts of various text types for the purpose of corpus-based work. Its components are:

- Documentation: Guidelines for text annotation
- Documentation: Guidelines for text recognition
- Schema: Formal specification in ODD and RNG schemas (for print and manuscript annotation)
- Schema: Schematron constraints for quality checks beyond the schema.

#### ***Step descriptive terms***

Research activity

- Discovering
- Designing
- Meta: Assessing

Standards

- XML
- TEI
- Schematron



## Resources

- Introduction to the DTABf, 2011  
Link: [http://www.deutschestextarchiv.de/doku/basisformat\\_en](http://www.deutschestextarchiv.de/doku/basisformat_en)
- DTA 'Base Format' (DTABf): Schema and Documentation, 2018  
Link: <https://github.com/deutschestextarchiv/dtabf>
- The DTA “Base Format”: A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources, 2015  
Author(s): Susanne Haaf, Alexander Geyken, Frank Wiegand  
Link: <https://journals.openedition.org/jtei/1114>
- Enabling the Encoding of Manuscripts within the DTABf: Extension and Modularization of the Format, 2017  
Author(s): Susanne Haaf, Christian Thomas  
Link: <http://journals.openedition.org/jtei/1650>
- Das DTA-Basisformat in neuem Gewand, 2017  
Link: <https://sprache.hypotheses.org/147>
- Customizations provided by the TEI Consortium, 2015  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/Guidelines/Customization/>

## 4- Digitization of resources for the research corpus

If written works should be included in the corpus which are not yet available as full texts, text recognition has to be considered. This step usually includes not only text recognition but also the enrichment of the text with logical and layout TEI structures. For text recognition, depending on the quality and condition of the sources, it can be feasible to perform Optical Character Recognition (OCR) and manually correct and enrich the output, or to carry out the whole text transcription and annotation task manually. In any case, text recognition should be performed with the goal of high recognition reliability, quality, and proximity to the text source in mind. There are several tools and resources to support this text processing step based on the DTABf:

- DTABf transcription and annotation guidelines
- a DTABf conformant template document containing the schema and the most important elements
- DTAoX: a framework for DTABf conformant text annotation in oXygen XML Author



**Step descriptive terms**

## Research activity

- Data recognition
- Transcription
- Annotating
- Clean up

## Standards

- XML
- TEI
- Schematron

**Resources**

- DTA 'Base Format' (DTABf): Schema and Documentation, 2018  
Link: <https://github.com/deutschestextarchiv/dtabf>
- Introduction to the DTABf, 2011  
Link: [http://www.deutschestextarchiv.de/doku/basisformat\\_en](http://www.deutschestextarchiv.de/doku/basisformat_en)

**5- Harmonization of existing digitized text for the corpus**

In case digitized full-texts were already available, most frequently it will be necessary to convert texts into the designated corpus format. In addition, it is necessary to evaluate the guidelines text recognition was based on and the resulting recognition quality. High error rates or severe interferences with the texts due to normalization might necessitate a text correction step. For format harmonization, the tool OxGarage, which allows for the conversion between various formats, can be used as a first step. Furthermore, there are several tools and resources to support this text harmonization step based on the DTABf:

DTAoX: a framework for DTABf conformant text annotation in oXygen XML Author

DTABf transcription and annotation guidelines

**Step descriptive terms**

## Research activity

- Assessing
- Annotating
- Clean up
- Converting



## **Standards**

- XML
- TEI
- Schematron

## **Resources**

- OxGarage Conversion  
Link: <http://www.tei-c.org/oxgarage/>
- Introduction to the DTABf, 2011  
Link: [http://www.deutschestextarchiv.de/doku/basisformat\\_en](http://www.deutschestextarchiv.de/doku/basisformat_en)

## **6- Recognition of metadata**

For each text in the corpus, it is necessary to record bibliographic information. The DTABf provides an extensive TEI Header to hold this information, containing not only title, author and year of publication, but also further information on the digital publication and licensing, notes about the text sources, responsibilities, etc. Metadata records based on the DTABf can easily be converted into other metadata formats, e.g. into similarly rich CMDI records. For this, there is a DTABf-based CMDI profile available. This way, the digitized documents can be made findable from other platforms, e.g. CLARINs Virtual Language Observatory. Several resources are provided to facilitate this step:

- A web form to record the necessary metadata in a WYSIWYG environment and to automatically create the DTABf conformant `teiHeader`
- Extensive documentation on the recording of metadata according to DTABf
- The DTABf-RNG-schema to validate the entire TEI-Document including its metadata
- The DTABf-based CMDI profile.

## **Step descriptive terms**

Research activity

- Data recognition

Standards

- XML
- TEI
- CMDI

## **Resources**

- Introduction to the DTABf, 2011  
Link: [http://www.deutschestextarchiv.de/doku/basisformat\\_en](http://www.deutschestextarchiv.de/doku/basisformat_en)
- Source files of the DTA 'Base Format' (DTABf) documentation and schema., 2018  
Link: <https://github.com/deutschestextarchiv/dtabf>
- DTA Base Format  
Link:  
[https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.x/profiles/clarin.eu:cr1:p\\_1381926654438/xsd](https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.x/profiles/clarin.eu:cr1:p_1381926654438/xsd)
- CLARIN Virtual Language Observatory  
Author(s): multiple creators  
Link: <https://vlo.clarin.eu>

## **7- Provision of the corpus**

The text corpus that has been created with the issue of interoperability in mind and according to the standardized DTA 'Base Format' can then be provided to the Digital Humanities' Community for re-use in other project contexts. It can be handed over to a CLARIN centre for persistent storage and availability via the CLARIN infrastructure. Because of the usage of the DTABf as a standardized format, conversion into CLARIN's input formats (CMDI, TCF) can be done automatically. This way, the corpus becomes findable via CLARIN's faceted search platform VLO (Virtual Language Observatory) and can be further analysed in WebLicht. German text corpora can be easily integrated in the Deutsches Textarchiv (DTA) platform, where they become part of a large historical corpus and connected to search and data analysis facilities on the platform. In any case, it is strongly recommended to respect the FAIR principles when providing the corpus.

## **Step descriptive terms**

### Research activity

- Archiving
- Preservation
- Publishing
- Sharing



## Standards

- CMDI
- TEI
- TCF

## Resources

- CLARIN Virtual Language Observatory  
Author(s): multiple creators  
Link: <https://vlo.clarin.eu>
- WebLicht: Getting Started, 2017  
Author(s): multiple authors  
Link: [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting\\_Started](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting_Started)
- TEI-TCF encoder+decoder  
Author(s): Bryan Jurish  
Link: <http://kaskade.dwds.de/tei-tcf/>
- The FAIR Data Principles  
Link: <https://www.force11.org/group/fairgroup/fairprinciples>
- CMDI specification, version 1.2, 2016  
Link: <https://www.clarin.eu/cmdi1.2-specification>  
Deutsches Textarchiv (DTA), 2007  
Author(s): multiple creators  
Link: <http://www.deutschestextarchiv.de/>

### 3.13 Extract textual content from images

Saranya Balasubramanian (OEAW-ACDH), Peter Catrie (Koninklijke Bibliotheek van België), Dario Kampkaspar (OEAW-ACDH), Tomasz Parkoła (Poznańskie Centrum Superkomputerowo-Sieciowe, Poznań), Charles Riondet (Inria), Pavel Stranak (Institute of Formal and Applied Linguistics, Charles University) and Daniel Schopper (OEAW-ACDH)

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_OCR](http://ssk.huma-num.fr/#/scenarios/SSK_sc_OCR)

#### Scenario descriptive terms

##### Discipline

- Literature
- History

##### Images

- Images
- Manuscript
- Text
- Text Bearing Objects

##### Techniques

- Pattern
- Recognition Scanning



Figure 2: Example of a text recognized with an OCR software, photo by Pietro.dipalma

#### 1- Define the characteristics of the outcome

First you need to precisely define the purpose of your text extraction task. There are many possible goals of text extraction, e.g. to get:

- searchable text: plain text OCR, without page segmentation, in order to use it in search engine or statistical analysis of textual data;
- “reading text” (reconstructing the main text flow excluding running heads etc.) e.g. to use it in further text collation exercises,
- reconstruct page layout with all the details, e.g. to use it for further annotation exercises.

It is also important to define text representation guidelines, or normalisation choices.



### ***Step descriptive terms***

Research activity

- Modelling

### ***Resources***

- Best Practices for TEI in Libraries, 2018  
Link: <https://github.com/kshawkin/Best-Practices-for-TEI-in-Libraries>
- SIG TEI in Libraries WIKI page  
Link: <https://wiki.tei-c.org/index.php/SIG:Libraries>

## ***2- Define the characteristics of the image***

The important characteristics of the image are the following:

- Layout elements (tables, graphics, ...)
- Script type (handwritten, Gothic script, Roman script, non-European)
- Mixed content (pre-printed documents with handwriting)

These characteristics determine the method:

- If it's a clean printed text, a classic OCR engine would be enough
- In other cases, i.e. handwritten, old printed texts with blackletter (gothic) scripts, or low quality scans, using a model is necessary.

If a model exists, or it works out of the box, or it needs to be retrained or adapted. Otherwise, you have to train a model from scratch.

### ***Step descriptive terms***

Research activity

- Modelling

## ***3- Survey existing experiences***

### ***Step descriptive terms***

Research activity

- Contextualizing

### ***Resources***

- Map of the digitisation landscape - IMPACT Centre of Competence

Link: <https://www.digitisation.eu/community/map-of-the-digitisation-landscape/>

- Europeana Newspapers

Link: <https://github.com/EuropeanaNewspapers>.

#### **4- Choose engine based on the type of the content**

If you choose a classic commercial OCR engine like ABBYY FineReader, the customization would be limited. Others tools allow for more options (training a model for instance). It may be interesting also to use separate tools to perform different steps of the recognition process (layout and character recognition).

#### **Step descriptive terms**

Standards

- hOCR
- ALTO-XML

#### **Resources**

- IMPACT: Report on the comparison of Tesseract and ABBYY FineReader OCR engines

Link: [https://www.digitisation.eu/download/website-files/IMPACT\\_D-EXT2\\_Pilot\\_report\\_PSNC.pdf](https://www.digitisation.eu/download/website-files/IMPACT_D-EXT2_Pilot_report_PSNC.pdf)

- Tools for text digitisation

Link: <https://www.digitisation.eu/tools-resources/tools-for-text-digitisation/?query=&search-filter-group=text+recognition&search-filter-type=&search-filter-subtype=>

- Tesseract OCR

Link: <https://github.com/tesseract-ocr/tesseract>

- Transkribus

Link: <https://transkribus.eu/Transkribus/>

- Språkbanken

Link: <https://spraakbanken.gu.se/eng/ocr>

#### **5- Layout analysis**

To identify specific text structures, it may be suitable to use a separate engine, for instance for table recognition, or to detect columns.



### ***Step descriptive terms***

Research activity

- Structural analysis

### ***Resources***

- Tools for text digitisation

Link: <https://www.digitisation.eu/tools-resources/tools-for-text-digitisation/>

- hocr-detect-columns, 2018

Link: <https://github.com/nypl-spacetime/hocr-detect-columns>

## ***6- Create manual transcriptions***

In case all existing engine give poor results or if the desired textual representation is very specific, one has to create his/her own training model or adapt an existing one. For this, one needs a number of manually transcribed pages: the larger part is used as the training set for an OCR/HTR engine; a smaller one is used to evaluate the result of the trained model. Such transcriptions can be obtained by several ways: It can be produced internally by the project, it could be outsourced to a contractor, or it could be produced by volunteers.

Caveats:

- Take care of the character encoding
- It can be good practice to test 2 engines and merge models (but be careful at the overfitting problem).
- Printed material: No need of a lot of training data (random 5 to 10 pages)

### ***Step descriptive terms***

Research activity

- Structural analysis

### ***Resources***

- Overfitting

Link: <https://web.ma.utexas.edu/users/mks/statmistakes/ovefitting.html>

- How To Transcribe Documents with Transkribus – Advanced Mode

Link: [https://transkribus.eu/wiki/images/d/db/HowToTranscribe\\_AdvancedMode.pdf](https://transkribus.eu/wiki/images/d/db/HowToTranscribe_AdvancedMode.pdf)

- Tesseract Open Source OCR Engine, 2019

Link: <https://github.com/tesseract-ocr/tesseract>

- Medieval Unicode Font Initiative



Link: <https://folk.uib.no/hnooh/mufi/>

- Virtual Transcription Laboratory

Link: <http://wlt.pcass.pl/wlt-web/index.xhtml>

- Transcribathon | transcribe europeana 1914-1918

Link: <https://transcribathon.com/en/>

- Trove

Link: <https://trove.nla.gov.au/>

## ***7- Training the model***

### ***Step descriptive terms***

Research activity

- Structural analysis

### ***Resources***

- Tesseract Open Source OCR Engine, 2019

Link: <https://github.com/tesseract-ocr/tesseract>

## ***8- Test on a subset and assess quality***

The first step to take is the definition of a target accuracy.

- Clean printed text: WER 70% is a minimum (can be 60% for a search engine)
- HTR of badly printed: The acceptable WER may be less. Then again, it depends on the outcomes.

To test the accuracy of the model, two metrics are available: Character Error rate (CER) and Word Error rate (WER). WER is more important because it defines the readability of the text. It is also important to assess the quality of the layout recognition.

### ***Step descriptive terms***

Research activity

- Structural analysis

Standards

- ISO 11554:2017

### ***Resources***



- ISO 28590:2017, 2017  
Link: <https://www.iso.org/fr/standard/64622.html>
- Error rates and ground truth - Text Digitisation  
Link: <https://sites.google.com/site/textdigitisation/qualitymeasures/basics>
- Results of the IMPACT project | Digital Libraries and Knowledge Platforms  
Department  
Link: <http://dl.psnc.pl/activities/projekty/impact/results/>

### **9- Correct output**

it may be a good practice to correct a subset of the output, in order to acquire additional training data, when the output is too far from the target accuracy. The post correction can be made either manually or with help of dedicated tools.

#### **Step descriptive terms**

Research activity

- Structural analysis

#### **Resources**

- The CIS OCR PostCorrectionTool, 2018  
Link: <https://github.com/cisocrgroup/PoCoTo>

### **10 Re-train the model with corrected output**

Usually, retraining the model with the corrected output from the first recognition iteration is very valuable. The number of pages to correct depends on the input, but the same amount of pages as the initial training data is reasonable.

### **11- Produce OCR output in standardized format**

Different data formats exist for OCR outputs. The most common formats are ALTO, hOCR and PAGE. All these formats are expressed in XML, so they are easily convertible to one another.

**Step descriptive terms****Standards**

- METS
- ALTO-XML
- XML
- hOCR

**Resources**

- Documentation and use cases for ALTO XML, 2019  
Link: <https://github.com/altoxml/documentation>
- Convert and transform various OCR formats (hOCR, ALTO, PAGE, FineReader), 2018  
Link: <https://github.com/kba/ocr-schemas>

**12- Extract the structure information from recognized blocks**

Convert the output to TEI, a format able to represent the semantics.

**Resources**

- Default Text Structure - The TEI Guidelines  
Link: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>

**13 - A few visualization options**

The visualization of OCR formats is made possible by various software applications. The choice is function of the desired outcome (searchable PDF, structured XML, etc.)

**Resources**

- Documentation and use cases for ALTO XML, 2019  
Link: <https://github.com/altoxml/documentation>
- View HOCR files with Mirador, 2019  
Link: <https://github.com/jbaiter/hocrviewer-mirador>
- HTML5 based DjVu file viewer, 2019  
Link: <https://github.com/mateusz-matela/djvu-html5>
- Goobi - Workflow management software for digitisation projects, 2019  
Link: <https://github.com/intranda/goobi>



- Kitodo: Kitodo. Key to digital objects

Link: <https://www.kitodo.org/>

- Kramerius

Link: <https://system-kramerius.cz/>





- Editing

#### Standards

- MEI

#### **Resources**

- Sibelius  
Author(s): Avid Technology  
Link: <http://www.avid.com/sibelius>
- Music Encoding Initiative Guidelines, 2016  
Link: [http://www.music-encoding.org/docs/MEI\\_Guidelines\\_v3.0.0.pdf](http://www.music-encoding.org/docs/MEI_Guidelines_v3.0.0.pdf)
- MEI Schemata  
Link: <https://github.com/music-encoding/music-encoding/tree/develop/schemata>
- MEI Score Editor –MEISE  
Link: <https://de.dariah.eu/mei-score-editor>
- MEISE Native note editor for the MEI encoding format  
Link: <https://sourceforge.net/projects/meise/>
- MEISE Wiki  
Link: <https://sourceforge.net/p/meise/wiki/Home/>

#### ***2- Change format into MEI files if necessary.***

Convert the Sibelius files into MEI files through the plugin SibMei.

#### ***Step descriptive terms***

##### Research activity

- Conversion

##### Standards

- MEI

#### **Resources**

- Sibelius  
Author(s): Avid Technology  
Link: <http://www.avid.com/sibelius>
- Music Encoding Initiative Guidelines, 2016  
Link: [http://www.music-encoding.org/docs/MEI\\_Guidelines\\_v3.0.0.pdf](http://www.music-encoding.org/docs/MEI_Guidelines_v3.0.0.pdf)

- MEI Schemata

Link: <https://github.com/music-encoding/music-encoding/tree/develop/schemata>

- Sibelius MEI Plugin, 2019

Link: <https://github.com/music-encoding/sibmei>

### ***3- Collect and organize metadata***

Add metadata according to Dublin Core standard. Create a structured database about people, places, sources, etc. related to the musical edition conforming to Dublin Core standards and organised through the Omeka CMS web publishing system.

#### ***Step descriptive terms***

Research activity

- Annotating

Standards

- DC

#### ***Resources***

- Omeka

Author(s): Roy Rosenzweig Center for History and New Media, George Mason University, Corporation for Digital Scholarship

Link: <https://omeka.org/>

- Omeka Classic User Manual

Link: <https://omeka.org/classic/docs/>

- DCMI Documents

Link: <http://dublincore.org/documents/>

- DCMI Encoding Guidelines

Link: <http://dublincore.org/resources/expressions/>

- DCMI Metadata Terms

Link: <http://dublincore.org/documents/dcmi-terms/>

- DCMI Schemas

Link: <http://dublincore.org/schemas/>

- Using Dublin Core

Link: <http://dublincore.org/documents/usageguide/>



#### **4- Enrich the metadata of MEI files**

Add metadata in the header of MEI files using MerMEId.

##### **Step descriptive terms**

Research activity

- Annotating

##### **Resources**

- MerMEId manual  
Link: <http://labs.kb.dk/editor/manual/index.html>
- MerMEId – Metadata Editor and Repository for MEI Data  
Link: <http://www.kb.dk/en/nb/dcm/projekter/mermeid.html>
- Tool for editing MEI header data, 2019  
Link: <https://github.com/Det-Kongelige-Bibliotek/MerMEId>

#### **5- Enrich the musical edition with critical apparatus metadata**

Add to music edition one staff for each source attesting variants or person suggesting emendation. Encode variants/emendations only in the measures where interventions occur.

##### **Step descriptive terms**

Research activity

- Annotating

Standards

- MEI

##### **Resources**

- Sibelius  
Author(s): Avid Technology  
Link: <http://www.avid.com/sibelius>
- Music Encoding Initiative Guidelines, 2016  
Link: [http://www.music-encoding.org/docs/MEI\\_Guidelines\\_v3.0.0.pdf](http://www.music-encoding.org/docs/MEI_Guidelines_v3.0.0.pdf)
- MEI Schemata  
Link: <https://github.com/music-encoding/music-encoding/tree/develop/schemata>

#### **6- Enrichment of MEI files**



Correct errors occurred during the conversion process. Enrich the files with MEIMessaging depending on the required output. Convert MEI files into drawing instructions for the VexFlow online music notation rendering API (MEItoVexFlow) and/or Verovio.

### ***Step descriptive terms***

Research activity

- Editing
- Conversion
- Clean up

### ***Resources***

- Django/python powered MEI messaging app, 2016  
Link: <https://github.com/DuChemin/MEIMessaging>
- JavaScript demo application displaying variant readings of historical musical pieces, 2015  
Link: <https://github.com/zolaemil/meiView>
- JavaScript library to render MEI notation using VexFlow, 2017  
Link: <https://github.com/TEI-Music-SIG/MEItoVexFlow>

## ***7- Present in an open source digital CMS the result of the musical edition.***

If the published musical corpus is a critical edition which is expected to be dynamic, publish on Omeka and VexFlow. Otherwise, for instance in the case of a library of music scores, publish on Omeka and Verovio (Verovio doesn't support dynamic critical notes for now). It is also possible to use Tikit, a customised distribution of Omeka with Verovio as a preconfigured plugin.

### ***Step descriptive terms***

Research activity

- Publishing



## **Resources**

- Omeka  
Author(s): Roy Rosenzweig Center for History and New Media, George Mason University, Corporation for Digital Scholarship  
Link: <https://omeka.org/>
- Omeka Classic User Manual  
Link: <https://omeka.org/classic/docs/>
- Verovio - Tutorial  
Link: <http://www.verovio.org/tutorial.xhtml>
- Verovio. A music notation engraving library  
Link: <http://www.verovio.org/index.xhtml>
- Music notation engraving library for MEI with MusicXML and Humdrum support and various toolkits (JavaScript, Python), 2019  
Link: <https://github.com/rism-ch/verovio>
- The VexFlow Tutorial  
Link: <https://github.com/0xfe/vexflow/wiki/The-VexFlow-Tutorial>
- VexFlow - Music Engraving in JavaScript and HTML5  
Link: <http://www.vexflow.com/>
- A JavaScript library for rendering music notation and guitar tablature., 2019  
Link: <https://github.com/0xfe/vexflow>
- tikit-musica, 2017  
Link: <https://github.com/Acatus-dev/tikit-musica>

## **8- Offer anyone interested the possibility to contribute**

Create on the website a dedicated form for submitting contributions (i.e. complete metadata, reconstitute parts with missing voices, suggest some analysis, etc.). Integrate, after validation by the editorial team, the contributions into the edition. Possibly make the name of the contributors visible (section listing the contributors for example).

## **Step descriptive terms**

Research activity

- Collaboration
- Sharing

***9- Have a reflexion on the quality of the produced data.***

Evaluate the quality of the data thanks to dedicated tools.

***Step descriptive terms***

Research activity

- Clean up
- Meta-Assessing

***Resources***

- Neuma Quality dashboard

Link: <http://neuma.huma-num.fr/quality/>

### 3.15 Aggregate different datasets to obtain a single reference resource

*Maurizio Sanesi (SISMEL), Roberta Giacomi (SISMEL) and Emiliano Degl'Innocenti (CNR-OVI)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_RUBRICA](http://ssk.huma-num.fr/#/scenarios/SSK_sc_RUBRICA)

A researcher has to confront two or more reference resources dataset, in order to find useful material and to satisfy his research question circumscribing and isolating a certain topic of interest. Therefore, he uses PARTHENOS specialized services (Resource Discovery Service, RUBRICA), as a support for his research needs.

#### **Scenario descriptive terms**

Research objects

- Metadata

Discipline

- History



#### **1- A researcher needs a dataset as reference resource.**

A researcher needs a dataset that serves as a complete reference resource (produced by an aggregation of different datasets) for his research needs.

#### **Step descriptive terms**

Research activity

- Interpretation

#### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## ***2- Search the Reference Resource via RDS***

Search the Reference Resource via Resource Discovery Service (RDS) to see if it is already present in the PARTHENOS project files.

### ***Step descriptive terms***

Research activity

- Analysis

### ***Resources***

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## ***3- Starting the research path***

The researcher can start the research path by already having one or more reference resources he wants to use. If no reference resources are present in the project files, the researcher uses the X3ML data integration platform.

### ***Step descriptive terms***

Research activity

- Creation

### ***Resources***

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## ***4- Discover other useful datasets***

In the PARTHENOS platform, through the semantic alignment performed through the BBT, it will be possible for the user, thanks to the RDS, to discover other useful datasets.

### ***Step descriptive terms***

Research activity

- Analysis



## Standards

- XMLResources
- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## **5- Use of RUBRICA through the Project VREs.**

RUBRICA is available through the PARTHENOS VREs on D4Science, and it is integrated into the project services.

### **Step descriptive terms**

Research activity

- Analysis

Standards

- XML
- RDFS

### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## **5- Choosing parameters**

The researcher chooses through the use of the parameters described in the RUBRICA documentation, the elements he wants to aggregate (example: names of people, names of places, dates ...). At the end of the procedure, RUBRICA provides the researcher with a unique Reference Resource.

### **Step descriptive terms**

Research activity

- Analysis

### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments  
Author(s): CNR-ISTI  
Link: <https://services.d4science.org/explore>

### **6- Dissemination of results**

Disseminate and publish the result files of RUBRICA research in the Infrastructure workspace.

### **Step descriptive terms**

Research activity

- Dissemination

### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments  
Author(s): CNR-ISTI  
Link: <https://services.d4science.org/explore>

### 3.16 Aggregate different datasets using RUBRICA

*Maurizio Sanesi (SISMEL), Roberta Giacomini (SISMEL) and Emiliano Degl'Innocenti (CNR-OVI)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_RUBRICA\\_technical](http://ssk.huma-num.fr/#/scenarios/SSK_sc_RUBRICA_technical)

A researcher wants to do a search and aggregation of reference resources using RUBRICA tool.

#### **Scenario descriptive terms**

Research objects

- Metadata

Discipline

- History



#### **1- Find reference resources**

If a researcher wants to find other reference resources beyond his own, he can look for them through the PARTHENOS registry.

#### **Step descriptive terms**

Research activity

- Discovering

#### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>



## ***2. Compression of the search result as a dataset***

The researcher compresses the search result as a dataset into his personal folder (Personal Workspace) of D4Science, in one of the formats supported by RUBRICA (XML-RDF, TTL, NTriples).

### ***Step descriptive terms***

Research activity

- Organizing

### ***Resources***

- D4Science.org Gateway - Explore Virtual Research Environments  
Author(s): CNR-ISTI  
Link: <https://services.d4science.org/explore>

## ***3- The researcher has to upload in RUBRICA his/her datasets***

If the researcher has an external dataset compatible with RUBRICA, he must enter it in his Personal Workspace of D4science to be able to consult it with RUBRICA.

### ***Step descriptive terms***

Research activity

- Organizing

Standards

- XML

### ***Resources***

- D4Science.org Gateway - Explore Virtual Research Environments  
Author(s): CNR-ISTI  
Link: <https://services.d4science.org/explore>

## ***4- Select Datasets***

The researcher must select two of the datasets available in the Personal Workspace, prepared in the previous steps.



### ***Step descriptive terms***

Research activity

- Content Analysis

### ***Resources***

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## ***5- Executing an experiment with RUBRICA***

The researcher has to follow this path on D4science: go to VRE RUBRICA or VRE PARTHENOS LAB, click on Community Provided Services, then Execute an experiment, the Data Aggregation and finally RUBRICA V2.

### ***Step descriptive terms***

Research activity

- Analysis

### ***Resources***

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## ***6- Downloading the tool usage manual***

The researcher has to download the tool usage manual to become familiar with the parameters required for its operation, and with the operation of the tool itself.

### ***Step descriptive terms***

Research activity

- Analysis

Standards

- XML
- RDFS

**Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

**7- Entering correct parameters for the research**

The researcher must enter the required and correct parameters to perform the search (the application provides indications to help the user).

**Step descriptive terms**

Research activity

- Analysis

Standards

- XML
- RDFS

**Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

**8- Reference Resources selection**

The researcher has to select the reference resources he wants to combine between the ones he previously placed in the Personal Workspace

**Step descriptive terms**

Research activity

- Analysis

**Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>



## 9- Starting RUBRICA

The researcher can start the application using the "Start Computation" command.

### **Step descriptive terms**

Research activity

- Analysis

Standards

- XML
- RDFS

### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

## 10- Downloading results

Once RUBRICA has ended its computation, the researcher can download the results choosing the most suitable output format.

### **Step descriptive terms**

Research activity

- Capture

Standards

- XML
- RDFS

### **Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

## 11- RUBRICA documentation analysis

The researcher downloads and analyzes the documentation provided by the application, which contains information on the procedure and actions performed on the data. Once the

data is mounted on a triple store, the documentation provides an example of a query to check the quality of the output data.

**Step descriptive terms**

Research activity

- Analysis

Standards

- XML
- RDFS

**Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

**12- Link for publications**

The user can save a link (equivalent get request) of the results that can be used for citations within publications / papers / other.

**Step descriptive terms**

Research activity

- Dissemination

Standards

- XML
- RDFS

**Resources**

- D4Science.org Gateway - Explore Virtual Research Environments

Author(s): CNR-ISTI

Link: <https://services.d4science.org/explore>

### 3.17 Creation of a TEI-based corpus

Klaus Illmayer (OEAW-ACDH), Susanne Haaf (BBAW) and Piotr Bański (IDS Mannheim, CLARIN)

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_corpusModellingInTEI](http://ssk.huma-num.fr/#/scenarios/SSK_sc_corpusModellingInTEI)

This scenario explains the steps to take to create a corpus based on the TEI tagset. As of today, the TEI guidelines have become a de facto standard for text annotation, providing solutions for a great variety of text and phrase structures, information on content types, linguistic information on words or phrases, etc. In many digital text collections and digital edition projects annotation has been based on the TEI. Linguistic corpora based on TEI may thus be re-used in projects of other disciplines as well or may themselves benefit from the wide range of already existing resources.

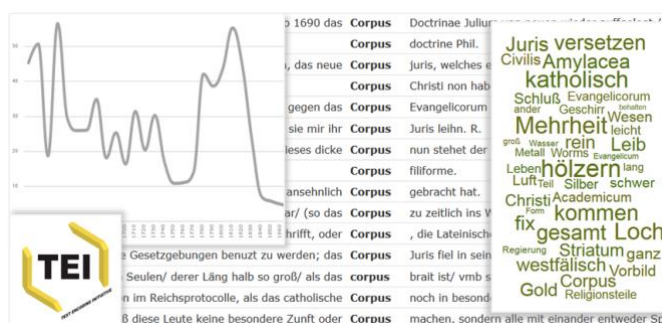
#### Scenario descriptive terms

Discipline

- Linguistics

Research objects

- Text



#### 1- Corpus Composition

For corpus composition a set of canonical decisions should be considered. As a starting point, the corpus type will have to be determined: Is it an opportunistic selection, a balanced corpus, etc.? If balanced, which would be the criteria for this? This is accompanied by the question of which domain the corpus should represent (e.g. should it only contain resources within a specific date range or of a certain text type or from a specific author?). Based on this decision, the resources that should be included in the corpus have to be selected, possibly by reference to a research question. After that, corpus composers have to decide on the type of annotation that should be applied: lightweight or with multiple layers, inline or standoff, etc. The next important step for corpus composition is the determination of a general tagset which the annotation should follow, i.e. the content of the mark up planned.

**Step descriptive terms**

Research activity

- Designing

Standards

- XML
- TEI
- Metadata
- CMDI

**Resources**

- CLARIN Virtual Language Observatory  
Author(s): multiple creators  
Link: <https://vlo.clarin.eu>
- CMDI SPECIFICATION, version 1.2, 2016  
Link: <https://www.clarin.eu/cmd1.2-specification>

**2- Verification and Clean up**

It is straightforward to include a phase of verification of results and clean up after each step towards corpus creation. This step is actually a "hidden" or "stepified" scenario which should take parameters. Verification can be automatic at first, and then manual, done by annotators. This step has the power of looping the entire process back to the previous step (after the/some errors have been corrected). Important: Preserve the data that is about to undergo clean up (simplest: zip with a date stamp, more intricate: use a versioning system).

**Step descriptive terms**

Research activity

- Meta-assessing
- Clean up

Standards

- XML
- TEI



### **3- Conversion to TEI**

For the project at hand, a TEI format has to be chosen or created (the latter by usage of the ODD language) which suits the mark up necessities defined in the corpus composition step. Thus, if digitized data from other sources are to be re-used for corpus creation, these may very likely be available only in formats that aren't similar to the TEI format selected for the corpus creation project at hand. External data may either come in completely different formats or at least in different TEI dialects. In any case, it will be necessary to convert the data from different formats into the TEI output format. Conversion may be conducted semi-automatically.

#### **Step descriptive terms**

Research activity

- Conversion

Standards

- XML
- TEI

#### **Resources**

- The TEI Guidelines. Documentation Elements, 2016  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>
- TEI: Getting Started with ODDs, 2013  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/Guidelines/Customization/odds.xml>
- W3C Recommendation : Extensible Markup Language (XML) 1.0 (Fifth Edition), 2008  
Author(s): World wide web consortium  
Link: <http://www.w3.org/TR/xml/>

### **4- Verification and Clean up**

It is straightforward to include a phase of verification of results and clean up after each step towards corpus creation. This step is actually a "hidden" or "stepified" scenario which should take parameters. Verification can be automatic at first, and then manual, done by annotators. This step has the power of looping the entire process back to the previous step (after



the/some errors have been corrected). Important: Preserve the data that is about to undergo clean up (simplest: zip with a date stamp, more intricate: use a versioning system).

### ***Step descriptive terms***

Research activity

- Meta-assessing
- Clean up

Standards

- XML
- TEI

## ***5- Create Workbench***

After the data has been gathered and homogenized in one similar format, it should be compiled into a proper corpus. This can be done by just bundling the data into one repository, either a file directory or a Github repository.

### ***Step descriptive terms***

Research activity

- Organizing

Standards

- XML
- TEI

## ***6- Verification and Clean up***

It is straightforward to include a phase of verification of results and clean up after each step towards corpus creation. This step is actually a "hidden" or "stepified" scenario which should take parameters. Verification can be automatic at first, and then manual, done by annotators. This step has the power of looping the entire process back to the previous step (after the/some errors have been corrected). Important: Preserve the data that is about to undergo clean up (simplest: zip with a date stamp, more intricate: use a versioning system).



### ***Step descriptive terms***

#### Research activity

- Meta-assessing
- Clean up

#### Standards

- XML
- TEI

### ***7- Linguistic Annotation***

The corpus texts containing TEI text structuring can be further enhanced by adding linguistic mark up, e.g. information on tokens, lemmas, Parts-of-Speech or results of higher level linguistic analysis). Here, a wide range of NLP tools are available to perform the task of linguistic analysis and tagging automatically. For example, the CLARIN infrastructure offers the WebLicht service which allows for data analysis with various NLP tools and for the building of analysis chains. In- and output format here is the Text Corpus Format (TCF; for conversion between TEI and TCF use the TEI2TCF-webservice). Another toolkit is e.g. Apache OpenNLP. As tools you may use e.g. the Stanford Tokenizer or TreeTagger for tokenization, POS tagging and more, or e.g. enrich corpus with Named Entity Recognition annotation (e.g. by usage of Apache Stanbol or Babelify). Manual annotation may e.g. be performed in WebAnno.

### ***Step descriptive terms***

#### Research activity

- Annotating

#### Standards

- XML
- TEI
- TCF
- LAF

### ***Resources***

- Welcome to WebLicht!  
Link: [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page)
- WebLicht: Getting Started, 2017

Author(s): multiple authors

Link: [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting\\_Started](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting_Started)

- CLARIN-D WebAnno

Author(s): Computer Science Department of the Technische Universität Darmstadt

Link: <https://webanno.sfs.uni-tuebingen.de/>

- Welcome to Apache OpenNLP

Link: <http://opennlp.apache.org/index.html>

Abstract: The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text.

- The Stanford Natural Language Processing Group

Link: <https://nlp.stanford.edu/software/>

- TreeTagger - a part-of-speech tagger for many languages

Link: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

- Welcome to apache standbol

Link: <https://stanbol.apache.org/>

- Babelfy

Link: <http://babelfy.org/>

- TEI-TCF encoder+decoder

Author(s): Bryan Jurish

Link: <http://kaskade.dwds.de/tei-tcf/>

## **8- Verification and Clean up**

It is straightforward to include a phase of verification of results and clean up after each step towards corpus creation. This step is actually a "hidden" or "stepified" scenario which should take parameters. Verification can be automatic at first, and then manual, done by annotators. This step has the power of looping the entire process back to the previous step (after the/some errors have been corrected). Important: Preserve the data that is about to undergo clean up (simplest: zip with a date stamp, more intricate: use a versioning system).

### **Step descriptive terms**

Research activity

- Meta-assessing
- Clean up



## Standards

- XML
- TEI

## 9- Finalize

Finalize corpus creation by providing the corpus, its format and bibliographic information on it to the community. Ensure that the FAIR principles are met by its publication. It should be findable (F), accessible (A), interoperable (I) and re-usable (R).

### ***Step descriptive terms***

#### Research activity

- Publishing

#### Standards

- XML
- TEI

### ***Resources***

- The FAIR Data Principles

Link: <https://www.force11.org/group/fairgroup/fairprinciples>

### 3.18 Project-centred EAD customization

*Charles Riondet (Inria)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_schemaCustomization](http://ssk.huma-num.fr/#/scenarios/SSK_sc_schemaCustomization)

TEI ODD can be used to document data models external to the TEI environment. Several projects working with archival standards (in particular EAD) use it as well. PARTHENOS created and maintain an instance of the EAD specification in ODD, that can be used to create project-oriented customizations.

#### ***Scenario descriptive terms***

Discipline

- Library and Information Sciences
- Cultural Heritage and Museology

Research objects

- Metadata

Research techniques

- Encoding
- Mapping



#### ***1- Express an XML schema with ODD***

ODD stands for One Document Does it all. It is a TEI-XML conformant specification format that allows one to customize TEI P5 in a literate programming fashion. Any XML schema can be described with ODD.

#### ***Step descriptive terms***

Research activity

- Modelling

Standards

- XML
- TEI
- EAD



## Resources

- Resolving the Durand Conundrum, 2013  
Author(s): Lou Burnard  
Link: <http://journals.openedition.org/jtei/842>
- The TEI Guidelines. Documentation Elements, 2016  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>
- ODD - TEIWiki  
Author(s): TEI consortium  
Link: <https://wiki.tei-c.org/index.php/ODD>
- TEI: Getting Started with ODDs, 2013  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/Guidelines/Customization/odds.xml>
- Introducing Pure ODD  
Author(s): Lou Burnard  
Link: <http://teic.github.io/TCW/pureODDtutorial.html>
- How to make an ODD automagically  
Author(s): Lou Burnard  
Link: <http://teic.github.io/TCW/howtoGenerate.html>
- How to Update your ODD  
Author(s): Lou Burnard  
Link: <http://teic.github.io/TCW/purifyDoc.html>
- ODD Chaining for Idiots  
Author(s): Lou Burnard  
Link: <http://teic.github.io/TCW/howtoChain.html>
- EAD: Encoded Archival Description (EAD Official Site, Library of Congress)  
Author(s): Library of Congress  
Link: <http://www.loc.gov/ead/>
- Encoded Archival Description Tag Library - Version 2002 (EAD Official Site, Library of Congress)  
Link: <http://www.loc.gov/ead/tglib/index.html>
- standards Library, 2018  
Link: <https://github.com/PARTHENOSWP4/standardsLibrary>

## **2- Express projects requirements in machine-readable format**

With ODD, it is possible to customize a schema in different ways. To add specific, or project-oriented, rules, the best way is to embed schematron rules in the Schema specification.

### **Step descriptive terms**

Research activity

- Modelling

Standards

- XML
- TEI
- EAD
- Schematron

### **Resources**

- EAD-custom-template, 2017  
Link:  
<https://github.com/PARTHENOSWP4/standardsLibrary/blob/master/archivalDescription/EAD/odd/EAD-custom-template.xml>
- Licence Agreement for Publicly Available Standards, 2016  
Link:  
[http://standards.iso.org/ittf/PubliclyAvailableStandards/c055982\\_ISO\\_IEC\\_19757-3\\_2016.zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/c055982_ISO_IEC_19757-3_2016.zip)
- Introduction to Schematron, 2008  
Link: <http://www.mulberrytech.com/papers/schematron-Philly.pdf>
- ISO Schematron Quick Reference , 2009  
Link: [http://www.mulberrytech.com/quickref/schematron\\_rev1.pdf](http://www.mulberrytech.com/quickref/schematron_rev1.pdf)
- ISO Scematron 2016 schemas, 2017  
Link: <https://github.com/Schematron/schema>
- TEI element constraintSpec (constraint on schema)  
Author(s): TEI consortium  
Link: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-constraintSpec.html>
- Library of Standard-based tools for Humanities, 2018  
Link: <https://github.com/PARTHENOSWP4/standardsLibrary>
- EAD-ODD: A solution for project-specific EAD schemes, 2018



Link: <https://hal.inria.fr/hal-01737568/document>

- Towards multiscale archival digital data, 2017

Link: <https://hal.inria.fr/hal-01586389/document>

- XPath and Schematron for TEI Customization

Link:

[http://www.wwp.neu.edu/outreach/seminars/\\_current/presentations/schematron/schematron\\_odd\\_tutorial.xhtml](http://www.wwp.neu.edu/outreach/seminars/_current/presentations/schematron/schematron_odd_tutorial.xhtml)

- Report on Standards, 2017

Link: <https://hal.inria.fr/hal-01503235/document>

### **3- Create associated documentation**

To make you customization easy to read and understand by others, it is important to document all the changes made and present them clearly. ODD allows you to generate such documentation in various formats (HTML, PDF, Docx, etc.)

#### **Step descriptive terms**

Research activity

- Annotating

Standards

- XML
- TEI
- EAD
- Schematron

Resources

- Tei-xsl - TEIWiki

Link: <https://wiki.tei-c.org/index.php/Tei-xsl>

- XSL stylesheets for TEI XML

Link: <http://www.tei-c.org/release/doc/tei-xsl/>

- Customization of EAD for the EHRI project

Author(s): Charles Riondet

Link: [https://cdn.rawgit.com/EHRI/data-validations/92c8e39f/ODD-RelaxNG/EAD/EHRI\\_EAD\\_doc.html](https://cdn.rawgit.com/EHRI/data-validations/92c8e39f/ODD-RelaxNG/EAD/EHRI_EAD_doc.html)

- Report on Standards, 2017



Link: <https://hal.inria.fr/hal-01503235/document>.

#### ***4- Test data against the schema***

An important debug phase is to test relevant data with the new schema to make sure that the behaviour of the added rules is appropriate.

##### ***Step descriptive terms***

Research activity

- Assessing

Standards

- XML
- TEI
- EAD

#### ***5- Share and publish data in this new format***

The data tested and validated with the customized schema can be published on the archival portal.

##### ***Step descriptive terms***

Research activity

- Publishing
- Sharing



## Heritage science scenarios

### ***Introduction***

Heritage Science datasets include multidisciplinary informative texts, records, reports, 2D-3D imaging data, 2D - 3D spectral/analytical data in diverse formats produced through multidisciplinary research and documentation approaches. Heritage Science data, although containing essential information for all domain experts from different disciplines present often accessibility and interoperability issues that are not related to legal protection issues but that are due to the low quality or incompatibility of metadata provided.

Within Task 4.4 and taking advantage of the SSK platform, an effort was put to create Scenarios explaining how to proceed to document step by step an analytical protocol followed for in situ characterization of heritage materials or art objects using, indicatively, widely established techniques as XRF, FTIR or microRaman spectroscopies. The produced and stored metadata schemas following a common general structure will enable progress in data compatibility and sharing in the domain.

The common outline for the documentation of analytical protocols includes the following steps:

#### ***1- Describe the artwork***

#### ***2- Describe the project***

#### ***3- Describe the method***

- ***Use (scope)***
- ***Operational environmental and safety conditions for the measurement***
- ***Equipment specifications***
- ***Measurement range, spot size and spatial resolution/ accuracy***

#### ***4- Describe the measurement.***

- ***Pre-processing/ safe limits***
- ***Pre-processing / selection and annotation of Measurement Sites on reference images***
- ***Pre- processing/preliminary tests - Measurement setup***
- ***Measurement / Spectra acquisition / Storage of data and metadata files***

**5 - Post process and interpret the data**

- ***Stray light subtraction/ Baseline correction.***
- ***Peak / spectral bands assessment.***
- ***Pigment/material identification***

**6 - Disseminate the results.**

- ***Data and metadata storage, publication and sharing***
- ***Prepare report***
- ***Publish the results***

### **3.19 Raman spectrometry for the analysis and identification of pigments on cultural heritage objects**

*Sophia Sotiropoulou (IESL-FORTH), Panayiotis Siozos (IESL-FORTH) and Demetrios Anglos (IESL-FORTH)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_RamanV2](http://ssk.huma-num.fr/#/scenarios/SSK_sc_RamanV2)

This scenario explains how to document step by step the analytical protocol followed for in situ characterization of heritage materials or art objects with microRaman spectroscopy. Raman spectrometry is a widely used technique for examining various types of cultural heritage objects, with significant contributions in studies concentrating on pigment identification. The Raman effect provides a quick and relatively straightforward molecular identification of a material under examination due to the fact that the Raman spectrum can be considered as a fingerprint that is used for compound identification. The method is considered significantly versatile therefore different types of equipment are commercially available such as bench top and laboratory instruments, probes/portable and handheld instruments. These advantages including the simplicity, accuracy and speed of the analysis has made Raman spectroscopy very successful in the field of cultural heritage. The current scenario refers to the documentation of a protocol for Raman analysis directly on cultural heritage object (no sampling).

#### **Scenario descriptive terms**

##### **Discipline**

- History
- Art and art history
- Cultural heritage and museology



#### **1- Describe the artwork**

- Description of the object/monument identity: the title and a short description and context/known history of the artwork/monument are included. Additional information

on the artwork can be also included or linked such as inventory number, photograph(s) and physical dimensions.

- Location of the Site / Museum: location of the permanent position/storage of the artwork/monument
- Owner of the object / Scholar - keeper: Information about the owner or the keeper or scholar owing permission for analysis.
- Date/period: information about the date or period that the artwork/monument was created/constructed including information about the creator.
- Type of object: information about the type of the object such as polychromy on sculpture, panel painting etc.
- Type of material(s) for examination: type(s) of materials for examination. e.g. substrate, coatings, pigments or other materials related to surface finishing decoration.

### ***Step descriptive terms***

Research activity

- Contextualizing

### ***Resources***

- CIDOC CRM | Home  
Link: <http://www.cidoc-crm.org/>
- CIDOC CRM Tutorial  
Link: <http://www.cidoc-crm.org/tutorialPageRes>
- CIDOC CRM, Current version  
Link: <http://www.cidoc-crm.org/Version/version-6.2.4>
- ICS - CRMsci : the Scientific Observation Model  
Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)
- CRMsci | Home  
Link: <http://www.cidoc-crm.org/crmsci/home-1>

## ***2- Describe the project***

- Framework, analytical context (motivation - aim of the analytical campaign); Short presentation of the main objectives of the analysis of the artworks/monument



- Preliminary analysis: Information about any preliminary visual examination and analysis of the artwork using other techniques such as photography, multispectral imaging, XRF, LIBS etc. The preliminary analysis provides information on the condition and the characteristics of the artwork useful for an effective planning and appropriate selection of measurement points for Raman analysis.

Possible source of error: wrong selection of measurement position, relating to previous measurements as LIBS (if ignored or not well-documented)

- Known conservation history / archives / oral communication: Research in digital resources, archives and even oral communication with experts concerning information of past analytical data or conservation interventions.

Possible source of error: possible interference of (aged) conservation materials

### ***Step descriptive terms***

Research activity

- Contextualizing

### ***Resources***

- CIDOC CRM | Home  
Link: <http://www.cidoc-crm.org/>
- CIDOC CRM Tutorial  
Link: <http://www.cidoc-crm.org/tutorialPageRes>
- CIDOC CRM, Current version  
Link: <http://www.cidoc-crm.org/Version/version-6.2.4>
- ICS - CRMsci : the Scientific Observation Model  
Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)
- CRMsci | Home  
Link: <http://www.cidoc-crm.org/crmsci/home-1>

### ***3- Describe the method - Use (scope)***

- Analysis expected outcome Information to be obtained /analysis outcome - research expected result: Identification of inorganic or organic compounds (mainly pigments, dyes, fillers) through the fingerprint vibrational spectra of the molecular structure (amorphous, crystalline and polymorph particles are detectable).

- Skills requirements (level of training of the operator): Scientist specialised in Raman Spectroscopy

### ***Step descriptive terms***

Research activity

- Contextualizing

### ***Resources***

- CIDOC CRM | Home  
Link: <http://www.cidoc-crm.org/>
- CIDOC CRM Tutorial  
Link: <http://www.cidoc-crm.org/tutorialPageRes>
- CIDOC CRM, Current version  
Link: <http://www.cidoc-crm.org/Version/version-6.2.4>
- ICS - CRMsci : the Scientific Observation Model  
Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)
- CRMsci | Home  
Link: <http://www.cidoc-crm.org/crmsci/home-1>

## ***4- Describe the method. Operational environmental and safety conditions for the measurement***

Assure preferably low level of illuminance or dark conditions to avoid any light interference. Assure stability of the device set up and the artwork under investigation - avoid vibrations. The analytical protocol does not involve any risk for the safety of the operators or the public. Possible source of error: Stray light interference: if stray light such as sunlight, fluorescent lamp light or incandescent light is present, should be reported. Moreover, other environmental conditions that can potentially affect the measurements are reported such as levels of temperature or relative humidity.

### ***Resources***

- Raman Spectroscopy of cultural heritage Materials: Overview of Applications and New Frontiers in Instrumentation, Sampling Modalities, and Data Processing, 2016  
Author(s): Francesca Casadio, Céline Daher, Ludovic Bellot-Gurlet



Link: <https://link.springer.com/article/10.1007/s41061-016-0061-z>

- The on-site/remote Raman analysis with mobile instruments: a review of drawbacks and success in cultural heritage studies and other associated fields - Colomban - 2012 - Journal of Raman Spectroscopy - Wiley Online Library  
Link: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrs.4042>

### ***5- Describe the method - Equipment specifications***

The equipment specifications of the instrument and the overall apparatus are documented such as type of equipment, manufacturing company, laser excitation wavelength, optical system (camera) /objective lens used and any other component used.

If possible, a proper excitation line should be used for specific materials. Both in consideration of stability under the laser, the scattering probability and also photochemical damage should be considered.

### ***6- Describe the method. Measurement range, spot size and spatial resolution/ accuracy***

The measurement distance from the object is short, however, a long working distance objective is preferable for in situ measurements to avoid any contact with the artwork surface. The measurement spot size (dimensions of the analysed spot) and the measurement resolution/accuracy depends on the magnification of the objective used.

### ***Step descriptive terms***

Research activity

- Designing

### ***7- Describe the measurement. Pre-processing/ safe limits***

Protocol for safe application (for the object damage protection). Particular attention is put to control the Laser intensity during measurement to avoid any risk of photochemical damage or local overheating of the irradiated area.

Possible source of error: Significant absorption of the laser beam by the material is likely to generate photochemical damage or local overheating, which in turn can produce chemical alteration and degrade the irradiated area of the sample under examination. The laser-



degraded area may also give rise to a Raman spectrum which can be incorrectly attributed as a different material.

### **Resources**

- Raman spectroscopy in cultural heritage: Background paper, 2015  
Author(s): AMCTB No 67 Analytical Methods Committee
- Interferences and Side-effects, 2013  
Link: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119961284.ch2>
- Raman spectroscopy as a means for the identification of plattnerite (PbO<sub>2</sub>), of lead pigments and of their degradation products, 2001  
Author(s): Lucia Burgio, Robin J. H. Clark, Steven Firth  
Link: <https://pubs.rsc.org/en/content/articlelanding/2001/an/b008302j>
- Mitigation strategies for radiation damage in the analysis of ancient materials, 2015  
Author(s): Bertrand Loïc, Schöeder Sebastian, Anglos Demetrios et al.  
Link: <https://doi.org/10.1016/j.trac.2014.10.005>

### **8- Describe the measurement. Pre- processing / selection and annotation of Measurement Sites on reference images**

Select the sites of analysis. Use high resolution detailed reference images of the entire object or of the part under study (in case of a monument) to annotate the points of analysis. Possible source of error: Wrong/ inaccurate indexing of measurement position on the object due to the low resolution of the reference image

### **Resources**

- Non invasive analysis of miniature paintings: proposal for an analytical protocol, 2012  
Author(s): M. Aceto, A. Agostino, G. Fenoglio, M. Gulmini, V. Bianco, E. Pellizzi  
Link: <http://europepmc.org/abstract/med/22391225>



### ***9- Describe the measurement. Pre- processing/preliminary tests - Measurement setup***

Preliminary tests operated on the specimen / area under examination using a minimum laser power. The laser power can be progressively increased as well as the integration time / average of scans can be optimised as needed to obtain a good spectrum.

#### ***Resources***

- Raman spectroscopy in cultural heritage: Background paper, 2015  
Author(s): AMCTB No 67 Analytical Methods Committee

### ***10- Describe the measurement. Measurement / Spectra acquisition / Storage of data and metadata files***

Acquire and store the Raman spectra taken on representative sites on the CH object. A preliminary evaluation of the spectrum is done (assessment of Raman peaks or fluorescence emission).

The conditions for each measurement constitute essential metadata which are documented either in a text file for all the measurements or in form of metadata file for each spectrum file. Laser excitation wavelength; laser power at the sample; spectral range (cut off); spectral resolution; integration time; number of scans.

### ***11- Post process and interpret the data. 1: Stray light subtraction/ Baseline correction.***

Straylight correction (dark subtraction) Baseline correction - Subtraction of the fluorescence background from the spectra.

Possible source of error:

- Fluorescence is the most commonly encountered issue when working with conventional Raman spectroscopy. For cultural heritage applications, it can be due to organic materials with chromophore groups or to inclusions and impurities in some mineral compounds.
- Some new equipment have algorithms for automatic background subtraction, this should be specified.

## Resources

- Fluorescence and Raman spectra on painting materials: reconstruction of spectra with mathematical methods, 2006  
Author(s): Osticioli Iacopo, Zoppi Angela, Castellucci Emilliano Mario
- Subtracted shifted Raman spectroscopy of organic dyes and lakes, 2010  
Author(s): Rosi Francesca, Paolantoni Marco, Clementi Catia, et al.  
Link: <https://doi.org/10.1002/jrs.2447>

### **12- Post process and interpret data. Peak / spectral bands assessment.**

In this step the Raman shift wavenumber and the relative intensity of the observed peaks in the Raman spectrum are identified and documented. The relative intensity can be expressed as weak, moderate or strong.

## Resources

- MOVIDA  
Author(s): Dr. Anna Amat  
Link: [http://www.ars-it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)
- GNU Octave  
Link: <https://www.gnu.org/software/octave/about.html>
- DIAGNOSIS - Cultural Heritage Laser Analysis and Diagnosis Documentation System  
Link: <http://politeia.iesl.forth.gr/diagnosis/Login?lang=en>
- Origin  
Link: <http://www.originlab.com/>

### **13- Post process and interpret the data. Pigment/material identification**

The accurate identification of pigments requires the comparison of the Raman peaks with those reported in databases of Raman spectra of interest for cultural heritage studies. Several databases are available that cover a wide range of materials, from pigments and minerals to organic media and plastics.

Possible source of error: Reference spectra acquired with a different laser with respect to the experimental one, may have variation in relative intensity of band.



## Resources

- Spectral Databases for Raman | IRUG  
Link: <http://www.irug.org/resources/spectral-databases-for-raman>
- UCL Raman Spectroscopic Library  
Author(s): Bell Ian M., Clark Robin J.H., Gibbs Peter J.  
Link: <http://www.chem.ucl.ac.uk/resources/raman/>
- Raman Database  
Link: <https://www.fis.unipr.it/phevix/ramandb.html>
- IBeA Raman Database of Inorganic Artists Material  
Author(s): Madariaga Juan Manuel, Castro Kepa, Perez-Alonso Maite  
Link: <http://www.ehu.eus/udps/database/index1.html>
- RRUFF Project website of integrated database of Raman spectra, X-ray diffraction and chemistry data for minerals.  
Link: <http://rruff.info/>
- Analytical study of 20th century paint  
Link: <https://modern.kikirpa.be/>
- Raman spectroscopic library of natural and synthetic pigments (pre- approximately 1850 AD)., 1997  
Author(s): Bell Ian M., Clark RJ, Gibbs PJ
- Library of FT-Raman spectra of pigments, minerals, pigment media and varnishes, and supplement to existing library of Raman spectra of pigments with visible excitation  
Author(s): Burgio Lucia, Clark Robin JH

### ***14- Disseminate the results. Data and metadata storage, publication and sharing***

Organize data and metadata in standard format for open access repositories. The metadata should incorporate descriptions of the equipment, instrumental setup and measurement conditions and information related to the artwork/monument. Finally, the (meta)data should follow the FAIR principles.

### ***Step descriptive terms***

Research activity

- Annotating

**Resources**

- The FAIR Guiding Principles for scientific data management and stewardship, 2016  
Author(s): Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg et al.  
Link: <http://www.nature.com/articles/sdata201618>
- H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, 2016  
Link:  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- The FAIR Data Principles  
Link: <https://www.force11.org/group/fairgroup/fairprinciples>

**15- Disseminate the results. Prepare report**

Prepare a document that includes the objectives of the project/study, information on the artwork, technical /experimental information, selection of representative data obtained, results and conclusions on the overall work.

**Step descriptive terms**

Research activity

- Gathering

**Resources**

- MOVIDA  
Author(s): Dr. Anna Amat  
Link:  
[http://www.arsit.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.arsit.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)
- POLYGNOSIS - A knowledge platform for laser analysis, diagnosis and conservation of Cultural Heritage  
Link: <http://politeia.iesl.forth.gr/polygnosis/>

**16- Disseminate the results.**

Publish the results according to the licencing/IPR / reuse restrictions



### ***Step descriptive terms***

Research activity

- Publishing

### ***Resources***

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

### ***3.20 In situ X-Ray fluorescence analysis for the characterisation of materials on cultural heritage objects***

*Laura Cartechini (CNR-ISTM)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_FTIR](http://ssk.huma-num.fr/#/scenarios/SSK_sc_FTIR)

This scenario explains how to document step by step the analytical protocol followed for in situ characterization of heritage materials or art objects using X-Ray fluorescence spectroscopy. X-ray fluorescence spectroscopy is a well-established method for the in situ investigation of archaeological objects and artworks through the characterization of the chemical (elemental) composition of materials. The fact that the method is essentially non-destructive makes it particularly attractive and widely used by heritage scientists, art historians, archaeologists, curators and conservators. The benefits of XRF include fast analysis, applicability to any type of object or composite material of various shapes and dimensions, suitability for the detection of major to trace elements and semi quantitative capabilities.

#### ***Scenario descriptive terms***

##### **Discipline**

- History
- Art and art history
- Cultural heritage and museology



#### ***1- Describe the artwork***

- Description of the object/monument identity: The title and a short description and context/known history of the artwork/monument are included. Additional information on the artwork can be also included or linked such as inventory number, photograph(s) and physical dimensions.
- Location of the Site / Museum: Location of the permanent position/storage of the artwork/monument
- Owner of the object / Scholar - keeper: Information about the owner or the keeper or scholar owing permission for analysis.



- Date/period: Information about the date or period that the artwork/monument was created/constructed including information about the creator.
- Type of object: Information about the type of the object such as polychromy on sculpture, panel painting etc.
- Type of material(s) for examination: Type(s) of materials for examination. e.g. substrate, coatings, pigments or other materials related to surface finishing decoration.

### ***Step descriptive terms***

#### Research activity

- Contextualizing

### ***Resources***

- CIDOC CRM Tutorial  
Link: [http://www.cidoc-crm.org/cidoc\\_tutorial/index.html](http://www.cidoc-crm.org/cidoc_tutorial/index.html)
- Definition of the CIDOC Conceptual Reference Model , 2011  
Link: [http://new.cidoc-crm.org/sites/default/files/cidoc\\_crm\\_version\\_5.0.4.pdf](http://new.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf)
- Home | CIDOC CRM  
Link: <http://www.cidoc-crm.org/>
- ICS - CRMsci : the Scientific Observation Model  
Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)
- Home | CRMsci  
Link: <http://www.cidoc-crm.org/crmsci/home-1>

## ***2- Describe the project***

- Framework, analytical context (motivation - aim of the analytical campaign); Short presentation of the main objectives of the analysis of the artworks/monument
- Preliminary analysis: Information about any preliminary visual examination and analysis of the artwork using other techniques such as photography, multispectral imaging, etc. The preliminary analysis provides information on the condition and the characteristics of the artwork useful for an effective planning and appropriate selection of measurement points for consequent analysis.
- Possible source of error: wrong selection of measurement position, relating to previous measurements (if ignored or not well-documented)



- Known conservation history / archives / oral communication: Research in digital resources, archives and even oral communication with experts concerning information of past analytical data or conservation interventions
- Possible source of error: possible interference of (aged) conservation materials

### ***Step descriptive terms***

Research activity

- Contextualizing

### ***Resources***

- CIDOC CRM Tutorial  
Link: [http://www.cidoc-crm.org/cidoc\\_tutorial/index.html](http://www.cidoc-crm.org/cidoc_tutorial/index.html)
- Definition of the CIDOC Conceptual Reference Model , 2011  
Link: [http://new.cidoc-crm.org/sites/default/files/cidoc\\_crm\\_version\\_5.0.4.pdf](http://new.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf)
- Home | CIDOC CRM  
Link: <http://www.cidoc-crm.org/>
- ICS - CRMsci : the Scientific Observation Model  
Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)
- Home | CRMsci  
Link: <http://www.cidoc-crm.org/crmsci/home-1>

### ***3- Describe the method - Use (scope)***

- Scope: (scientific analysis? Archaeological excavation? conservation in course? Demonstration? Promotion? Service?)
- Analysis expected outcome: Identification of inorganic compounds through elemental analysis.
- Skills requirements (level of training of the operator): Scientist specialised in XRF Spectroscopy

### ***Step descriptive terms***

Research activity

- Contextualizing



## **Resources**

- Use of microscopic XRF for non-destructive analysis in art and archaeometry, 2000  
Author(s): K. Janssens, G. Vittiglio, I. Deraedt, et al.  
Link: <http://doi.wiley.com/10.1002/%28SICI%291097-4539%28200001/02%2929%3A1%3C73%3A%3AAID-XRS416%3E3.0.CO%3B2-M>
- An Introduction to X-Ray Fluorescence (XRF) Analysis in Archaeology, 2011  
Author(s): M. Steven Shackley  
Link: [https://doi.org/10.1007/978-1-4419-6886-9\\_2](https://doi.org/10.1007/978-1-4419-6886-9_2)
- Non-Invasive and Non-Destructive Examination of Artistic Pigments, Paints, and Paintings by Means of X-Ray Methods, 2017  
Author(s): Koen Janssens, Geert Van der Snickt, et al.  
Link: [https://doi.org/10.1007/978-3-319-52804-5\\_3](https://doi.org/10.1007/978-3-319-52804-5_3)

## **4- Describe the method. Operational environmental and safety conditions for the measurement**

Bearing in mind the high penetrating power of X-Rays,

- Choose carefully the support/check also the back of the artwork to avoid interferences.
- Determine and mark carefully the security restricted area and define the set-up geometry (X-ray beam direction) for safety of the operators and public.

## **5- Describe the method - Equipment specifications**

The equipment specifications of the instrument and the overall apparatus are documented such as type of equipment, manufacturing company, anode (Rh, Ro, Ag, Au, W etc.), XRF tube operational range parameters (Voltage, Current, acquisition time), optical system (camera) /objective lens used and any other component used.

## **6- Describe the method. Reference image (scale / multiscale / map / 2D / 3D)**

Use (or Acquire) high resolution images of the entire object and of the details under study as references. The detailed reference images will be used to document the ROIs of analysis.

### ***7- Describe the measurement: Pre-processing/ safe limits***

A protocol for safe application (for the user and the object damage protection) is necessary to ensure safe operating procedures to avoid human exposure to X-ray radiation

#### ***Resources***

- MOVIDA

Author(s): Dr. Anna Amat

Link: <http://www.ars->

it.org/index.php?option=com\_content&view=article&id=71:movida&catid=78&Itemid=471

### ***8- Describe the measurement: Pre-processing / selection and annotation of measurement sites on reference images.***

- Select the sites of analysis.
- Use (or Acquire) high resolution detailed images of the entire object and of the part under study as references.
- Zoom in and acquire high resolution image of the detail under examination. Select the area of measurement and use the detailed reference images to annotate the points of analysis.

### ***9- Describe the measurement. Pre-processing / preliminary tests - Measurement setup***

The instrument head must access the point of analysis, taking into account that the working distance must be optimized checking the signal amplitude. A tentative spectrum is obtained on the specimen / area under examination acquiring the number of scans that seems appropriate. The scans may be increased according to the final output.

Depending on surface morphology and composition, it may be necessary to optimize the acquisition parameters to improve the spectral profile.

### ***10- Describe the measurement. Measurement / Spectra acquisition / Storage of data and metadata files***

- Record data taken on representative sites on the CH object. A preliminary evaluation of the spectrum is relevant (assessment of major, minor elements present)



- Store data and metadata files
- The measurement conditions (Voltage, current, acquisition time, any filter used) are documented in a text file for all the measurements or in form of metadata file for each spectrum file.

### **Resources**

- XGLab - X and Gamma Ray Electronics

Author(s): XGLab S.R.L

Link: <https://www.xglab.it/>

### **11- Post-process and interpret data. Peak / spectral bands assessment**

Assess the spectra and assign the peaks to major, minor or trace elements

### **Resources**

- MOVIDA

Author(s): Dr. Anna Amat

Link: <http://www.ars->

[it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

### **12- Post-process and interpret data. Material identification/peak assignment**

Interpret the detected elements for the characterisation of materials

### **13- Disseminate the results. Data and metadata storage, publication and sharing**

Organize data and metadata in standard format for open access repositories. The metadata should incorporate descriptions of the equipment, instrumental setup and measurement conditions and information related to the artwork/monument. Finally, the (meta)data should follow the FAIR principles.

### **Step descriptive terms**

Research activity

- Annotating

**Resources**

- The FAIR Guiding Principles for scientific data management and stewardship, 2016  
Author(s): Mark D. Wilkinson, Michel Dumontier, et al.  
Link: <http://www.nature.com/articles/sdata201618>
- H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, 2016  
Link:  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- The FAIR Data Principles  
Link: <https://www.force11.org/group/fairgroup/fairprinciples>

**14- Disseminate the results. Prepare report**

Prepare a document that includes the objectives of the project/study, information on the artwork, technical /experimental information, selection of representative data obtained, results and conclusions on the overall work.

**Step descriptive terms**

Research activity

- Gathering

**Resources**

- MOVIDA  
Author(s): Dr. Anna Amat  
Link: [http://www.ars-it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)
- POLYGNOSIS - A knowledge platform for laser analysis, diagnosis and conservation of Cultural Heritage  
Link: <http://politeia.iesl.forth.gr/polygnosis/>

**15. Disseminate the results. Publish the results**

Publish the results according to the licencing/IPR / reuse restrictions



### ***Step descriptive terms***

Research activity

- Publishing

### ***Resources***

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

### 3.21 In situ FTIR Analysis for the characterisation of materials on archaeological or cultural heritage objects

*Barbara Salvadori, CNR-ICVBC*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_FTIR](http://ssk.huma-num.fr/#/scenarios/SSK_sc_FTIR)

This scenario explains how to document step by step the analytical protocol followed for in situ characterization of heritage materials or art objects using FTIR spectroscopy in reflectance mode.

FTIR spectroscopy is a widely used technique for examining various types of cultural heritage objects, with significant contributions in studies concerning chemical characterization of materials, both organic (binding media, varnishes, adhesives, coatings, consolidants, etc.) and inorganic (pigments, corrosion products, salts, etc.). The main goals are the evaluation of the state of conservation of artworks, characterization of their constituent materials as well as the monitoring of restoration interventions. FTIR spectroscopy can be applied to all the substrates (paintings, metals, stone, paper,...). The FTIR spectroscopy provides quick and relatively straightforward molecular identification of materials under examination, due to the fact that the collective position and pattern of absorption bands in the FTIR spectrum designate the combination of molecular groups found in any specific compound. The method is significantly versatile, therefore different types of equipment are commercially available such as benchtop instruments, probes/portable and handheld instruments. These advantages including the simplicity, accuracy and speed of the analysis has made FTIR spectroscopy very successful in the field of cultural heritage. The current scenario presents a methodology for FTIR analysis directly on cultural heritage objects (no sampling).





## ***Scenario descriptive terms***

### Research techniques

- Scanning

### Disciplines

- History
- Art and art history
- Cultural heritage and museology

## ***1 - Describe the artwork***

- Description of the object/monument identity: The title and a short description and context/known history of the artwork/monument are included. Additional information on the artwork can be also included or linked such as inventory number, photograph(s) and physical dimensions.
- Location of the Site/ Museum: Location of the permanent position/storage of the artwork/monument
- Owner of the object/ Scholar - keeper: Information about the owner or the keeper or scholar owing permission for analysis.
- Date/period: Information about the date or period that the artwork/monument was created/constructed including information about the creator.
- Type of object: Information about the type of the object such as polychromy on sculpture, panel painting etc.
- Type of material(s) for examination: Type(s) of materials for examination. e.g. substrate, coatings, pigments or other materials related to surface finishing decoration.

## ***Step descriptive terms***

### Research activity

- Contextualizing

## ***Resources***

- CIDOC CRM Tutorial  
Link: [http://www.cidoc-crm.org/cidoc\\_tutorial/index.html](http://www.cidoc-crm.org/cidoc_tutorial/index.html)
- Definition of the CIDOC Conceptual Reference Model, 2011  
Link: [http://new.cidoc-crm.org/sites/default/files/cidoc\\_crm\\_version\\_5.0.4.pdf](http://new.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf)



- Home | CIDOC CRM

Link: <http://www.cidoc-crm.org/>

- ICS - CRMsci : the Scientific Observation Model

Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)

- Home | CRMsci

Link: <http://www.cidoc-crm.org/crmsci/home-1>

## **2- Describe the project**

- Framework, analytical context (motivation - aim of the analytical campaign); Short presentation of the main objectives of the analysis of the artworks/monument
- Preliminary analysis: Information about any preliminary visual examination and analysis of the artwork using other techniques such as photography, multispectral imaging, XRF, LIBS etc. The preliminary analysis provides information on the condition and the characteristics of the artwork useful for an effective planning and appropriate selection of measurement points for Raman analysis.
- Possible source of error: wrong selection of measurement position, relating to previous measurements as LIBS (if ignored or not well-documented)
- Known conservation history / archives / oral communication: Research in digital resources, archives and even oral communication with experts concerning information of past analytical data or conservation interventions
- Possible source of error: possible interference of (aged) conservation materials

### **Step descriptive terms**

Research activity

- Contextualizing

### **Resources**

- CIDOC CRM Tutorial

Link: [http://www.cidoc-crm.org/cidoc\\_tutorial/index.html](http://www.cidoc-crm.org/cidoc_tutorial/index.html)

- CIDOC-CRM current version

Link: <http://www.cidoc-crm.org/versions-of-the-cidoc-crm>

- ICS - CRMsci : the Scientific Observation Model

Link: [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](https://www.ics.forth.gr/isl/index_main.php?l=e&c=663)



- Home | CRMsci  
Link: <http://www.cidoc-crm.org/crmsci/home-1>
- Home | CIDOC CRM  
Link: <http://www.cidoc-crm.org/>

### **3- Describe the method - Use (scope)**

- Scope: non invasive scientific analysis for material characterisation of the artwork, individuation of alteration products, detection and/or monitoring of materials for conservation
- Information obtained /analysis outcome - research expected result: Identification of inorganic or organic compounds (pigments, dyes, binders, coatings, salts, etc.) through the excitation of vibrations of bonds in molecules by infrared radiation / vibrational spectra of the molecules.

### **Step descriptive terms**

Research activity

- Contextualizing

### **Resources**

- Portable Raman versus portable mid-FTIR reflectance instruments to monitor synthetic treatments used for the conservation of monument surfaces, 2013  
Author(s): Claudia Conti, Jana Striova, Irene Aliatis et al.
- An in situ multi-analytical approach in the restoration of bronze artefacts, 2016  
Author(s): Paola Letardi, Barbara Salvadori, Monica Galeotti et al.  
Link: <http://www.sciencedirect.com/science/article/pii/S0026265X1500288X>
- Characterization of Pigment-Binding Media Systems - Comparison of Non-invasive In-situ Reflection FTIR with Transmission FTIR Microscopy, 2011  
Author(s): Wilfried Vetter, Manfred Schreiner  
Link: <https://core.ac.uk/display/27719081>
- Non-invasive identification of plastic materials in museum collections with portable FTIR reflectance spectroscopy: Reference database and practical applications, 2016  
Author(s): Daniela Saviello, Lucia Toniolo, Sara Goidanich, Francesca Casadio  
Link: <http://www.sciencedirect.com/science/article/pii/S0026265X15001617>

#### ***4- Describe the method - Equipment specifications***

The equipment specifications of the instrument and the overall apparatus are documented such as type of equipment, manufacturing company, laser excitation wavelength, optical system (camera) /objective lens used and any other component used.

#### ***5- Describe the method. Measurement range, spot size and spatial resolution/ accuracy***

Contact or short distance from the surface is required: The instrument head must access the point of analysis, taking into account that the working distance must be optimized checking the signal amplitude. The measurement spot size (dimensions of the analysed spot) and the measurement resolution/accuracy are documented".

#### ***Step descriptive terms***

Research activity

- Designing

#### ***6-Describe the measurement. Pre-processing / selection and annotation of measurement sites on reference images***

- Select the sites of analysis. The point of analysis (3 to 5 mm diameter) has to be optically flat to inhibit diffuse reflection. Acquire high resolution detailed reference images of the entire object or of the part under study (in case of a monument). Use the detailed reference images to document the points of analysis.
- Possible source of error: Diffuse reflection results from a combined process including reflection, refraction and diffraction of the incident radiation due to interactions with particles of the sample. Diffuse reflection spectra appear to be similar to transmission spectra derived from poorly prepared KBr disks.

#### ***7- Describe the measurement. Pre-processing / preliminary tests - Measurement setup***

The instrument head must access the point of analysis, taking into account that the working distance must be optimized checking the signal amplitude. A tentative spectrum is obtained on the specimen / area under examination acquiring the number of scans that seems appropriate. The scans may be increased according to the final output.



Depending on surface morphology and composition, it may be necessary to optimize the acquisition parameters to improve the spectral profile.

#### ***8- Describe the measurement. Measurement / Spectra acquisition / Storage of data and metadata files***

Recording of the FTIR spectra taken on representative sites on the CH object. First evaluation of the spectrum (assessment of FTIR peaks and eventually distortions). The measurement conditions (number of scans, acquisition spectral range, spectral resolution) are documented in a text file for all the measurements or in form of metadata file for each spectrum file.

The position of the instrument must be kept as fixed as possible to avoid artefacts and data rejection by the processing system.

#### **Resources**

- JCAMP-DX range of scientific standards from the Joint Committee on Atomic and Molecular Physical Data (JCAMP)  
Link: <http://www.jcamp-dx.org/>
- Revised JCAMP-DX Spectral File Format for Submissions to the Infrared & Raman Users Group (IRUG) Spectral Database  
Link: <http://irug.org/uploads/documentation/irug-jcamp-dx-revised-white-paper-text-only-with-2b-version-1-26-sept-2013.pdf>
- MOVIDA  
Author(s): Dr. Anna Amat  
Link: [http://www.ars-it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)
- DIAGNOSIS - Cultural Heritage Laser Analysis and Diagnosis Documentation System  
Link: <http://politeia.iesl.forth.gr/diagnosis/Login?lang=en>

### **9- Post-process and interpret data. Baseline correction / Kramers-Kronig transformation**

Baseline correction. In addition, when specular reflection leads to first derivative-like spectral features and strong inverted bands (so called Reststrahlen bands), mathematical transformation (Kramers-Kronig transformation) has to be applied in order to gain absorbance-like spectra which may be compared to actual database spectra. Some drawbacks of this method mainly concern the distortion of the spectra in the fingerprint region. These distortions are caused by anomalous dispersion due to variations in the refractive index across an absorption band

#### **Resources**

- Fourier Transform Infrared Spectrometry, 2006  
Link: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470106310.fmatter>

### **10- Post-process and interpret data. Peak / spectral bands assessment**

In this step the band position (wavenumber) and the relative intensity of the observed peaks in the FTIR spectrum are identified and assigned to functional groups vibrations. The relative intensity can be expressed as weak, moderate or strong.

#### **Resources**

- Infrared Spectroscopy in Conservation Science, 1999  
Authors: Michele R. Derrick, Dusan C. Stulik, James M. Landry  
Link: <http://www.getty.edu/publications/virtuallibrary/0892364696.html>
- MOVIDA  
Author(s): Dr. Anna Amat  
Link: [http://www.ars-it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

### **11- Post-process and interpret data. Material identification**

Following bands assignment, the accurate identification of materials requires the comparison of the FTIR bands with those reported in databases of FTIR spectra of interest for cultural heritage studies. Some databases are available that cover a wide range of materials, from pigments and minerals to organic media and plastics.



The assignment of bands obtained in total reflection mode is a critical issue due to the possible distortion of the bands in the fingerprint region. Special care must be taken to correctly interpret derivative-like peaks and inverted bands.

### **Resources**

- PolIReS IFAC-CNR - Polimers Infrared Reflectance Spectra database  
Link: <http://polires.ifac.cnr.it/index.php>
- Search IRUG Spectral Database | IRUG  
Link: <http://www.irug.org/search-spectral-database>  
Link: [http://lisa.chem.ut.ee/IR\\_spectra/](http://lisa.chem.ut.ee/IR_spectra/)

### **12- Disseminate the results. Data and metadata storage, publication and sharing**

Organize data and metadata in standard format for open access repositories. The metadata should incorporate descriptions of the equipment, instrumental setup and measurement conditions and information related to the artwork/monument. Finally, the (meta)data should follow the FAIR principles.

### **Step descriptive terms**

Research activity

- Annotating

### **Resources**

- The FAIR Guiding Principles for scientific data management and stewardship, 2016  
Author(s): Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg et al.  
Link: <http://www.nature.com/articles/sdata201618>
- H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, 2016  
Link:  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- The FAIR Data Principles  
Link: <https://www.force11.org/group/fairgroup/fairprinciples>

**13- Disseminate the results. Prepare report**

Prepare a document that includes the objectives of the project/study, information on the artwork, technical /experimental information, selection of representative data obtained, results and conclusions on the overall work.

**Step descriptive terms**

Research activity

- Gathering

**Resources**

- MOVIDA

Author(s): Dr. Anna Amat

Link: [http://www.ars-](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

[it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

- POLYGNOSIS - A knowledge platform for laser analysis, diagnosis and conservation of Cultural Heritage

Link: <http://politeia.iesl.forth.gr/polygnosis/>

**14- Disseminate the results. Publish the results**

Publish the results according to the licencing/IPR / reuse restrictions.

**Step descriptive terms**

Research activity

- Publishing

**Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

### 3.22 Macro-X-ray fluorescence mapping of materials on cultural heritage objects,

*Laura Cartechini, CNR-ISTM*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_XRFmapping](http://ssk.huma-num.fr/#/scenarios/SSK_sc_XRFmapping)

This scenario explains how to document step by step the analytical protocol followed for the application of Scanning Macro-X-ray fluorescence for mapping materials on art objects. Scanning X-ray fluorescence spectroscopy (MAXRF) encompasses all the benefits of conventional point analysis XRF for the in situ characterization of the chemical (major to trace elements) composition of materials in archaeological objects and artworks. In addition, it offers the invaluable information about distribution of these materials on the investigated surface, which can provide a unique insight into the creative process of the artist (revealing hidden sub-surface layers) but also can uncover restorations on the surface. Then, elemental maps can be used for direct comparison with other spectral imaging data.



#### 1- Describe the artwork

- Description of the object/monument identity: The title and a short description and context/known history of the artwork/monument are included. Additional information on the artwork can be also included or linked such as inventory number, photograph(s) and physical dimensions.



- Location of the Site / Museum: Location of the permanent position/storage of the artwork/monument
- Owner of the object / Scholar - keeper: Information about the owner or the keeper or scholar owing permission for analysis.
- Date/period: Information about the date or period that the artwork/monument was created/constructed including information about the creator.
- Type of object: Information about the type of the object such as polychromy on sculpture, panel painting etc.
- Type of material(s) for examination: Type(s) of materials for examination. e.g. substrate, coatings, pigments or other materials related to surface finishing decoration.

### ***Step descriptive terms***

Research activity

- Contextualizing

## ***2- Describe the project***

- Framework, analytical context (motivation - aim of the analytical campaign); Short presentation of the main objectives of the analysis of the artworks/monument
- Preliminary analysis: Information about any preliminary visual examination and analysis of the artwork using other techniques such as photography, multispectral imaging, pXRF, etc. The preliminary analysis provides information on the condition and the characteristics of the artwork useful for an effective planning of consequent analysis.
- Known conservation history / archives / oral communication: Research in digital resources, archives and even oral communication with experts concerning information of past analytical data or conservation interventions.  
Possible source of error: possible interference of (aged) conservation materials.

### ***Step descriptive terms***

Research activity

- Contextualizing



### **3- Describe the method - Use (scope)**

- Scope: (scientific analysis? Archaeological excavation? conservation in course? Demonstration? Promotion? Service?)
- Skills requirements (level of training of the operator): Scientist trained in macroXRF mapping analysis.

### **Step descriptive terms**

Research activity

- Contextualizing

### **Resources**

- CRONO: a fast and reconfigurable macro X-ray fluorescence scanner for in-situ investigations of polychrome surfaces  
Author(s): R. Alberti, T. Frizzi, L. Bombelli  
Link: <https://onlinelibrary.wiley.com/doi/full/10.1002/xrs.2741>
- Integrated X-ray fluorescence and diffuse visible-to-near-infrared reflectance scanner for standoff elemental and molecular spectroscopic imaging of paints and works on paper  
Author(s): John K. Delaney, Damon M. Conover  
Link: <https://doi.org/10.1186/s40494-018-0197-y>
- XRF and reflectance hyperspectral imaging on a 15th century illuminated manuscript: combining imaging and quantitative analysis to understand the artist's technique  
Authors(s): Laurence de Viguierie, Sophie Rochut, Matthias Alfeld  
Link: <https://doi.org/10.1186/s40494-018-0177-2>

### **4- Describe the method. Operational environmental and safety conditions for the measurement**

Bearing in mind the high penetrating power of X-Rays,

- Choose carefully the support/check also the back of the artwork to avoid interferences.
- Determine and mark carefully the security restricted area and define the set-up geometry (X-ray beam direction) for safety of the operators and public.

### ***5- Describe the method - Equipment specifications***

The equipment specifications of the instrument and the overall apparatus are documented such as type of equipment, manufacturing company, anode (Rh, Ro, Ag, Au, W etc.), XRF tube operational range parameters (Voltage, Current, acquisition time), optical system (camera)/ objective lens used and any other component used.

### ***6- Describe the method. Measurement range, spot size and spatial resolution/ accuracy***

The measurement conditions are documented and stored. In particular, the measurement distance from the object (contact / short / long / remote), the spot size (dimensions of the analysed area) and the measurement resolution/accuracy

### ***7- Describe the measurement: Pre-processing/ safe limits***

A protocol for safe application (for the user and the object damage protection) is necessary to ensure safe operating procedures to avoid human exposure to X-ray radiation.

### ***8- Describe the measurement. Pre-processing / selection and annotation of measurement sites on reference images***

Select the sites of analysis. The point of analysis (3 to 5 mm diameter) has to be optically flat to inhibit diffuse reflection. Acquire high resolution detailed reference images of the entire object or of the part under study (in case of a monument). Use the detailed reference images to document the areas of analysis.

Possible source of error: Diffuse reflection results from a combined process including reflection, refraction and diffraction of the incident radiation due to interactions with particles of the sample. Diffuse reflection spectra appear to be similar to transmission spectra derived from poorly prepared KBr disks.

### ***9- Describe the measurement. Pre-processing / preliminary tests - Measurement setup***

The instrument head must access the point of analysis, taking into account that the working distance must be optimized checking the signal amplitude. A tentative spectrum is obtained on the specimen / area under examination acquiring the number of scans that seems appropriate. The scans may be increased according to the final output.



Depending on surface morphology and composition, it may be necessary to optimize the acquisition parameters to improve the spectral profile.

#### ***10- Describe the measurement. Measurement / Spectra acquisition / Storage of data and metadata files***

- Record data taken on representative sites on the CH object. A preliminary evaluation of the spectrum is relevant (assessment of major, minor elements present)
- Store data and metadata files
- The measurement conditions (Voltage, current, acquisition time, any filter used) are documented in a text file for all the measurements or in form of metadata file for each spectrum file.

#### ***Resources***

- XGLab - X and Gamma Ray Electronics  
Author(s): XGLab S.R.L  
Link: <https://www.xglab.it/>  
Abstract: The core-business of XGLab is the development of high-performance electronics and instrumentation for X-ray and Gamma-ray applications.

#### ***11- Post-process and interpret data. Peak / spectral bands assessment***

Assess the spectra and assign the peaks to major, minor or trace elements

#### ***Resources***

- MOVIDA  
Author(s): Dr. Anna Amat  
Link: [http://www.ars-it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

#### ***12- Post-process and interpret data. Material identification/peak assignment***

Interpret the detected elements for the characterisation of materials

**13- Disseminate the results. Data and metadata storage, publication and sharing**

Organize data and metadata in standard format for open access repositories. The metadata should incorporate descriptions of the equipment, instrumental setup and measurement conditions and information related to the artwork/monument. Finally, the (meta)data should follow the FAIR principles.

**Step descriptive terms**

Research activity

- Annotating

**Resources**

- The FAIR Guiding Principles for scientific data management and stewardship, 2016  
Author(s): Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, et al.  
Link: <http://www.nature.com/articles/sdata201618>
- H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, 2016  
Link:  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- The FAIR Data Principles  
Link: <https://www.force11.org/group/fairgroup/fairprinciples>

**14- Disseminate the results. Prepare report**

Prepare a document that includes the objectives of the project/study, information on the artwork, technical / experimental information, selection of representative data obtained, results and conclusions on the overall work.

**Step descriptive terms**

Research activity

- Gathering



## **Resources**

- MOVIDA

Author(s): Dr. Anna Amat

Link: <http://www.ars->

[it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

- POLYGNOSIS - A knowledge platform for laser analysis, diagnosis and conservation of Cultural Heritage

Link: <http://politeia.iesl.forth.gr/polygnosis/>

## **15- Disseminate the results. Publish the results**

Publish the results according to the licencing/IPR / reuse restrictions

## **Step descriptive terms**

Research activity

- Publishing

## **Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

Abstract: HAL is an open archive where authors can deposit scholarly documents from all academic fields.

### 3.23 Multispectral imaging for the in-situ characterization of painted works of art

*Panayotis Siozos (IESL-FORTH), Demetrios Anglos (IESL-FORTH)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_multispectralImaging](http://ssk.huma-num.fr/#/scenarios/SSK_sc_multispectralImaging)

Multispectral imaging is a commonly used technique that is currently available to the scientist, conservator, archaeologist and art historian for the non-invasive investigation of works of art. It has a relatively low cost, since inexpensive and widely available equipment exists that provides qualitative, non-invasive, affordable and portable tools for the spatial characterization of specific materials or types of materials. Furthermore, extensive research has been undertaken to develop optimized methodologies for the acquisition and processing of multispectral images as published in the literature but also in the context of the CHARISMA project. The main objective of the scenario is to establish standardized procedures for multispectral imaging analysis in order to accomplish qualitative results during the study of artworks.

#### **Scenario descriptive terms**

##### Disciplines

- Art and art history
- Cultural heritage and museology

##### Research techniques

- Photography

##### Research objects

- Artifacts
- Data
- Images



#### **1 - Research on Artwork**

Research in digital resources concerning information of the painting (pigments, drawing, etc.)

#### **Step descriptive terms**

##### Research activity

- Discovering



## 2 - Multispectral imaging setup

The setup for the acquisition of multispectral images consists of:

- One or more radiation sources which provide the incident radiation to the object being studied and, in the case of luminescence techniques, also provide the means to induce the luminescence (the excitation source) in the materials under study;
- A filter or set of filters in order to allow the transmittance of radiation in the wavelength range under study and exclude unwanted radiation from being recorded;
- A detector or recording device which in this work is a commercially available digital camera with a silicon-based sensor, modified by removal of the IR-blocking filter to provide sensitivity from approximately 350 nm to 1100 nm;
- A set of standards to enable the post-processing methods;

### **Step descriptive terms**

Research activity

- Gathering

### **Resources**

- Multispectral Imaging Reflectance and Photo-induced Luminescence modes: A User Manual  
Authors: Joanne Dyer, Giovanni Verri, John Cupitt  
Link:<https://www.britishmuseum.org/pdf/charisma-multispectral-imaging-manual-2013.pdf>

## 3 - Preparation

The preparation procedure consists of the steps for effective image recording. These steps are: a) Positioning the object/artwork b) positioning the calibration targets and c) positioning the camera and illumination sources.

### **Step descriptive terms**

Research activity

- Gathering



**Resources**

- The AIC Guide to Digital Photography and Conservation Documentation  
Authors: Jeffrey Warda, Franziska Frey, Dawn Heller, et al.  
Link: <http://www.conservation-us.org/resources/our-publications/special-projects/the-aic-guide#.WON542iLSUk>

**4 - Image Capture 1: Capturing VIS images**

Visible-reflected (VIS) images corresponds to standard photography and record the reflected light in the visible region (400-700 nm) from a subject when this is illuminated with visible light.

**Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001

**Resources**

- GIMP cross-platform image editor  
Link: <https://www.gimp.org/>
- Electronic still-picture imaging -- Removable memory -- Part 2: TIFF/EP image data format  
Link: <https://www.iso.org/standard/29377.html>

**5 - Image Capture 2: Capturing IRR images**

Infrared-reflected (IRR) images record the reflective properties of the object in the IR range (>700 nm) when this is illuminated with infrared radiation.

**Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001



### **6 - Image Capture 3: Capturing UVR images**

Ultraviolet-reflected (UVR) record the reflected radiation in the ultraviolet region (200-400 nm) from a subject when this is illuminated with ultraviolet radiation.

#### **Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001

### **7 - Image Capture 4: Capturing UVL images**

Ultraviolet-induced luminescence (UVL) images record the emission of light (luminescence) in the visible region (400-700 nm) from a subject when this is illuminated with UV radiation. This image is used to investigate the distribution of luminescent materials, such as organic binders and colourants.

#### **Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001

### **8 - Image Capture 5: Capturing VIL images**

Visible-induced infrared luminescence (VIL) image record the emission of radiation (luminescence) in the infrared region (700-1100 nm) from a subject when this is illuminated with visible light.

#### **Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001

## 9 - Image Capture 6: Capturing VIVL images

Visible-induced visible luminescence (VIVL) image record the emission of light in the visible region (500-700 nm) from a subject when this is illuminated with visible light (400-500 nm).

### **Step descriptive terms**

Research activity

- Imaging

Standards

- ISO 12234-2:2001

## 10 - Image Post-Processing

Applying post-processing methods in order to extract the maximum available information from the images but also to produced standardise (device-independent) images that can be compared directly compared.

### **Step descriptive terms**

Research activity

- Visualization

### **Resources**

- Adobe Photoshop

Link: <http://www.adobe.com/products/photoshop.html>

- ImageJ

Link: <https://imagej.net/Welcome>

## 11 - Prepare report

Organizing data to prepare a document with the results and the conclusion of the multispectral imaging analysis.

### **Step descriptive terms**

Research activity

- Organizing

## 12 - Publish



Publish the results of the artwork's multispectral imaging analysis.

***Step descriptive terms***

Research activity

- Publishing

### 3.24 Documenting application of lasers in conservation/restoration

*Marta Castillejo (CSIC) and Mikel Sanz (CSIC)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_laserconservationrestoration](http://ssk.huma-num.fr/#/scenarios/SSK_sc_laserconservationrestoration)

The standard will indicate a methodology to document the kind, extent, and objectives of laser cleaning, the employed laser cleaning parameters and the obtained results including cleaning quality as well as side effects or induced damage. The standard will contain guidelines to report the aforementioned required steps in the use of lasers for conservation/restoration in order to allow the comparison of results with different treatments performed in different studies.

#### **Scenario descriptive terms**

Research techniques

- Technology preservation

Research objects

- Research results
- Standards

Discipline

- Art and Art History
- Cultural Heritage and Museology



#### **1- Research laser cleaning parameters**

Selecting info for specific analytical studies of the use of laser in conservation.

#### **Step descriptive terms**

Research activity

- Gathering

#### **Resources**

- Handbook on the use of lasers in conversation and conversation Science  
Link: <http://www.science4heritage.org/COSTG7/booklet/>



- Laser conservation of art, 2007

Author(s): A Nevin, P Pouli, S Georgiou, C Fotakis

Link: <http://www.nature.com/nmat/journal/v6/n5/full/nmat1895.html>

## **2- Determining laser characteristics**

Testing and characterization of lasers.

### **Step descriptive terms**

Research activity

- Content analysis

### **Resources**

- Optics and photonics - Lasers and laser-related equipment - Test methods for laser beam power, energy and temporal characteristics

Link: <https://www.iso.org/standard/69232.html>

## **3- Structural analysis**

Obtaining the relevant parameters related to the laser-material interaction for the CH conservation/restoration area.

### **Step descriptive terms**

Research activity

- Content analysis

### **Resources**

- ISO/TR 21254-4:2011

Link: <https://www.iso.org/standard/43004.html>

## **4- Characterization of laser effects**

Assessing the physicochemical effects of the laser on the material.

### **Step descriptive terms**

Research activity

- Interpretation

**Resources**

- libvips

Link: <https://libvips.github.io/libvips/>

**5- Modelling Laser-Material Interactions**

Modelling Laser-Material Interactions in COMSOL Multiphysics.

**Step descriptive terms**

Research activity

- Modelling

**Resources**

- Modelling Laser-Material Interactions in COMSOL Multiphysics

Link: <https://www.comsol.com/blogs/modeling-laser-material-interactions-in-comsol-multiphysics/>

**6- Documentation of laser effects**

Systematically documenting undesired side effects or collateral induced damage.

**Step descriptive terms**

Research activity

- Archiving

**Resources**

- CEN/TC 346 - Conservation of Cultural Heritage

Link:

[https://standards.cen.eu/dyn/www/f?p=204:32:0::::FSP\\_ORG\\_ID,FSP\\_LANG\\_ID:411453,25&cs=17D2D76D6596BE0CAD81B69108A090D68](https://standards.cen.eu/dyn/www/f?p=204:32:0::::FSP_ORG_ID,FSP_LANG_ID:411453,25&cs=17D2D76D6596BE0CAD81B69108A090D68)

**7- Publishing**

Developing a report that contains all the relevant data related to the laser application in conservation.



### ***Step descriptive terms***

Research activity

- Writing

### ***Resources***

- MOVIDA

Author(s): Dr. Anna Amat

Link: <http://www.ars->

[it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

## ***8- Presenting results***

Facilitating the reuse of the report.

### ***Step descriptive terms***

Research activity

- Publishing

### ***Resources***

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>



### 3.25 Standardizing "Laser Induced Breakdown Spectroscopy" analysis of glasses

*Marta Castillejo (CSIC) and Mikel Sanz (CSIC)*

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_LIBS](http://ssk.huma-num.fr/#/scenarios/SSK_sc_LIBS)

A Researcher is interested in accessing systematic works related to Laser-Induced Breakdown Spectroscopy (LIBS) analysis of glass substrates and its application to cultural heritage glass artefacts given the speed, absence of sampling and sample preparation, microscopic resolution and in-situ characterization proper to the technique.

#### **Scenario descriptive terms**

Research techniques

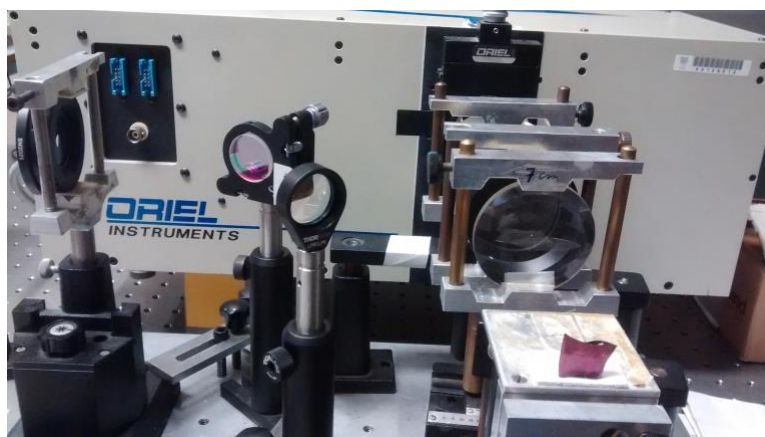
- Technology preservation

Research objects

- Research results
- Standards

Discipline

- Art and art history
- Cultural heritage and museology



#### **1- Research on glasses**

Selecting information for specific analytical studies of glass characterization by Laser-Induced Breakdown Spectroscopy (LIBS).

#### **Step descriptive terms**

Research activity

- Gathering

#### **Resources**

- Laser-Induced Breakdown Spectroscopy in Art and Archaeology, 2001  
Author(s): D. Anglos  
Link: <http://journals.sagepub.com/doi/pdf/10.1366/0003702011952398>



- Laser-Induced Breakdown Spectroscopy in Art and Archaeology, 2001  
Author(s): D. Anglos  
Link: <http://journals.sagepub.com/doi/pdf/10.1366/0003702011952398>

## ***2- Spectra acquisition***

Using LIBS technique in glass substrates for own research studies.

### ***Step descriptive terms***

Research activity

- Imaging

### ***Resources***

- MultiSpec A Freeware Multispectral Image Data Analysis System  
Author(s): David Landgrebe, Larry Biehl  
Link: <https://engineering.purdue.edu/~biehl/MultiSpec/index.html>

## ***3- Pre-processing of spectra***

Applying corrections to spectra for subsequent analysis and interpretation.

### ***Step descriptive terms***

Research activity

- Clean up

### ***Resources***

- Spectrum Analyzer Home Page  
Link: <http://www.physics.muni.cz/~zdenek/span/index.php>

## ***4- Identifying relevant elements***

Qualitative identification of elemental components of glass substrates.

### ***Step descriptive terms***

Research activity

- Contextualizing

**Resources**

- NIST Atomic Spectra Database Lines Form  
Author(s): NIST  
Link: [https://physics.nist.gov/PhysRefData/ASD/lines\\_form.html](https://physics.nist.gov/PhysRefData/ASD/lines_form.html)
- Atomic spectral line database  
Author(s): R. L. Kurucz  
Link: <https://www.cfa.harvard.edu/amp/ampdata/kurucz23/sekur.html>

**5- Analyzing the spectra**

Quantitative analysis (composition from major to minor and trace elements) of glass substrate under study.

**Step descriptive terms**

Research activity

- Content analysis

**Resources**

- MultiSpec A Freeware Multispectral Image Data Analysis System  
Author(s): David Landgrebe, Larry Biehl  
Link: <https://engineering.purdue.edu/~biehl/MultiSpec/index.html>

**6- Glass characterization**

Determining glass composition, glass classification, provenance, chromophores, opacifiers, surface alteration and degradation of usefulness for preservation and restoration of glass artefacts in the CH field.

**Step descriptive terms**

Research activity

- Theorizing

**Resources**

- Evaluation of the analytical potential of laser-induced breakdown spectrometry (LIBS) for the analysis of historical glasses, 2003  
Author(s): K. Muller, H. Stege



Link: <http://onlinelibrary.wiley.com/doi/10.1111/1475-4754.00119/pdf>

- Lead determination in glasses by laser-induced breakdown spectroscopy, 2007

Author(s): Carmona, Oujja, Gaspard, García-Heras, Villegas, Castillejo

Link: <https://doi.org/10.1016/j.sab.2007.01.003>

## **7- Comparison with other techniques**

Compare the obtained results by LIBS with results from other analytical techniques employed in the characterization of the same glass object or other glasses.

### **Step descriptive terms**

Research activity

- Contextualizing

### **Resources**

- Multianalytical characterization of Late Roman glasses including nanosecond and femtosecond laser induced breakdown spectroscopy

Author(s): Oujja, Sanz, Agua, Conde, García-Heras, Dávila, Oñate, Sanguino, Vázquez de Aldana, Moreno, Villegas, Castillejo

Link:

<http://pubs.rsc.org/en/Content/ArticleLanding/2015/JA/C5JA00150A#!divAbstract>

## **8- Elaborate a systematic report**

Developing a report that contains all the relevant data related to the LIBS characterization of glass artefacts.

### **Step descriptive terms**

Research activity

- Writing

## **Resources**

- MOVIDA

Author(s): Dr. Anna Amat

Link: <http://www.ars->

[it.org/index.php?option=com\\_content&view=article&id=71:movida&catid=78&Itemid=471](http://www.ars-it.org/index.php?option=com_content&view=article&id=71:movida&catid=78&Itemid=471)

## **9- Publishing**

Disseminating the characterization report.

### **Step descriptive terms**

Research activity

- Publishing

## **Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>

## **10- Keep access to the report**

Using a sustainable repository.

### **Step descriptive terms**

Research activity

- Archiving

## **Resources**

- The open archive HAL

Link: <https://hal.archives-ouvertes.fr>



## 4 Related Heritage Science standardization activities

As reported in Deliverable 4.2, in the domain of Heritage Science and, more specifically, the domain of analytical investigation of Heritage objects, standardization is not well advanced. Although characterization of materials and objects based on scientific analysis is well-established as integral part of the study, documentation and conservation of objects or monuments in every institution related to CH, there is a rather limited number of procedures or protocols that have been approved as standards.

Institutions or domain experts follow well defined protocols, either based on institution tradition or expert's own experience as well as on generally adopted best practices, which are, however, often designed for specific ends or adapted to available instrumentation or case specific conditions in which the analytical campaign is carried out.

Despite this common practice there is, however, no question about the impact of standard procedures in upgrading the safety of users/operators and objects, the performance of the methods, the quality, reliability of the results.

Further, in the perspective of turning the European Open Science Cloud (EOSC) into reality for Heritage Science, several preliminary measures are required to be taken<sup>1</sup>. To obtain effective interoperability between different Heritage Science communities and integration of inter/cross/trans disciplinary information, rules and standards need to be established for the data produced from the different communities. The development and adoption of “universal” protocols and standards are intended as reference framework to generate valuable information and suitable metadata to be stored, accessed, queried, shared and reused among the institutions, scientists and conservation professionals in various contexts and research scenarios. Exchange and creation of new knowledge across disciplines, actions, research and services put in central focus the efficient interoperability of data (Wilkinson et al. 2016).

The main inhibitors for the interoperability of data across institutions, domain experts and disciplines is the incompatibility of data formats, heterogeneity of data representation (metadata schemas) and lack of common interdisciplinary understanding/interpretation of

---

<sup>1</sup> <https://www.openaire.eu/guides>; <https://www.eosc-portal.eu/>

the data. During this “pre-standardization” phase, documents describing established analytical protocols and best practices need to be shared, assessed and be widely accepted through consensus among experts in the field of Heritage Science.

In this context, the SSK platform provides an excellent forum for domain experts to gather best practices, exchange and agree on specifications and criteria to be applied consistently towards standardization procedures. On the same time, the platform supports educational purposes disseminating scenarios of research for scientists and professionals new in the field or in the use of a specific method, in particular.

It includes:

- Protocols and procedures for the analysis of materials using a specific technique or a group of techniques,
- Protocols and procedures for data management and dissemination.

A synchronized, equally major effort is put towards standardization of documentation schemas that will enable interoperability of data produced and related information obtained through standard protocols. To this end, within Task 4.4, documents describing best practices in applying different established analytical methods for gathering scientific data about the material composition of a cultural heritage objects/monuments have been compiled and evaluated comparatively.

The scope of this task was to generate a common general scheme of documentation and metadata structure adaptable to be specific of the different analytical protocols but retaining the standard outline enabling exchange of information in a proper way. More precisely, SSK scenarios that describe in detail the steps to follow in the application and documentation of specific analytical protocols for the in situ characterization of materials on heritage objects have been enriched with the related structure of metadata for data representation. Furthermore, the metadata have been mapped in CIDOC-CRM (and compatible models) as a common and extensible semantic framework to formulate relevant information on the analytical protocols, related projects, analytical or conservation campaigns and heritage

objects involved in a structured manner compatible for information systems<sup>2</sup>. CIDOC-CRM (Core model + domain-specific extensions, see Figure 1) has been used as a semantic reference model which reinforces compatibility and enables exchange of digital information resources produced and stored in different management systems.

As representative example, the SSK on Raman Analysis is presented in Appendix 1, including a corpus of essential metadata fields responding to domain needs, required to be documented at each step of the SSK and displayed in correspondence with the related CIDOC-CRM terms.

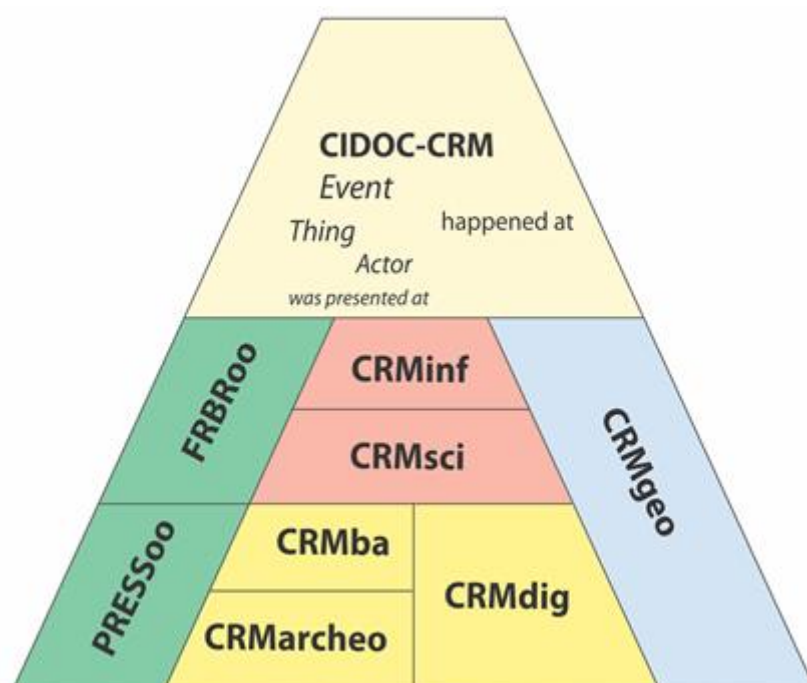


Figure 1: CIDOC-CRM family of models, [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=229](https://www.ics.forth.gr/isl/index_main.php?l=e&c=229)

The metadata format used for this purpose is indicative but representative of the documentation scheme followed by domain experts, in this case by scientists involved in the

<sup>2</sup> CIDOC CRM (ISO 21127:2006) - <http://cidoc-crm.org> ; Core model + domain-specific extensions as CRMsci Scientific observation information - <http://cidoc-crm.org/crmsci>; CRMdig Provenance metadata for digital objects - <http://cidoc-crm.org/crmdig>



application of Raman Spectroscopy for the characterization of heritage materials or art objects. The proposed mapping to CIDOC–CRM is intended to promote a shared understanding of information gathered through an analytical investigation and is explored as a common and extensible semantic framework that any cultural heritage information can be mapped to.

The metadata presented correspond to existing CIDOC-CRM classes, properties and interrelations among them. However, this work is ongoing and is aimed to promote exchange among domain experts towards necessary extensions and definition of new classes and properties on top of the core model and existing compatible models to create a domain ontology, more appropriate for the management, integration, mediation, interchange and access to heritage science data.



## ***Bibliography***

Riondet, Charles, Dorian Seillier, Lionel Tadjou, and Laurent Romary, 2018, [Standardization Survival Kit \(Final\)](https://hal.inria.fr/hal-01925144). <https://hal.inria.fr/hal-01925144>, accessed February 6, 2019.

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

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al., 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship" in *Scientific Data* 3: 160018.

## Annex 1. Heritage Science specific activities

As a representative example, the SSK scenario entitled “entitled “Raman spectrometry for the analysis and identification of pigments on cultural heritage objects” [http://ssk.humanum.fr/#/scenarios/SSK\\_sc\\_RamanV2](http://ssk.humanum.fr/#/scenarios/SSK_sc_RamanV2), is presented here, including a corpus of essential metadata fields responding to domain needs, required to be documented at each step of the SSK and displayed in correspondence with the related CIDOC-CRM terms.

The metadata format presented here is indicative but representative of the documentation scheme followed by domain experts, in this case by scientists involved in the application of Raman Spectroscopy for the characterization of heritage materials or art objects. The proposed mapping to CIDOC–CRM is intended to promote a shared understanding of information gathered through an analytical investigation and is explored as a common and extensible semantic framework that any cultural heritage information can be mapped to.

The metadata presented correspond to existing CIDOC-CRM classes, properties and interrelations among them. However, this work is ongoing and is aimed to promote exchange among domain experts towards necessary extensions and definition of new classes and properties on top of the core model and existing compatible models to create a domain ontology, more appropriate for the management, integration, mediation, interchange and access to heritage science data.



## **Raman spectrometry for the analysis and identification of pigments on cultural heritage objects**

SSK authors: Sophia Sotiropoulou (IESL-FORTH), Panayiotis Siozos (IESL-FORTH) and Demetrios Anglos (IESL-FORTH)

Metadata mapping: Sophia Sotiropoulou, Chrysoula Bekiari (ICS-FORTH), Athina Kritsotaki (ICS-FORTH).

[http://ssk.huma-num.fr/#/scenarios/SSK\\_sc\\_RamanV2](http://ssk.huma-num.fr/#/scenarios/SSK_sc_RamanV2)

This scenario explains how to document step by step, the analytical protocol followed for in situ characterization of heritage materials or art objects with microRaman spectroscopy.

Raman spectrometry is a widely used technique for examining various types of cultural heritage objects, with significant contributions in studies concentrating on pigment identification. The Raman effect provides a quick and relatively straightforward molecular identification of a material under examination due to the fact that the Raman spectrum can be considered as a fingerprint that is used for compound identification. The method is considered significantly versatile, therefore, different types of equipment are commercially available such as bench top and laboratory instruments, probes/portable and handheld instruments. These advantages, including the simplicity, accuracy and speed of the analysis, has made Raman spectroscopy very successful in the field of cultural heritage. The current scenario refers to the documentation of a protocol for Raman analysis directly on cultural heritage objects (no sampling).

### Step 1 - Describe the artwork

Description of the object/monument identity: The title and a short description and context/known history of the artwork/monument are included. Additional information on the artwork can be also included or linked such as inventory number, photograph(s) and physical dimensions.

<b>Identity of the artwork</b> <identification number> Artwork identifier: The object identifier is typically composed of alphanumeric sequence assigned to an object in order to identify it uniquely and permanently within the context of one or more organisations. Such codes are often known as inventory numbers, registration codes, etc.	
Metadata field/term	CIDOC CRM related terms
<Value>: The value of the code number is documented in this field. (Additional metadata fields for the code may be included as <Code>; <Category>; <Attributedby>; <AttributionDate>)	E22 Physical Man-made object: P 48 has preferred identifier: E42 Identifier
<ObjectTitle>: the main title of the Object - the principal denomination with which the art work/ monument is currently known (Additional metadata fields for the Title may be included as <otherObjectTitle>; <Attributedby>)	E22 Physical Man-made object: P102 has title: E 35 Title E22 Physical Man-made object: P1 is identified by: E 41 Appellation
<ShortDescription> A text describing in brief the object or monument <ObjectDescription>: a detailed description of the Object	E22 Physical Man-made object: P62 depicts: E1 CRM Entity E22 Physical Man-made object: P3 has note: E62 String



<p>&lt;ObjectDimensions&gt;: contains information about the dimensions of the object</p> <p>&lt;Type&gt; Contains the type of the dimension (eg. depth, height etc.)</p> <p>&lt;Unit&gt; Contains the counting unit of the dimension (eg. cm, pixels, etc.),</p> <p>&lt;Value&gt; Contains the value of the dimension</p>	<p>E22 Physical Man-made object: P43 has dimension: E54 Dimension</p> <p>E54 Dimension: P2 has type: E55 Type</p> <p>E54 Dimension: P91 has unit: E58 Measurement Unit</p> <p>E54 Dimension: P90 has value: E60 Number</p>
<p>&lt;CharacteristicImage&gt;: a characteristic image of the Object</p>	<p>E22 Physical Man-made object: P138b has representation: E38 Image</p>
<p>&lt;DigitalFile&gt;: a characteristic image of the Object, link to a &lt;Digital Object&gt;</p> <p>&lt;DigitalFileFormat&gt;; &lt;Dimensions&gt;;</p> <p>&lt;Storage Location&gt; Contains the storage location of the &lt;digitalfile&gt;</p> <p>&lt;Note&gt; : any comment about the image of the Object</p>	<p>D1 Digital Object: P70 documents: E22 Physical Man-made object</p> <p>D1 Digital Object: P2 has type:E55 Type</p> <p>D1 Digital Object: P43 has dimension: E54 Dimension</p> <p>D1 Digital Object: P1 is identified by: E41 Appellation</p> <p>D1 Digital Object: P3 has note: E62 String</p>

Location of the Site / Museum: Location of the permanent position/storage of the artwork/monument

Metadata field/term	CIDOC CRM related terms
<p>&lt;CurrentLocation&gt;: contains information about the current location of the object</p>	<p>E19 Physical Object - P54 has current permanent location: E53 Place</p>

Owner of the object / Scholar - keeper: Information about the owner or the keeper or scholar owing permission for analysis.

Metadata field/term	CIDOC CRM related terms
<p>&lt;CurrentOwner&gt;: contains information about the current owner of the object</p> <p>&lt;PersonInvolved&gt; contains the link to the individual current owner or the keeper or scholar owing permission for analysis of the object</p>	<p>E 39 Actor - E 40 Legal body or F10 Person: P105B has right on: E22 Physical Man-made object</p> <p>OR</p> <p>P50B is current keeper of E22 Physical Man-made object</p> <p>OR</p> <p>P52B is current owner of :E22 Physical Man-Made Object</p>

Date/period: Information about the date or period that the artwork/monument was created/constructed including information about the creator.

Metadata field/term	CIDOC CRM related terms
<p>&lt;CreationInformation&gt; contains information about the creation of the object</p> <p>&lt;Creator&gt; contains information about the creator of the object</p> <p>&lt;Creation date&gt; contains the creation date or period of the object</p>	<p>E22 Physical Man-made object: P108 was produced by:E12 Production: P117 occurred during: E4 Period</p> <p>OR</p> <p>P108 was produced by:E12 Production:P14 carried out by:E39 Actor /in the role of: E55 Type</p> <p>OR</p> <p>P108 was produced by:E12 Production:P4 has time-span:E52 Time-span</p>



Type of object: Information about the type of the object such as polychromy on sculpture, panel painting etc.

Metadata field/term	CIDOC CRM related terms
<ObjectCategory>: the type or category of the Object	E22 Physical Man-made object - P2 has type: E55 Type (e.g. Painting)

Type of material(s) for examination: Type(s) of materials for examination. e.g. substrate, coatings, pigments or other materials related to surface finishing decoration.

Metadata field/term	CIDOC CRM related terms
<MaterialCategory>: contains information about the type of materials under investigation	E22 Physical Man-made object: P46 is composed of: E18 Physical Thing P57 has number of parts: E60 Number P45 consists of: E57 Material P44 has condition: E3 Condition State E19 Physical Object P56 - bears feature (is found on) E26 Physical Feature



## Step 2 - Describe the project

Framework, analytical context (motivation - aim of the analytical campaign); Short presentation of the main objectives of the analysis of the artworks/monument

Metadata field/term	CIDOC CRM related terms
<p>&lt;Identity of the project&gt;: information about the project in which the object is investigated.</p> <p>(Additional metadata fields for the code may be included as:</p> <p>&lt;Project Title&gt;: the title of the Project - framework</p> <p>&lt;Dates&gt; : information about the dates of the project;</p> <p>&lt;Starting Date&gt; /&lt;Ending Date&gt;</p> <p>&lt;Project Type&gt;: the type of the Project</p> <p>&lt;Project Goal&gt; : the goal of the project</p> <p>&lt;Agreement&gt; the agreement file of the Project</p>	<p>E7 Activity: P1 is identified by: E41 Appellation</p> <p>E7 Activity: P4 has time-span: E52 Timespan</p> <p>E7 Activity: P2 has type: E55 Type</p> <p>E7 Activity: P21 had general purpose: E55 Type</p> <p>E7 Activity: P17 was motivated by: E1 CRM Entity</p> <p>E7 Activity: P70B is documented in: E31 Document</p>

Preliminary analysis: Information about any preliminary visual examination and analysis of the artwork using other techniques such as photography, multispectral imaging, XRF, LIBS etc. The preliminary analysis provides information on the condition and the characteristics of the artwork useful for an effective planning and appropriate selection of measurement points for Raman analysis.

Metadata field/term	CIDOC CRM related terms
<p>&lt;Condition Report&gt; contains information about the condition of the Object</p> <p>&lt;MacroscopicExamination&gt; contains information about the condition of the object as observed without the use of devices</p>	<p>E14 Condition Assessment</p> <p>E7 Activity: P2 has type: E55 type</p> <p>S4 Observation</p> <p>E7 Activity :P2 has type:E55 type</p>



<p>&lt;Observation&gt; contains information about the condition of the object as observed without the use of devices</p> <p>&lt;RelatedTechnicalExamination&gt; includes information of the related Technical examination that was preliminary applied to the Object during the specific Project</p> <p>&lt;Method 1&gt; the name of the Technical Examination</p> <p>&lt;LinkToMethod&gt; link of the related the specific technical examination file including all related information about application, results, etc. with emphasis on</p> <p>&lt;Map&gt; the map of spots or map of areas that were examined during the specific examination</p> <p>&lt;File&gt; the image of the examined object where the examined areas or spots are annotated</p> <p>&lt;Note&gt; any comment related to the map of spots/areas</p> <p>&lt;MethodResult&gt; Contains information about the outcomes of examinations on the different spots</p> <p>&lt;Method 2&gt; the name of the Technical Examination .....</p>	<p>E7 Activity :P33 used specific technique: E29 Design or Procedure : P69 has association with: E29 Design or Procedure</p> <p>E29 Design or Procedure: P67B is referred to by: E73 Information Object</p> <p>E29 Design or Procedure: P3 has note: E62 String</p>
---	---

Known conservation history / archives / oral communication: Research in digital resources, archives and even oral communication with experts concerning information of past analytical data or conservation interventions.

(Possible source of error: possible interference of (aged) conservation materials)

Metadata field/term	CIDOC CRM related terms
<p>&lt;Interventions&gt; It refers to the interventions performed on the object</p> <p>&lt; Type&gt; contains the type of the intervention</p> <p>&lt;Date&gt; contains the date or period that the intervention happened</p> <p>&lt;PersonInvolved&gt; or &lt;OrganisationInvolved&gt; the person that participated in the intervention</p> <p>&lt;Conservation Report&gt; contains information about any detailed intervention on the Object</p>	<p>E11 Modification - P31 has modified: E22 Physical Man-Made Object</p> <p>E11 Modification: P2 has type: E55 Type "previous"</p> <p>E11 Modification: P4 has time-span: E52 Timespan</p> <p>E11 Modification: P14 carried out by: E39 Actor</p> <p>E11 Modification: P67B is referred to by:E73 Information Object</p> <p>E5 (Past) Event / E7 Activity/E11 Modification: P126 employed: E57 Material</p>
<p>&lt;Method&gt; is a link to a Glossary Entity describing the Method that is documented</p> <p>&lt;IdentityOfMethod&gt;: contains basic information about the method</p>	<p>S21 Measurement - P2 has type - E55 Type (Method Type as in situ / non-invasive)</p>



### Step 3 - Describe the method - Use (scope)

#### Method Identity and Scope

Scientific analysis? Archaeological excavation? conservation in course? Demonstration? Promotion? Service?

e.g. pigments identification to be integrated in the overall material characterisation of the artwork

Metadata field/term	CIDOC CRM related terms
<Method> is a link to a Glossary Entity describing the Method that is documented <IdentityOfMethod>: contains basic information about the method	S21 Measurement - P2 has type - E55 Type (Method Type as in situ / non-invasive)
<LinkToProject> information about the project in which the object is investigated (Additional metadata fields for the code may be included as: <Project Type>: the type of the Project <Project Goal> : the goal of the project )	E7 Activity: P2 has type: E55 Type E7 Activity = S21 Measurement - P21 had general purpose (was purpose of) - E55 Type

Analysis expected outcome Information to be obtained /analysis outcome - research expected result: Identification of inorganic or organic compounds (mainly pigments, dyes, fillers) through the fingerprint vibrational spectra of the molecular structure (amorphous, crystalline and polymorph particles are detectable).

Metadata field/term	CIDOC CRM related terms
<MethodReasoning>: the reason why the specific Method was chosen to examine the Object	S21 Measurement: P20 had specific purpose: E13 Attribute assignment or E14 Condition Assessment S21 Measurement: O24 measured: S15 Observable entity S21 Measurement: P140 assigned attribute to: E57 material

Skills requirements (level of training of the operator): Scientist specialised in Raman Spectroscopy

Metadata field/term	CIDOC CRM related terms
<Involved PersonSkills> / <Involved PersonProfessionalProfile>	S21 Measurement: P14 carried out by (performed): E39 Actor (E21 Person): P2 has type E55 Type (Profession Type)
<Involved Person> contains all information about the Person involved (<Appellation>; <Affiliation>, <Role>; <ContactInfo>)	E21 Person: P1 is identified by: E41 Appellation E21 Person: P3 has note: E62 String E21 Person: P2 has type: E55 Type E21 Person: P1 is identified by: E51 Contact Point
<Starting date> the day that the examination started, <Ending date> the day that the examination ended	S21 Measurement: P4 has time-span: E52 Timespan
<Location> is the Location where the examination took place, in situ or ex situ	S21 Measurement: P7 took place at: E53 Place



#### ***Step 4 - Describe the method. Operational environmental and safety conditions for the measurement***

Assure preferably low level of illuminance or dark conditions to avoid any light interference. Assure stability of the device set up and the artwork under investigation - avoid vibrations. The analytical protocol does not involve any risk for the safety of the operators or the public.

Metadata field/term	CIDOC CRM related terms
<MethodApplicationConditions> information about the environmental and operational conditions requirements for the application of the method. <Note> Any information about the conditions in which the measurements took place	S21 Measurement: P3 has note: E62 String

### Step 5 - Describe the method - Equipment specifications

The equipment specifications of the instrument and the overall apparatus are documented such as type of equipment, manufacturing company, laser excitation wavelength, optical system (camera) /objective lens used and any other component used.

Metadata field/term	CIDOC CRM related terms
<p>&lt;Device&gt;: information on the device and the overall apparatus</p> <p>&lt;DetectorType&gt;; type of detector. (e.g. CCD (charge coupled device), InGaAs (indium gallium arsenide) or other (specify))</p> <p>&lt;SpectrometerClass&gt;: e.g. dispersive (spectrograph with gratings/prisms) or FT (Fourier Transform) within the interferometer.</p> <p>&lt;Grating type&gt;: type of grating, such as holographically recorded diffraction grating (HRDG), classically ruled mechanical grating (CRMG);</p> <p>&lt;Grating density&gt;:line density of grating in dispersive Raman instrument. In format: lines per mm (lines/mm).</p> <p>&lt;SpectrometerManufacturer&gt;;&lt;Model&gt;: the name of spectrometer manufacturer and model used to collect spectral data</p> <p>&lt;Description&gt;: information describing the main characteristics of the device and the overall apparatus</p> <p>&lt;InstrumentOperationalSoftware;version;release&gt;: name, version and release of the Instrument software used</p>	<p>E7 Activity = E16 Measurement - P16 used specific object</p> <p>- E24 Physical man made thing: P2 has type: E55 Type (Instrument type)</p> <p>S21 Measurement: P33 used specific technique:</p> <p>E29 Design or procedure: P68 foresees use of : E55 Type: P2B is type of: E24 Physical man made thing has type E55 Type</p> <p>E29 Design or procedure: P3 has note: E62 String</p> <p>S21 Measurement: P16 used specific object: D14 Software</p> <p>D14 Software: P1 is identified by: E41 Appellation</p> <p>D14 Software: P1 is identified by: E42 Identifier</p>



### **Step 6- Describe the method. Measurement range, spot size and spatial resolution/ accuracy**

The measurement distance from the object is short, however, a long working distance objective is preferable for in situ measurements to avoid any contact with the artwork surface. The measurement spot size (dimensions of the analysed spot) and the measurement resolution/accuracy depends on the magnification of the objective used.

Metadata field/term	CIDOC CRM related terms
<p>&lt;MethodApplication&gt; information about the mode of application of the method, as distance from the object, spot size</p> <p>&lt;Confocal&gt;:use of confocal optical arrangement. In format: Y or N.</p> <p>&lt;Working distance&gt;: distance between objective lens and sample given in microns or mm.</p> <p>&lt;Spot size&gt;: laser spot diameter (nm) as determined by laser wavelength and microscope objective.</p> <p>&lt;MethodDescription&gt;: information on the method (spectral range, spatial resolution and accuracy)</p>	<p>S21 Measurement: P33 used specific technique: E29 Design or Procedure</p>



**Step 7 - Describe the measurement. Pre-processing/ safe limits**

Protocol for safe application (for the object damage protection). Particular attention is put to control the Laser intensity during measurement to avoid any risk of photochemical damage or local overheating of the irradiated area.

Metadata field/term	CIDOC CRM related terms
<MethodTesting>; <DeviceTesting>; <SpectrometerTesting>; <SetupTesting> Contains information on standard procedures on testing the setup preliminary to the measurements on the object; <Calibration> standards used to calibrate system (wavelength/intensity) <SampleCalibrationStandard> Contains information on the object or material used for the spectrometer testing	E29 Design or procedure: P33 was used by: S21 Measurement S9 Property Type: O9 property type was observed by S4 Observation S7 Simulation: P141 assigned: E54 dimension: P2 has type: E55 Type S7 Simulation: O23 defines: E29 Design or procedure
<PreprocessingSafeProtocol> contains information about the protocol applied to respect safe limits in the following measurements	D1 Digital Object: P70 documents: E73 Information Object



## Step 8 - Describe the measurement. Pre-processing/selection and annotation of Measurement Sites on reference images

Select the sites of analysis.

Use high resolution detailed reference images of the entire object or of the part under study (in case of a monument) to annotate the points of analysis.

Metadata field/term	CIDOC CRM related terms
<p>&lt;SpotInformation&gt; the information about the spot that is going to be examined</p> <p>&lt;PlacementwithintheObject&gt; defines the exact location of the Measurement Site on the Object.</p> <p>&lt;SpotPhoto&gt; a characteristic macro-photograph of the detail in which the spot is annotated or a micro-photograph of the examined area</p> <p>&lt;SurfaceDescription&gt; a description of the spot before starting the examination,</p> <p>&lt;PositionDescription&gt; a description of the position of the spot within the object</p> <p>&lt;Color&gt; the main colour of the spot that is identified with naked eye before the examination</p>	<p>E47 Spatial coordinates / E53 Place: P1 is identified by: - E42 Identifier (numerical Value or Code)</p> <p>E15 Identifier Assignment P37 assigned: E42 Identifier</p> <p>E53 Place/E26 Physical Feature: P3 has note: E62 String</p> <p>E26 Physical Feature: P2 has type: E55 Type(color)</p>
<p>&lt;Representation&gt; representation of the measurement Site in a photograph or drawing or</p> <p>&lt;Map&gt; the map of spots or map of areas that were examined during the specific examination</p>	<p>D30 Annotation Event: L48 created annotation: D29 Annotation Object</p>

<p>&lt;File&gt; the image of the examined object where the examined areas or spots are annotated</p> <p>&lt;Note&gt; any comment related to the map of spots/areas</p>	
--	--



### ***Step 9 - Describe the measurement. Pre- processing/preliminary tests - Measurement setup***

Preliminary tests operated on the specimen / area under examination using a minimum laser power. The laser power can be progressively increased as well as the integration time / average of scans can be optimised as needed to obtain a good spectrum.

<b>Metadata field/term</b>	<b>CIDOC CRM related terms</b>
<Preprocessing> Contains information about the preparation of the measurement	S21 Measurement - P20 had specific purpose (was purpose of) E5 Event

### Step 10 - Describe the measurement. Measurement / Spectra acquisition / Storage of data and metadata files

Acquire and store the Raman spectra taken on representative sites on the CH object. A preliminary evaluation of the spectrum is done (assessment of Raman peaks or fluorescence emission)

Metadata field/term	CIDOC CRM related terms
<p>&lt;MeasurementEvent&gt; Contains information about the measurement</p> <p>&lt;MeasurementDate&gt;: information about the date the measurement took place</p> <p>&lt;SpotAnalysis&gt; Contains all information about the examination of a specific spot</p>	<p>S15 Observable Entity - O24 was measured by - S21 Measurement</p> <p>S21 Measurement: P4 has time-span: E52 Timespan</p>

The conditions for each measurement constitute essential metadata which are documented either in a text file for all the measurements or in form of metadata file for each spectrum file. Laser excitation wavelength; laser power at the sample; spectral range (cut off); spectral resolution; integration time; number of scans.

Metadata field/term	CIDOC CRM related terms
<p>&lt;Settings&gt; Contains all information about the settings of the device that is used for the measurement of the specific spot</p>	



<p>(related metadata:</p> <p>&lt;LaserExcitationWavelength&gt;: excitation source (laser line) wavelength in nanometers (nm)</p> <p>&lt;LaserPowerAtTheSample&gt;: power incident on sample (not of source) in milliwatts (mW)</p> <p>&lt;Filter&gt;: optical filters used to prevent stray light from reaching spectrometer/detector, such as edge (long or short pass) or notch. (Example 1: holographic notch; Example 2: dielectric edge)</p> <p>&lt;Lens&gt; - &lt;ObjectiveMagnification&gt;: collecting lens magnification (Example: 50x)</p> <p>&lt;Lens&gt;-&lt;Numerical aperture&gt;: numerical aperture of collecting lens (Example: 0.75)</p> <p>&lt;Accessories&gt; any accessories used</p> <p>&lt;SpectralRange&gt;: data collection range in wavenumbers (cm<sup>-1</sup>)</p> <p>&lt;SpectralResolution&gt;: Resolution of spectrum as measured in wavenumbers</p> <p>&lt;Accumulations&gt;: number of accumulations used (co-added) to produce spectrum.</p> <p>&lt;Data collection type&gt;: Raman acquisition mode, static or scanned</p> <p>&lt;IntegrationTime&gt;: integration (dwell) time in seconds (sec)</p> <p>&lt;Cut-off frequency&gt;: low frequency cut-off in cm<sup>-1</sup> (1/cm)</p>	<p>E7 Activity: P70B is documented in: D1 Digital Object</p> <p>D10 Software Execution: L13 used parameters: D1 Digital Object</p> <p>D11 Digital Measurement Event: P2 has type: E55 Type (Raman spectrum acquisition)</p> <p>D11 Digital Measurement Event: L17 measured thing of type E55 Type (Raman spectrum)</p>
--	--

<p>&lt;Digital Object Title&gt; Contains the title of the digital object</p> <p>&lt;Digital Object Type&gt; Contains the type of the digital object, e.g. Raman Spectrum</p> <p>&lt;Digital Rights&gt;/&lt;Owner&gt; Contains the digital rights of the digital object</p>	<p>D1 Digital Object: P70 documents: E73 Information Object</p> <p>D1 Digital Object: P2 has type:E55 Type (type of file)</p> <p>D1 Digital Object: P1 is identified by: E41 Appellation</p> <p>D1 Digital Object: P3 has note: E62 String</p>
<p>&lt;Digital File&gt; or &lt;SpectrumFile&gt; Contains the digital file of the digital object</p> <p>    &lt;SpectrumFileType&gt;: Native /Raw/Proprietary or non-Proprietary?</p> <p>    &lt;Digital FileStandard&gt; Contains the type of standard Format of the digital file (e.g. spc; txt...)</p> <p>    &lt;Storage Location&gt; Contains the storage location of the digital object</p>	<p>D9 Data Object: L20 was created by: D11 Digital Measurement Event: L23 used software or firmware: D14 Software</p>



**Step 11 - Post process and interpret the data. 1: Stray light subtraction/ Baseline correction.**

Straylight correction (dark subtraction) Baseline correction - Subtraction of the fluorescence background from the spectra.

Metadata field/term	CIDOC CRM related terms
<Postprocessing> contains information about the postprocessing procedure (e.g. dark subtraction; baseline correction; cosmic ray removal:filtering of spectral artifacts (spikes/peaks) in Raman spectrum produced by cosmic rays rather than Raman emissions) <OtherDataPostProcessing>	E29 Design or procedure



**Step 12 - Post process and interpret data. Peak / spectral bands assessment.**

In this step the Raman shift wavenumber and the relative intensity of the observed peaks in the Raman spectrum are identified and documented. The relative intensity can be expressed as weak, moderate or strong.

Metadata field/term	CIDOC CRM related terms
<Postprocessing> contains information about the findings of the measurement based on the evaluation of the data	S6 Data Evaluation: O11 described: S15 Observable entity S15 Observable entity: O11 was described by: S6 Data evaluation



### ***Step 13 - Post process and interpret the data. Pigment/material identification***

The accurate identification of pigments requires the comparison of the Raman peaks with those reported in databases of Raman spectra of interest for cultural heritage studies. Several databases are available that cover a wide range of materials, from pigments and minerals to organic media and plastics.

<b>Metadata field/term</b>	<b>CIDOC CRM related terms</b>
<Postprocessing> contains information about the findings of the measurement based on the evaluation of the data	S6 Data Evaluation: O11 described: S15 Observable entity S15 Observable entity: O11 was described by: S6 Data evaluation

### Step 14 - Disseminate the results. Data and metadata storage, publication and sharing

Organize data and metadata in standard format for open access repositories. The metadata should incorporate descriptions of the equipment, instrumental setup and measurement conditions and information related to the artwork/monument. Finally, the (meta)data should follow the FAIR principles.

Metadata field/term	CIDOC CRM related terms
<p>&lt;Digital File&gt; or &lt;SpectrumFile&gt; Contains the digital file of the digital object</p> <p>&lt;SpectrumFileType&gt;: Native /Raw/Proprietary or non-Proprietary?</p> <p>&lt;Digital FileStandard&gt; Contains the type of standard Format of the digital file (e.g. spc; txt...)</p> <p>&lt;Storage Location&gt; Contains the storage location of the digital object</p>	<p>E73 Information object / E31 Document / E 36 Visual item</p> <p>E72 Legal Object: P105 right held by: E21 Person</p> <p>D1 Digital Object: L19 is stored on: D13 Digital Information Carrier</p>



### Step 15 - Disseminate the results. Prepare report

Prepare a document that includes the objectives of the project/study, information on the artwork, technical /experimental information, selection of representative data obtained, results and conclusions on the overall work.

Metadata field/term	CIDOC CRM related terms
<ExportReport> Offers to the editor the possibility to export at the local file system a report of the analytical process. The export file can be in PDF format. Reports include data from a number of predefined fields. In order to perform this action the editor should select the desired record from a list and then click to the action “Export report”. The editor can export a full report which means that data from all predefined fields of the record will be saved into a specific structure in the PDF file or a customized report by selecting which fields he/she wants to be included in the report.	E31 Document: P70 - documents: E7 Activity E 31 Document S5 Inference Making S4 Observation / S21 Measurement / E13 Attribute Assignment: O8 observed: S15 Observable Entity O9 observed property type: S9 Property Type O16 observed value: E1 CRM Entity I1 Argumentation: J2 concluded that I2 Belief E39 Actor (E21 Person OR E74 Group): P75 possesses: E30 Right



## Annex 2. Creation of a standards and methods Knowledge base

This Knowledge base represents the standards and methods used in the different SSK scenarios. They also serve as a keywords to describe scenarios, steps and resources, in order to help the SSK user to find the most accurate content.

<b>Standard name (Acronym)</b>	<b>The Text Encoding Initiative (TEI)</b>
<b>Standard description</b>	The Text Encoding Initiative (TEI) is a text-centric community of practice in the academic field of digital humanities, operating continuously since the 1980s. The community currently runs a mailing list, meetings and conference series, and maintains an eponymous technical standard, a journal, a wiki, a GitHub repository and a toolchain.
<b>Link to specification</b>	<a href="http://www.tei-c.org/index.xml">www.tei-c.org/index.xml</a>
<b>Standard name (Acronym)</b>	<b>Open Archival Information System (OAIS)</b>
<b>Standard description</b>	An Open Archival Information System (or OAIS) is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. The term OAIS also refers, by extension, to the ISO OAIS Reference Model for an OAIS. This reference model is defined by recommendation CCSDS 650.0-B-2 of the Consultative Committee for Space Data Systems; this text is identical to ISO 14721:2012. The CCSDS's purview is space agencies, but the OAIS model it developed has proved useful to a wide variety of other organizations and institutions with digital archiving needs.
<b>Link to specification</b>	<a href="http://www.iso.org/standard/57284.html">www.iso.org/standard/57284.html</a>



<b>Standard name (Acronym)</b>	<b>Encoded Archival Context (EAC)</b>
<b>Standard description</b>	Encoded Archival Context - Corporate bodies, Persons and Families (EAC-CPF) is an XML standard for encoding information about the creators of archival materials -- i.e., a corporate body, person or family -- including their relationships to (a) resources (books, collections, papers, etc.) and (b) other corporate bodies, persons and families. The goal is to provide contextual information regarding the circumstances of record creation and use. EAC-CPF can be used in conjunction with Encoded Archival Description (EAD) for enhancement of EAD's capabilities in encoding finding aids, but can also be used in conjunction with other standards or for standalone authority file encoding.
<b>Link to specification</b>	<a href="http://eac.staatsbibliothek-berlin.de/">eac.staatsbibliothek-berlin.de/</a>
<b>Standard name (Acronym)</b>	<b>Encoded Archival Guide (EAG)</b>
<b>Standard description</b>	Encoded Archival Guide (EAG) is intended to provide a standard for the description of archives. The Spanish Ministry of Culture has developed EAG 2002 for the Spanish-language archive portal "Censo-Guía de Archivos de España e Iberoamérica", together with the corresponding XML structure, to encode information about archives, such as address, opening hours, technical equipment, etc. A corresponding ICA standard was published in 2008 as the International Standard for Describing Institutions with Archival Holdings (ISDIAH).
<b>Link to specification</b>	<a href="http://www.bundesarchiv.de/archivgut_online/standards_werkzeug/e/02255/index.html">http://www.bundesarchiv.de/archivgut_online/standards_werkzeug/e/02255/index.html</a>
<b>Standard name (Acronym)</b>	<b>CIDOC Conceptual Reference Model (CIDOC_CRM)</b>
<b>Standard description</b>	The CIDOC Conceptual Reference Model (CRM) provides an extensible ontology for concepts and information in cultural heritage and museum documentation. It is the international standard (ISO 21127:2014) for the controlled exchange of cultural heritage information. Galleries, libraries, archives, museums (GLAMs), and other cultural institutions are encouraged to use the CIDOC CRM to enhance accessibility to museum-related information and knowledge.
<b>Link to specification</b>	<a href="http://www.cidoc-crm.org/">www.cidoc-crm.org/</a>



<b>Standard name (Acronym)</b>	<b>Component MetaData Infrastructure (CMDI)</b>
<b>Standard description</b>	Component MetaData Infrastructure(CMDI)describes a model that enables the flexible construction of interoperable metadata schemas for Language Resources (LRs). The metadata schemas based on this model can be used to describe resources at different levels of granularity (e.g. descriptions both on the collection level and on the level of individual resources).
<b>Link to specification</b>	<a href="http://www.clarin.eu/content/component-metadata">www.clarin.eu/content/component-metadata</a>
<b>Standard name (Acronym)</b>	<b>Analyzed Layout and Text Object (ALTO-XML)</b>
<b>Standard description</b>	ALTO (Analyzed Layout and Text Object) is an open XML Schema developed by the EU-funded project called METAE. The standard was initially developed for the description of text OCR and layout information of pages for digitized material. The goal was to describe the layout and text in a form to be able to reconstruct the original appearance based on the digitized information - similar to the approach of a lossless image saving operation. ALTO is often used in combination with Metadata Encoding and Transmission Standard (METS) for the description of the whole digitized object and creation of references across the ALTO files, e.g. reading sequence description.
<b>Link to specification</b>	<a href="http://www.loc.gov/standards/alto/">www.loc.gov/standards/alto/</a>
<b>Standard name (Acronym)</b>	<b>Music Encoding Initiative (MEI)</b>
<b>Standard description</b>	The Music Encoding Initiative (MEI) is an open-source effort to create a system for representation musical documents in a machine- readable structure. MEI closely mirrors work done by text scholars in the Text Encoding Initiative (TEI) and while the two encoding initiatives are not formally related, they share many common characteristics and development practices. The term "MEI", like "TEI", describes the governing organization and the mark up language.
<b>Link to specification</b>	<a href="http://music-encoding.org/">music-encoding.org/</a>



<b>Standard name (Acronym)</b>	<b>Morpho-syntactic annotation framework (MAF)</b>
<b>Standard description</b>	Morpho-syntactic annotation framework(MAF)provides a framework for the representation of annotations of word-forms in texts; such annotations concern tokens, their relationship with lexical units, and their morpho-syntactic properties.
<b>Link to specification</b>	<a href="http://www.iso.org/standard/51934.html">www.iso.org/standard/51934.html</a>
<b>Standard name (Acronym)</b>	<b>Extensible Markup Language (XML)</b>
<b>Standard description</b>	Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human- readable and machine- readable.
<b>Link to specification</b>	<a href="http://www.w3.org/standards/xml/core">www.w3.org/standards/xml/core</a>
<b>Standard name (Acronym)</b>	<b>Encoded Archival Description (EAD)</b>
<b>Standard description</b>	EAD stands for Encoded Archival Description, and is a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment. Finding aids are inventories, indexes, or guides that are created by archival and manuscript repositories to provide information about specific collections. While the finding aids may vary somewhat in style, their common purpose is to provide detailed description of the content and intellectual organization of collections of archival materials. EAD allows the standardization of collection information in finding aids within and across repositories.
<b>Link to specification</b>	<a href="http://www.loc.gov/ead/">www.loc.gov/ead/</a>
<b>Standard name (Acronym)</b>	<b>Historical International Standard of Classification of Occupations (HISCO)</b>
<b>Standard description</b>	The Historical International Standard of Classification of Occupations or HISCO is a theoretical model used to code social class and occupational status. Formulated in 2002, the model complements the ILO's ISCO68 scheme, as it prescribes a universal code system for examining occupation descriptions.
<b>Link to specification</b>	<a href="http://socialhistory.org/en/projects/hisco-history-work">socialhistory.org/en/projects/hisco-history-work</a>





<b>Standard name (Acronym)</b>	<b>Lightweight Information Describing Objects (LIDO)</b>
<b>Standard description</b>	LIDO (Lightweight Information Describing Objects) is an XML schema for describing museum objects. It is applicable to all kinds of objects: art, natural history, technology, cultural objects. LIDO is widely used for harvesting museum data in union catalogues, and as a preliminary step for converting to semantic formats like CIDOC CRM or EDM.
<b>Link to specification</b>	<a href="http://www.cidoc-crm.org/Resources/the-lido-model">www.cidoc-crm.org/Resources/the-lido-model</a>
<b>Standard name (Acronym)</b>	<b>Metadata Encoding and Transmission Standard (METS)</b>
<b>Standard description</b>	The Metadata Encoding and Transmission Standard (METS) is a metadata standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium (W3C).
<b>Link to specification</b>	<a href="http://www.loc.gov/standards/mets/">www.loc.gov/standards/mets/</a>
<b>Standard name (Acronym)</b>	<b>Metadata Object Description Schema (MODS)</b>
<b>Standard description</b>	The Metadata Object Description Schema (MODS) is an XML-based bibliographic description schema developed by the United States Library of Congress' Network Development and Standards Office. MODS was designed as a compromise between the complexity of the MARC format used by libraries and the extreme simplicity of Dublin Core metadata.
<b>Link to specification</b>	<a href="http://www.loc.gov/standards/mods/">www.loc.gov/standards/mods/</a>
<b>Standard name (Acronym)</b>	<b>Data Documentation Initiative (DDI)</b>
<b>Standard description</b>	The Data Documentation Initiative (also known as DDI or DDI Metadata) is an International standard for describing surveys, questionnaires, statistical data files, and social sciences study-level information. This information is described as metadata by the standard.
<b>Link to specification</b>	<a href="http://www.ddialliance.org/">www.ddialliance.org/</a>



<b>Standard name (Acronym)</b>	<b>Dublin Core Schema (DC)</b>
<b>Standard description</b>	The Dublin Core Schema is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks. The full set of Dublin Core metadata terms can be found on the Dublin Core Metadata Initiative (DCMI) website.
<b>Link to specification</b>	<a href="http://dublincore.org/">dublincore.org/</a>
<b>Standard name (Acronym)</b>	<b>Schematron (Schematron)</b>
<b>Standard description</b>	Schematron is a rule-based validation language for making assertions about the presence or absence of patterns in XML trees. It is a structural schema language expressed in XML using a small number of elements and XPath.
<b>Link to specification</b>	<a href="http://schematron.com/">schematron.com/</a>
<b>Standard name (Acronym)</b>	<b>Europeana Data Model (EDM)</b>
<b>Standard description</b>	The Europeana Data Model (EDM) is a richer metadata standard introduced to help give users more and better information.
<b>Link to specification</b>	<a href="http://pro.europeana.eu/page/edm-documentation">pro.europeana.eu/page/edm-documentation</a>
<b>Standard name (Acronym)</b>	<b>Data Model to Europeana (DM2E)</b>
<b>Standard description</b>	DM2E (Data Model to Europeana) is building the tools and communities to enable humanities researchers to work with manuscripts in the Linked Open Web.
<b>Link to specification</b>	<a href="http://dm2e.eu/">dm2e.eu/</a>
<b>Standard name (Acronym)</b>	<b>CENDARI Archival Extension</b>
<b>Standard description</b>	The standard installation of AtoM is used in conjunction with a CENDARI theme. In addition, CENDARI developed a Shibboleth authentication extension and a script to upload AtoM archival descriptions in EAD and EAG to the CENDARI repository.
<b>Link to specification</b>	<a href="http://docs.cendari.dariah.eu/developer/atom.html">docs.cendari.dariah.eu/developer/atom.html</a>



<b>Standard name (Acronym)</b>	<b>MeDieVal Extension of DM2E (MDV)</b>
<b>Standard description</b>	MDV is MeDieVal Extension of DM2E model developed by SISMELE for the study of medieval manuscripts.
<b>Link to specification</b>	<a href="http://www.digitalhumanist.net/wiki/images/9/9c/7_09_2015_ONTOLOGI_A.pdf">www.digitalhumanist.net/wiki/images/9/9c/7_09_2015_ONTOLOGI A.pdf</a>
<b>Standard name (Acronym)</b>	<b>Friend of a friend (FOAF)</b>
<b>Standard description</b>	FOAF (an acronym of friend of a friend) is a machine-readable ontology describing persons, their activities and their relations to other people and objects. Anyone can use FOAF to describe themselves. FOAF allows groups of people to describe social networks without the need for a centralised database. FOAF is a descriptive vocabulary expressed using the Resource Description Framework (RDF) and the Web Ontology Language (OWL).
<b>Link to specification</b>	<a href="http://www.foaf-project.org/">www.foaf-project.org/</a>
<b>Standard name (Acronym)</b>	<b>Gemeinsame Normdatei (GND)</b>
<b>Standard description</b>	The Integrated Authority File (German: Gemeinsame Normdatei, also known as: Universal Authority File) or GND is an international authority file for the organisation of personal names, subject headings and corporate bodies from catalogues. It is used mainly for documentation in libraries and increasingly also by archives and museums.
<b>Link to specification</b>	<a href="http://www.dnb.de/EN/Standardisierung/GND/gnd.html">www.dnb.de/EN/Standardisierung/GND/gnd.html</a>
<b>Standard name (Acronym)</b>	<b>Virtual International Authority File (VIAF)</b>
<b>Standard description</b>	The Virtual International Authority File (VIAF) is an international authority file. It is a joint project of several national libraries and operated by the Online Computer Library Center (OCLC). The aim is to link the national authority files (such as the German Name Authority File) to a single virtual authority file. In this file, identical records from the different data sets are linked together.
<b>Link to specification</b>	<a href="http://viaf.org">viaf.org</a>



<b>Standard name (Acronym)</b>	<b>Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)</b>
<b>Standard description</b>	The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol developed for harvesting (or collecting) metadata descriptions of records in an archive so that services can be built using metadata from many archives. An implementation of OAI-PMH must support representing metadata in Dublin Core, but may also support additional representations.
<b>Link to specification</b>	<a href="http://www.openarchives.org/pmh/">www.openarchives.org/pmh/</a>
<b>Standard name (Acronym)</b>	<b>Text Corpus Format (TCF)</b>
<b>Standard description</b>	The TCF format (Text Corpus Format) is used by WebLicht services as a machine-readable format for representing and exchanging linguistically annotated texts. It enables interoperability of linguistic tools. WebLicht developer manual provides detailed information on the TCF format and its background.
<b>Link to specification</b>	<a href="http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format">weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format</a>
<b>Standard name (Acronym)</b>	<b>Linguistic annotation framework (LAF)</b>
<b>Standard description</b>	LAF, linguistic annotation framework, is a standard ISO for representing linguistic annotations of language data such as corpora, speech signal and video. The framework includes an abstract data model and an XML serialization of that model for representing annotations of primary data. The serialization serves as a pivot format to allow annotations expressed in one representation format to be mapped onto another.
<b>Link to specification</b>	<a href="http://www.iso.org/standard/37326.html">www.iso.org/standard/37326.html</a>
<b>Standard name (Acronym)</b>	<b>International Standard Classification of Occupations (ISCO)</b>
<b>Standard description</b>	The International Standard Classification of Occupations (ISCO) is an International Labour Organization (ILO) classification structure for organizing information on labour and jobs.
<b>Link to specification</b>	<a href="http://www.ilo.org/public/english/bureau/stat/isco/">http://www.ilo.org/public/english/bureau/stat/isco/</a>



<b>Standard name (Acronym)</b>	<b>Laser-induced breakdown spectroscopy (LIBS)</b>
<b>Standard description</b>	Laser Induced Breakdown Spectroscopy (LIBS) is an analytical technique that enables the determination of the elemental composition of materials on the basis of the characteristic atomic emission from a micro-plasma produced by focusing a high-power laser on or in a material.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Raman spectroscopy (Raman spectroscopy)</b>
<b>Standard description</b>	Raman spectroscopy is a spectroscopic technique used to observe vibrational, rotational, and other low-frequency modes in a system. Raman spectroscopy is commonly used in chemistry to provide a structural fingerprint by which molecules can be identified.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Laser-induced fluorescence (LIF)</b>
<b>Standard description</b>	Laser-induced fluorescence (LIF) is a spectroscopic method in which an atom or molecule is excited to a higher energy level by the absorption of laser light followed by spontaneous emission of light. LIF is used for studying structure of molecules, detection of selective species and flow visualization and measurements. The excitation wavelength is often selected to be the one at which the species has its largest cross section. The excited species decay usually after a few nanoseconds to microseconds, and emit light at a wavelength longer than the excitation wavelength.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>X-ray fluorescence spectroscopy (XRF)</b>
<b>Standard description</b>	X-ray fluorescence is the emission of characteristic "secondary" (or fluorescent) X-rays from a material that has been excited by bombarding with high-energy X-rays or gamma rays. The phenomenon is widely used for elemental analysis and chemical analysis, particularly in the investigation of metals, glass, ceramics and building materials, and for research in geochemistry, forensic science, archaeology and art objects such as paintings and murals.
<b>Link to specification</b>	



<b>Standard name (Acronym)</b>	<b>Multispectral Imaging (Multispectral Imaging)</b>
<b>Standard description</b>	Multispectral imaging, like other spectral imaging, collects and processes information from across the electromagnetic spectrum. The goal of Multispectral imaging is to obtain the spectrum for each pixel in the image of a scene, with the purpose of finding objects, identifying materials, or detecting processes.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Electronic still-picture imaging -- Removable memory -- Part 2: TIFF/EP image data format (ISO 12234-2:2001)</b>
<b>Standard description</b>	Tag Image File Format/Electronic Photography (TIFF/EP) is a digital image file format standard – ISO 12234-2, titled "Electronic still-picture imaging – Removable memory – Part 2: TIFF/EP image data format". The TIFF/EP standard is based on a subset of the Adobe TIFF standard, and a subset of the JEITA Exif standard, with some differences and extensions.
<b>Link to specification</b>	<a href="https://www.iso.org/standard/29377.html">https://www.iso.org/standard/29377.html</a>
<b>Standard name (Acronym)</b>	<b>Conservation of cultural heritage - Cleaning of porous inorganic materials - Laser cleaning techniques for cultural heritage (CEN/TS 16163:2014)</b>
<b>Standard description</b>	This European standard applies to porous inorganic materials constituting cultural heritage. It provides the fundamental requirements of the laser parameters and guidelines for the choice of the laser operational parameters, in order to optimize the cleaning procedure.
<b>Link to specification</b>	<a href="https://standards.cen.eu/dyn/www/f?p=204:110:0:::FSP_PROJECT:38496&amp;cs=15154AA9F085EB7344C3DFD9FA87693B0">https://standards.cen.eu/dyn/www/f?p=204:110:0:::FSP_PROJECT:38496&amp;cs=15154AA9F085EB7344C3DFD9FA87693B0</a>
<b>Standard name (Acronym)</b>	<b>Optics and photonics -- Lasers and laser-related equipment -- Test methods for laser beam power, energy and temporal characteristics (ISO 11554:2017)</b>
<b>Standard description</b>	ISO 11554:2017 specifies test methods for determining the power and energy of continuous wave and pulsed laser beams, as well as their temporal characteristics of pulse shape, pulse duration and pulse repetition rate. Test and evaluation methods are also given for the power stability of cw-lasers, energy stability of pulsed lasers and pulse duration stability.
<b>Link to specification</b>	<a href="https://www.iso.org/standard/69232.html">https://www.iso.org/standard/69232.html</a>



<b>Standard name (Acronym)</b>	<b>Lasers and laser-related equipment -- Test methods for laser-induced damage threshold -- Part 4: Inspection, detection and measurement (ISO/TR 21254-4:2011)</b>
<b>Standard description</b>	ISO/TR 21254-4:2011 describes selected techniques for the inspection of optical surfaces prior to and after damage testing, and damage detection techniques integrated in detection facilities. The described damage detection methods are examples of practical solutions tested and often applied in detection facilities. Also, this direct information on the state of damage can be processed in the course of the running test to determine energy levels for the following interrogations optimised to minimise detection uncertainties. This inspection of the interrogated sites is essential for an accurate determination of the damage thresholds because it is the final and most sensitive assessment of the state of damage.
<b>Link to specification</b>	<a href="https://www.iso.org/standard/69232.html">https://www.iso.org/standard/69232.html</a>
<b>Standard name (Acronym)</b>	<b>AutoCAD Drawing Exchange Format (AutoCAD DXF)</b>
<b>Standard description</b>	AutoCAD DXF (Drawing Interchange Format, or Drawing Exchange Format) is a CAD data file format developed by Autodesk for enabling data interoperability between AutoCAD and other programs. DXF was originally introduced in December 1982 as part of AutoCAD 1.0, and was intended to provide an exact representation of the data in the AutoCAD native file format, DWG (Drawing), for which Autodesk for many years did not publish specifications. Because of this, correct imports of DXF files have been difficult. Autodesk now publishes the DXF specifications as a PDF on its website. Versions of AutoCAD from Release 10 (October 1988) and up support both ASCII and binary forms of DXF. Earlier versions support only ASCII. As AutoCAD has become more powerful, supporting more complex object types, DXF has become less useful. Certain object types, including ACIS solids and regions, are not documented. Other object types, including AutoCAD 2006's dynamic blocks, and all of the objects specific to the vertical market versions of AutoCAD, are partially documented, but not well enough to allow other developers to support them. For these reasons many CAD applications use the DWG format which can be licensed from Autodesk or non-natively from the Open Design Alliance. DXF coordinates are always without dimensions so that the reader or user needs to know the drawing unit or has to extract it from the textual comments in the sheets.
<b>Link to specification</b>	





<b>Standard name (Acronym)</b>	<b>Joint Photographic Experts Group (JPEG)</b>
<b>Standard description</b>	JPEG ( JAY-peg) is a commonly used method of lossy compression for digital images, particularly for those images produced by digital photography. The degree of compression can be adjusted, allowing a selectable tradeoff between storage size and image quality. JPEG typically achieves 10:1 compression with little perceptible loss in image quality. JPEG compression is used in a number of image file formats. JPEG/Exif is the most common image format used by digital cameras and other photographic image capture devices; along with JPEG/JFIF, it is the most common format for storing and transmitting photographic images on the World Wide Web. These format variations are often not distinguished, and are simply called JPEG. The term "JPEG" is an initialism/acronym for the Joint Photographic Experts Group, which created the standard. The MIME media type for JPEG is image/jpeg, except in older Internet Explorer versions, which provides a MIME type of image/pjpeg when uploading JPEG images. JPEG files usually have a filename extension of .jpg or .jpeg. JPEG/JFIF supports a maximum image size of 65,535×65,535 pixels, hence up to 4 gigapixels for an aspect ratio of 1:1.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>stereolithography (STL (file format))</b>
<b>Standard description</b>	STL (an abbreviation of "stereolithography") is a file format native to the stereolithography CAD software created by 3D Systems. STL has several after-the-fact backronyms such as "Standard Triangle Language" and "Standard Tessellation Language". This file format is supported by many other software packages; it is widely used for rapid prototyping, 3D printing and computer-aided manufacturing. STL files describe only the surface geometry of a three-dimensional object without any representation of colour, texture or other common CAD model attributes. The STL format specifies both ASCII and binary representations. Binary files are more common, since they are more compact. An STL file describes a raw, unstructured triangulated surface by the unit normal and vertices (ordered by the right-hand rule) of the triangles using a three-dimensional Cartesian coordinate system. In the original specification, all STL coordinates were required to be positive numbers, but this restriction is no longer enforced and negative coordinates are commonly encountered in STL files today. STL files contain no scale information, and the units are arbitrary.





<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Portable Network Graphics (PNG)</b>
<b>Standard description</b>	<p>Portable Network Graphics (PNG, pronounced PEE-en-JEE or PING) is a raster graphics file format that supports lossless data compression. PNG was created as an improved, non-patented replacement for Graphics Interchange Format (GIF), and is the most widely used lossless image compression format on the Internet. PNG supports palette-based images (with palettes of 24-bit RGB or 32-bit RGBA colours), grayscale images (with or without alpha channel for transparency), and full-colour non-palette-based RGB/RGBA images (with or without alpha channel). PNG was designed for transferring images on the Internet, not for professional-quality print graphics, and therefore does not support non-RGB colour spaces such as CMYK. A PNG file contains a single image in an extensible structure of "chunks", encoding the basic pixels and other information such as textual comments and integrity checks documented in RFC 2083. PNG files nearly always use file extension PNG or png and are assigned MIME media type image/png. PNG was published as informational RFC 2083 in March 1997 and as an ISO/IEC standard in 2004.</p>
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>3D studio (.3ds)</b>
<b>Standard description</b>	<p>3DS is one of the file formats used by the Autodesk 3ds Max 3D modelling, animation and rendering software. It was the native file format of the old Autodesk 3D Studio DOS (releases 1 to 4), which was popular until its successor (3D Studio MAX 1.0) replaced it in April 1996. Having been around since 1990 (when the first version of 3D Studio DOS was launched), it has grown to become a de facto industry standard for transferring models between 3D programs, or for storing models for 3D resource catalogues (along with OBJ, which is more frequently used as a model archiving file format). While the 3DS format aims to provide an import/export format, retaining only essential geometry, texture and lighting data, the related MAX format (now superseded by the PRJ format) also contains extra information specific to Autodesk 3ds Max, to allow a scene to be completely saved/loaded.</p>
<b>Link to specification</b>	



<b>Standard name (Acronym)</b>	<b>Tagged Image File Format (TIFF)</b>
<b>Standard description</b>	Tagged Image File Format, abbreviated TIFF or TIF, is a computer file format for storing raster graphics images, popular among graphic artists, the publishing industry, and photographers. TIFF is widely supported by scanning, faxing, word processing, optical character recognition, image manipulation, desktop publishing, and page-layout applications. The format was created by Aldus Corporation for use in desktop publishing. It published the latest version 6.0 in 1992, subsequently updated with an Adobe Systems copyright after the latter acquired Aldus in 1994. Several Aldus or Adobe technical notes have been published with minor extensions to the format, and several specifications have been based on TIFF 6.0, including TIFF/EP (ISO 12234-2), TIFF/IT (ISO 12639), TIFF-F (RFC 2306) and TIFF-FX (RFC 3949).
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Wavefront .obj file (OBJ)</b>
<b>Standard description</b>	OBJ (or .OBJ) is a geometry definition file format first developed by Wavefront Technologies for its Advanced Visualizer animation package. The file format is open and has been adopted by other 3D graphics application vendors. The OBJ file format is a simple data-format that represents 3D geometry alone — namely, the position of each vertex, the UV position of each texture coordinate vertex, vertex normals, and the faces that make each polygon defined as a list of vertices, and texture vertices. Vertices are stored in a counter-clockwise order by default, making explicit declaration of face normals unnecessary. OBJ coordinates have no units, but OBJ files can contain scale information in a human readable comment line.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>bitmap image file (BMP)</b>
<b>Standard description</b>	The BMP file format, also known as bitmap image file or device independent bitmap (DIB) file format or simply a bitmap, is a raster graphics image file format used to store bitmap digital images, independently of the display device (such as a graphics adapter), especially on Microsoft Windows and OS/2 operating systems. The BMP file format is capable of storing two-dimensional digital images both monochrome and color, in various color depths, and optionally with data compression, alpha channels, and color profiles. The Windows Metafile (WMF) specification covers the



	BMP file format. Among others, wingdi.h defines BMP constants and structures.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Polygon File Format (PLY)</b>
<b>Standard description</b>	PLY is a computer file format known as the Polygon File Format or the Stanford Triangle Format. It was principally designed to store three-dimensional data from 3D scanners. The data storage format supports a relatively simple description of a single object as a list of nominally flat polygons. A variety of properties can be stored, including: colour and transparency, surface normals, texture coordinates and data confidence values. The format permits one to have different properties for the front and back of a polygon. There are two versions of the file format, one in ASCII, the other in binary.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Collaborative Design Activity (COLLADA)</b>
<b>Standard description</b>	COLLADA (COLLABorative Design Activity) is an interchange file format for interactive 3D applications. It is managed by the non-profit technology consortium, the Khronos Group, and has been adopted by ISO as a publicly available specification, ISO/PAS 17506. COLLADA defines an open standard XML schema for exchanging digital assets among various graphics software applications that might otherwise store their assets in incompatible file formats. COLLADA documents that describe digital assets are XML files, usually identified with a .dae (digital asset exchange) filename extension.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>Data Documentation Initiative (DDI)</b>
<b>Standard description</b>	The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioural, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding, interpretation, and use -- by people, software systems, and computer networks.



<b>Link to specification</b>	<a href="https://www.ddialliance.org/">https://www.ddialliance.org/</a>
<b>Standard name (Acronym)</b>	<b>Cascading Style Sheets (CSS)</b>
<b>Standard description</b>	Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript. CSS is designed to enable the separation of presentation and content, including layout, colours, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural content.
<b>Link to specification</b>	<a href="https://www.w3.org/Style/CSS/">https://www.w3.org/Style/CSS/</a>
<b>Standard name (Acronym)</b>	<b>XSL Formatting Objects (XSL-FO)</b>
<b>Standard description</b>	XSL-FO (XSL Formatting Objects) is a markup language for XML document formatting that is most often used to generate PDF files. XSL-FO is part of XSL (Extensible Stylesheet Language), a set of W3C technologies designed for the transformation and formatting of XML data. The other parts of XSL are XSLT and XPath. Version 1.1 of XSL-FO was published in 2006.
<b>Link to specification</b>	<a href="https://www.w3.org/Style/XSL/">https://www.w3.org/Style/XSL/</a>
<b>Standard name (Acronym)</b>	<b>Extensible Stylesheet Language (XSL)</b>
<b>Standard description</b>	XSL is a language for expressing style sheets. An XSL style sheet is, like with CSS, a file that describes how to display an XML document of a given type. XSL shares the functionality and is compatible with CSS2 (although it uses a different syntax). It also adds a transformation language for XML documents: XSLT. Originally intended to perform complex styling operations, like the generation of tables of contents and indexes, it is now used as a general purpose XML processing language. XSLT is thus widely used for purposes other than XSL, like generating HTML web pages from XML data.
<b>Link to specification</b>	<a href="https://www.w3.org/Style/XSL/">https://www.w3.org/Style/XSL/</a>



<b>Standard name (Acronym)</b>	<b>Apache Formatting Objects Processor (Apache FOP)</b>
<b>Standard description</b>	Apache™ FOP (Formatting Objects Processor) is a print formatter driven by XSL formatting objects (XSL-FO) and an output independent formatter. It is a Java application that reads a formatting object (FO) tree and renders the resulting pages to a specified output. Output formats currently supported include PDF, PS, PCL, AFP, XML (area tree representation), Print, AWT and PNG, and to a lesser extent, RTF and TXT. The primary output target is PDF
<b>Link to specification</b>	<a href="https://xmlgraphics.apache.org/fop/">https://xmlgraphics.apache.org/fop/</a>
<b>Standard name (Acronym)</b>	<b>InDesign Markup Language (IDML)</b>
<b>Standard description</b>	IDML is an XML-based format for representing InDesign content. Essentially, it is a major revision of INX, InDesign's previous XML-based file format. IDML addresses INX shortcomings, in an effort to represent InDesign content in a human-readable format—something that can be reasonably assembled and disassembled by a competent XML programmer.
<b>Link to specification</b>	<a href="http://www.images.adobe.com/content/dam/acom/en/devnet/indesign/sdk/cs6/idml/idml-cookbook.pdf">http://www.images.adobe.com/content/dam/acom/en/devnet/indesign/sdk/cs6/idml/idml-cookbook.pdf</a>
<b>Standard name (Acronym)</b>	<b>Comma-separated values (CSV)</b>
<b>Standard description</b>	A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. A CSV file stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format. RFC 4180 formalized CSV. It defines the MIME type "text/csv", and CSV files that follow its rules should be very widely portable.
<b>Link to specification</b>	<a href="https://tools.ietf.org/html/rfc4180">https://tools.ietf.org/html/rfc4180</a>
<b>Standard name (Acronym)</b>	<b>American Standard Code for Information Interchange (ACSCII)</b>
<b>Standard description</b>	ASCII abbreviated from American Standard Code for Information Interchange, is a character encoding standard for electronic communication. ASCII codes represent text in computers, telecommunications equipment, and other devices. Most modern character-encoding schemes are based on ASCII, although they support many additional characters. ASCII is the traditional name for the encoding system.



<b>Link to specification</b>	<a href="https://www.iso.org/standard/4777.html">https://www.iso.org/standard/4777.html</a>
<b>Standard name (Acronym)</b>	<b>Thermo Scientific SPC file format (SPC)</b>
<b>Standard description</b>	The Thermo Scientific SPC file format is a generic format used in all of Thermo Scientific's software products. In addition, many spectrometer vendors' software packages offer capabilities to export spectral data to the Thermo Scientific SPC format or use it as their native file format. All the converters and translators used on this site create Thermo Scientific SPC files from the submitted spectral data files before presenting them to the library search engine on this site. Submitting data files in the Thermo Scientific SPC format is the most reliable way to guarantee that your data is interpreted correctly for searching against the databases. The Thermo Scientific SPC file format is completely public domain information.
<b>Link to specification</b>	
<b>Standard name (Acronym)</b>	<b>JCAMP-DX: A Standard Form for the Exchange of Infrared Spectra in Computer Readable Form (JCAMP-DX)</b>
<b>Standard description</b>	CAMP-DX is a standard file form for exchange of infrared spectra and related chemical and physical information between spectrometer data systems of different manufacture, main-frame time-sharing systems, general purpose lab computers, and personal computers. It is compatible with all media: telephone, magnetic and optical disk, magnetic tape, and even the printed page (via optical reader). All data are stored as labelled fields of variable length using printable ASCII characters. A JCAMP-DX spectrum is a text file which can be viewed, corrected, and annotated with a text editor. The present focus is on infrared spectra, but JCAMP-DX can easily accommodate Raman, UV, NMR, mass, and other types of spectra, x-ray powder patterns, chromatograms, thermograms, and other plots which require the capability of representing contours as well as peak position and intensity. JCAMP-DX also provides for combining adequate information about the sample and method of observation with its spectrum.
<b>Link to specification</b>	<a href="http://jcamp-dx.org">http://jcamp-dx.org</a>



<b>Standard name (Acronym)</b>	<b>Page Analysis and Ground-Truth Elements (PAGE XML)</b>
<b>Standard description</b>	There is a plethora of established and proposed document representation formats but none that can adequately support individual stages within an entire sequence of document image analysis methods (from document image enhancement to layout analysis to OCR) and their evaluation. This paper describes PAGE, a new XML-based page image representation framework that records information on image characteristics (image borders, geometric distortions and corresponding corrections, binarisation etc.) in addition to layout structure and page content. The suitability of the framework to the evaluation of entire workflows as well as individual stages has been extensively validated by using it in high-profile applications such as in public contemporary and historical ground-truthed datasets and in the ICDAR Page Segmentation competition series.
<b>Link to specification</b>	<a href="http://dx.doi.org/10.1109/ICPR.2010.72">http://dx.doi.org/10.1109/ICPR.2010.72</a>
<b>Standard name (Acronym)</b>	<b>hOCR (hOCR)</b>
<b>Standard description</b>	hOCR is an open standard for representing document layout analysis and OCR results as a subset of HTML. The goal is to reuse as much existing technology as possible, and to arrive at a representation that makes it easy to store, share, process and display OCR results. This specification defines many features that can represent a variety of OCR-related information. However, being built on top of HTML, hOCR is designed to make it easy to start simple and gradually use more complex constructs when necessary. Consider you have an HTML document that encodes a book: Wrapping page elements in <div class="ocr_page"> tags will convey the page boundaries to hOCR-capable agents and turn the HTML document into an hOCR document.
<b>Link to specification</b>	<a href="http://kba.cloud/hocr-spec/1.2/">http://kba.cloud/hocr-spec/1.2/</a>