PARCOURS project: **CRMcr** - a CIDOC-CRM extension for supporting semantic interoperability in the conservation and restoration domain

**Claudia MARINICA**

claudia.marinica@u-cergy.fr

*ETIS UMR 8051, Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS, C2RMF Centre de Recherche et de Restauration des Musées de France

**LRMH** Laboratoire de Recherche des Monuments Historiques

**CRCC** Centre de Recherche sur la Conservation des Collections (CRCC) USR3224 CRC DAVID, Université de Versailles Saint-Quentin-en-Yvelines

**IPANEMA**

**INP** Institut national du patrimoine

12-13 September 2019, London, UK

**Linked Conservation Data network: Modelling workshop**
Thank you!

Inès Bannour (post-doc)

Cheikh Niang (post-doc)
Table of contents

(1) Context and motivations
(2) Database: semantics and terminology
(3) An ontological framework for conservation
(4) CRMcr: the proposed ontological model
(5) Information system: data query, knowledge base, development
(6) Conclusion and perspectives
Detailed table of contents

(1) Context and motivations
   PARCOURS Project
   Motivations
   Objectives
   Questions

(2) Database: semantics and terminology

(3) An ontological framework for conservation
   Definitions
   Requirements of the cultural institutions
   Conceptual architecture

(4) CRMcr: the proposed ontological model
   Join institutions’ requirements
   Overview of the main concepts
   Identification and physical features
   Degrading events: alterations
   Non-degrading events: activities
   Analysis techniques

(5) Information system: data query, knowledge base, development

(6) Conclusion and perspectives
PARCOURS Project

- **PARCOURS**: Cultural Heritage and Conservation Restoration - an Ontology for the Use of a common Reference for different Data Sources
  
  http://projet-parcours.eu/

- Started in 2014 and supported by the Foundation of Heritage Sciences (PATRIMA Excellence Research Laboratory), France

- Partners:
  - French Museum’s Research and Restoration Center (C2RMF)
  - Research Laboratory for Historical Monuments (LRMH)
  - ETIS Research Lab, Cergy-Pontoise (Paris area), France
  - Research Center on the Conservation of Collections (CRCC)
  - DAVID Research Lab, Université de Versailles Saint-Quentin-en-Yvelines
  - Laboratoire IPANEMA
  - Institut national du patrimoine (INP)

- Aims at providing users from different CH Institutions with enhanced **unified global access to data** related to studies, analysis, techniques, and preservation practices in **conservation processes**
Motivations

- Diversity of the data generated in the **conservation** field
- Difficulty to put together the overall information concerning the artworks’ life cycle

⇒ Important to solve interoperability issues:

- **Distributed data:** isolated data due to restrictions
- **Databases heterogeneity:** relational, NoSql, bibliographic records...
- **Lack of semantic power:** data with implicit semantics
- **Terminology heterogeneity and ambiguity:** databases do not share the same meaning about the terms they use in conservation tasks
Motivation

► **C2RMF**: EROS: relational database

► **LRMH**: CASTOR: bibliographical NoSql database
PARCOURS’ objective

- **Goal:** Query at the same time multiple data sources
- **Constraints:**
  - have a minimum of structure and of semantics in the databases
  - have a unique conceptualization of the data for interoperability
  - have queries that can query at the same time
PARCOURS’ interoperability goals

▶ Provide a **common shared reference** for **distributed** and **heterogeneous** cultural heritage data
  ▶ to have a common understanding
  ▶ to have a common evaluation, etc.

(1) **A common conceptualization: a conceptual model**

▶ Provide a **common unified global access** to data sources, based on their **meta-data**:
  ▶ solve data incompleteness
  ▶ enhance data complementarity
  ▶ produce, diffuse, share and reuse efficiently data by stakeholders, etc.

(2) **An information system**
Questions to fulfill the objectives

1. Is the database **semantically** ready to integrate a unified query system?

2. Are there any terminological **heterogeneity** and **ambiguity** problems in the data?

3. What constraints and steps to produce the **shared conceptual model**?

4. How to query the **heterogeneous** databases by using the shared conceptual model?

5. How to query **distributed** databases and which architecture for the query system?
Detailed table of contents

(1) Context and motivations
PARCOURS Project
Motivations
Objectives
Questions

(2) Database: semantics and terminology

(3) An ontological framework for conservation
Definitions
Requirements of the cultural institutions
Conceptual architecture

(4) CRMcr: the proposed ontological model
Join institutions’ requirements
Overview of the main concepts
Identification and physical features
Degrading events: alterations
Non-degrading events: activities
Analysis techniques

(5) Information system: data query, knowledge base, development

(6) Conclusion and perspectives
Is the database **semantically** ready to integrate a unified query system?
Lack of semantics in the database

Bibliographical base CASTOR

Issue:

- The field *keywords* contains the knowledge concerning the conservation and restoration
- Impossible to know which keyword corresponds to: analyses, restorations, alterations, etc.

Goal:

- Class each keyword in a category: Restoration (*Rentoilage*), Analysis (*spectrometry, IR, microscopy, chimical analysis, MEB EDS*), Type(*Cathedral, Religious building, Church*)

Result:

- 7000/10,500 keywords classed in 12 categories
Are there any terminological **heterogeneity** and **ambiguity** problems in the data?
Terminological heterogeneity and ambiguity

Issues:

- **Synonym terms and abbreviations** of the terms: spectrometry/spectroscopy, etc.
- **Polysemic terms**: artificial aging (analysis technique) / natural aging (alteration of natural origin)
- **Rich vocabulary**: presence of generic terms and specific terms: MEB / MEB EDS, etc.

Objectives:

- Development of a set of thesauri
Thesauri: Common terminology
Reference software of the Ministry of Culture (France): **GINCO**

- Thesaurus of alterations
- Thesaurus of manufacturing techniques
- **Thesaurus of analysis and examination techniques**
- Thesaurus of intervention techniques
- Thesaurus of materials
Example: Thesaurus of analysis and examination techniques

- **Top level concepts**: Chromatography, Scientific imagery, Chemical measurement, Physical measurement, Spectrometry
What constraints and steps to produce the shared conceptual model?
Detailed table of contents

(1) Context and motivations
   PARCOURS Project
   Motivations
   Objectives
   Questions

(2) Database: semantics and terminology

(3) An ontological framework for conservation
   Definitions
   Requirements of the cultural institutions
   Conceptual architecture

(4) CRMcr: the proposed ontological model
   Join institutions’ requirements
   Overview of the main concepts
   Identification and physical features
   Degrading events: alterations
   Non-degrading events: activities
   Analysis techniques

(5) Information system: data query, knowledge base, development

(6) Conclusion and perspectives
Ontology:

- **Conceptualization**: produce an abstract model for the conservation field, by semantically classifying the domain concepts.

- **Formalization**: make this conceptualization understandable and exploitable by information systems.
Requirements of the cultural institutions

What kind of static information could describe the artwork and its identity?
- identifiers, author, typology, form, dimensions, manufacturing material, location, owner...

What kind of alterations can affect an artwork and what do you need to keep about them?
- Degradation, aging, attacks...
- Origin, Nature, Factors, Area...

What kind of scientific studies can be conducted on the artwork and what kind of data and results can be collected?
- Diagnosis, studies, sample taking, imagery...
- Measurements, spectrums, interpretations, images...

What kind of interventions can be undertaken on an artwork after possible analysis?
- restoration, preventive conservation, remedial Conservation...
CIDOC-CRM (Conceptual Reference Model): reference conceptual model for information related to cultural heritage [Class (E), Property (P), Blue]

CRMsci: extension of CIDOC-CRM for scientific observation [Class (S), Property (O), Green]

CRMcr: proposed extension of CIDOC CRM for the Conservation-Restoration of cultural objects [Class (C), Property (R), Orange]
Requirements’ validation
(1) Context and motivations
   PARCOURS Project
   Motivations
   Objectives
   Questions

(2) Database: semantics and terminology

(3) An ontological framework for conservation
   Definitions
   Requirements of the cultural institutions
   Conceptual architecture

(4) CRMcr: the proposed ontological model
   Join institutions’ requirements
   Overview of the main concepts
   Identification and physical features
   Degrading events: alterations
   Non-degrading events: activities
   Analysis techniques

(5) Information system: data query, knowledge base, development

(6) Conclusion and perspectives
Join institutions’ requirements

Objects identification

- S15 Observable Entity
- CIDOC-CRM

- P48 has preferred identifier
- P7 took place at
- Affected by
- CRMcr
- P14 carried out by
- P4 has time-span

Scientific studies

- CRMsci

- S4 Observation
- E16 Measurement
- S2 Sample Taking
- S21 Measurement
- 024 measured
- Non-invasive Scientific analysis: Imagery, spectroscopy...
- S11 Amount of Matter

- E38 Image
- Spectrum, curve...

Interventions

- CRM
- E7 Activity
- E73 Visual Item
- E57 Material
- E5 Event
- E2 Temporal Entity
- E52 Time-Span
- E53 Place
- E39 Actor
- P46 is composed of
- E73 Visual Item

- Images, spectra...
- Measurements

- Techniques
- Instruments
- E57 Material
- Measurements

Conservation, Restoration

- Reports, studies, diagnosis proposals...

Needed intervention

- Interventions
- Documentation

E16 Measurement

Followed

E5 Event

Initiated

E5 Event

Measuring instruments

- Techniques
- Instruments
- E57 Material
- Measurements
CRMcr: overview of the main concepts

Non-degrading events: Activities

Measuring Instruments

Degrading event: Alteration

C1 Cultural Object

C2 Immovable Cultural Object

C3 Movable Cultural Object

E5 Event

C8 Organization

E53 Location

C10 Procedure

C11 Technique

C12 Instrument

C16 Scientific Study

C17 Intervention

C39 Documentation

E7 Activity

S10 Material Substantial

R6 needed intervention

R15 affected area

R16 detected

R17 affected by

R18 was affected by

R9 required the use of

R10 intervened

P1 is identified by

P81 was witnessed

P52 has current owner

R1 belongs to

R6 used instrument

R7 used technique

S18 Alteration

E41 Appellation

E78 Collection

E4 Period

Identification

Measurement Instruments

CRMcr: overview of the main concepts

Non-degrading events: Activities

Measuring Instruments

Degrading event: Alteration

C1 Cultural Object

C2 Immovable Cultural Object

C3 Movable Cultural Object

E5 Event

C8 Organization

E53 Location

C10 Procedure

C11 Technique

C12 Instrument

C16 Scientific Study

C17 Intervention

C39 Documentation

E7 Activity

S10 Material Substantial

R6 needed intervention

R15 affected area

R16 detected

R17 affected by

R18 was affected by

R9 required the use of

R10 intervened

P1 is identified by

P81 was witnessed

P52 has current owner

R1 belongs to

R6 used instrument

R7 used technique

S18 Alteration

E41 Appellation

E78 Collection

E4 Period

Identification

Measurement Instruments

CRMcr: overview of the main concepts
Identification and physical features: Simple paths

Identification

P1: Is identified by

Physical features

C1 Cultural Object

E41 Appellation

E35 Title

E42 Identifier

C52 Inventory Number

« AM1988-882 »

« Adoration of shepherds »

E78 Collection

R1 Belongs To

E84 Type

C9 Shape

E57 Material

« Canvas »

R4 Has shape

P45 Consists of

R2 Falls within the domain of

« Painting »

http://data.c2mrff/oeuvre/F11878

E55 Type

E54 Dimension

http://../3220#height

http://../2370#width

E3 Condition State

C4 Conservation State

R23 Has conservation state

R5 Has stage

S16 State

R102 Has title

P8i Witnessed

E5 Material

P45 Consists of

R2 Has inventory number

P48 Has preferred identifier

C7 Stage

R1 Belongs To

E4 Period

P8i Witnessed

« Adoration of shepherds »

« AM1988-882 »
Identification and physical features: Complex paths

The author:
- **C1 Cultural Object** → **P108i was produced by** → **E12 Production**
- **E12 Production** → **P14 Carried Out By** → **E39 Actor**

The manufacturing technique:
- **C1 Cultural Object** → **P108i Was Produced By** → **E12 Production**
- **E12 Production** → **P32 Used General Technique** → **C7 Manufacturing Technique**
Degrading events: alterations

http://data.c2rmf.fr/oeuvre/F11788

«former restorations»

«aging»

E18 Physical Thing

C1 Cultural Object

018i was altered by

http://data.c2rmf.fr/alteration/F11788-d001

C14 Alteration Factor

C62 Area

C15 Altered Area

R14 was caused by

R15 has affected area

E11 Modification

S17 Physical Genesis

E7 Activity

C22 Intervention

E55 Type

C48 Altered Support

C49 Altered Surface

«rust and hole»

«yellowed varnish»

«aging»

«former restorations»
Non-degrading events: activities
Non-degrading events: activities

- **(1) C16 Scientific Study:** classification of analysis
  - **S4 Observation:** observing a cultural object
  - **E16 Measurement:** taking measurement by doing some experimental analysis
    - **Invasive analysis:** S3 Measurement By Sampling
    - **Non-invasive analysis:** C19 Scientific Imagery, C55 Spectrometry
    - **Chemical analysis:** C18 Chromatography
    - **Physical analysis:** C55 Spectrometry
Non-degrading events: activities’ results
Non-degrading events: activities’ results

(1) **C16 Scientific Study**: an example of analysis results
Non-degrading events: activities

(2) C22 Intervention: classification of interventions: International Council of Museums - Committee for Conservation
Analysis techniques

- E29 Design And Procedure
- E7 Activity
- C10 Procedure
- C12 Instrument
- C11 Technique
- C21 Intervention Technique
- C13 Analysis Technique
- C60 Chromatography Technique
- C20 Scientific Imagery Technique
- C17 Sample Taking Technique
- C61 Spectrometry Technique

Techniques:
- Scanning Electron Microscopy
- Infrared Phorography
- Monoethyl Ether
- Ethylene Glycol
- Varnish removal
- Stratigraphic section

Resources:
- http://data.c2mrff/technique/
- MB
- IR
- http://c2rmf.f/technique/10803

Requirements:
- R9 required the use of
- R7 used technique
- R8 supported by
- R10 followed

Materials:
- E57 Material

Notes:
- P3 has note on "Scanning Electron Microscopy" and "Infrared Phorography"
(1) Context and motivations
   PARCOURS Project
   Motivations
   Objectives
   Questions

(2) Database: semantics and terminology

(3) An ontological framework for conservation
   Definitions
   Requirements of the cultural institutions
   Conceptual architecture

(4) CRMcr: the proposed ontological model
   Join institutions’ requirements
   Overview of the main concepts
   Identification and physical features
   Degrading events: alterations
   Non-degrading events: activities
   Analysis techniques

(5) Information system: data query, knowledge base, development

(6) Conclusion and perspectives
How to query the **heterogeneous** databases by using the shared conceptual model?
Knowledge base development
How to query **distributed** databases and which architecture for the query system?
Information system development

- Translation of user queries from natural language to SPARQL
- Send the queries to knowledge bases (mediator)
- Send back the answers to the user
Information system
(1) Context and motivations
   PARCOURS Project
   Motivations
   Objectives
   Questions

(2) Database: semantics and terminology

(3) An ontological framework for conservation
   Definitions
   Requirements of the cultural institutions
   Conceptual architecture

(4) CRMcr: the proposed ontological model
   Join institutions’ requirements
   Overview of the main concepts
   Identification and physical features
   Degrading events: alterations
   Non-degrading events: activities
   Analysis techniques

(5) Information system: data query, knowledge base, development

(6) Conclusion and perspectives
Conclusion et perspectives

Conclusions:

- Semantization of CASTOR base
- Production and integration of thesauri
- Production, validation et documentation of the CRMcr ontology
- Mappings building
- Knowledge bases building
- Test and implementation of different reasoners
- Implementation of the information system
Conclusion et perspectives

Perspectives:

- Validate the CRMcr ontology by the CIDOC-CRM consortium
- Natural language queries processing
- Query optimization
- Development of a tool for online knowledge data production under the PARCOURS format (ongoing - we just got the funding!)
THANK YOU FOR YOUR ATTENTION!

http://projet-parcours.eu
Lack of semantics in the database

Bibliographical base CASTOR
Méthodologie - sémantisation :

- Mise en place d’un outil permettant d’annoter sémantiquement (affecter des classes) les mots-clés en se basant sur :
  - les thesauri du domaine culturel ou des branches de ces thesauri existants sur GINCO ou dans la base EROS
  - les catégories existantes dans la conceptualisation mises en place pour la restauration-conservation.

- Si un mot-clé est classé dans plusieurs classes, mise en place d’un système de classement (ranking) ou choix arbitraire de la ou des catégories correspondantes
Réalisation :

- Mots-clés à classer : ≈ 10 500 mots-clés
- Classes :
  - .TAE : Techniques d’analyse et d’examen
  - .AL : Altérations
  - .IN : Interventions
  - .MMO : Matière de mise en oeuvre
  - .TMO : Technique de mise en oeuvre
  - .DHA : Domaine ou Histoire de l’art
  - .MODEN : Dénomination des objets culturels mobiliers
  - .IMODEN : Dénomination des objets culturels immobiliers
  - .CEC : Composition en élément chimique (résultats d’analyses)
  - .ECON : État de conservation
  - .MI : Marque/inscription
  - .EAR : Élément architectural

- Mots-clés classés : ≈ 7000 mots
- Mots-clés non classés : ≈ 3500 mots
- Production d’une copie de la base de données avec les nouvelles classes créées pour éclater le champ *Mots-Clés* de CASTOR
Example: Thesaurus of analysis and examination techniques