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

Detailed table of contents

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

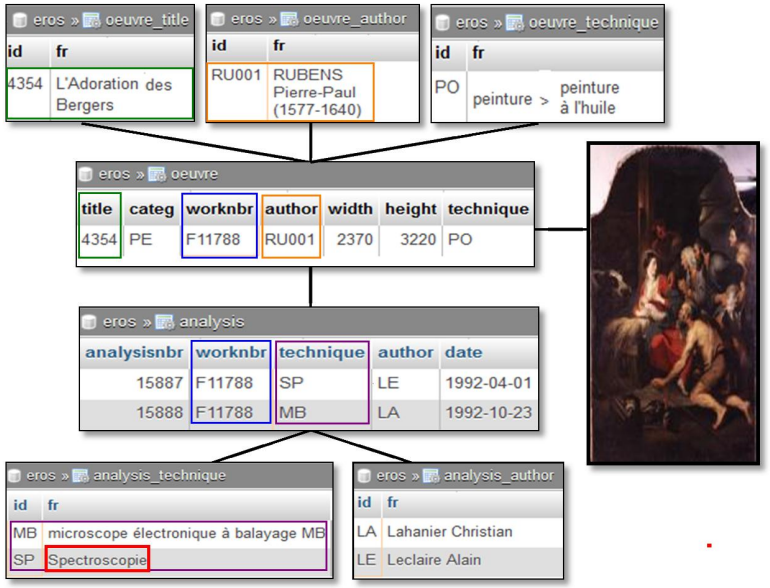
- Diversity of the data generated in the **conservation** field
- Difficulty to put together the overall information concerning the artworks' life cycle

⇒ Important to solve interoperability issues:

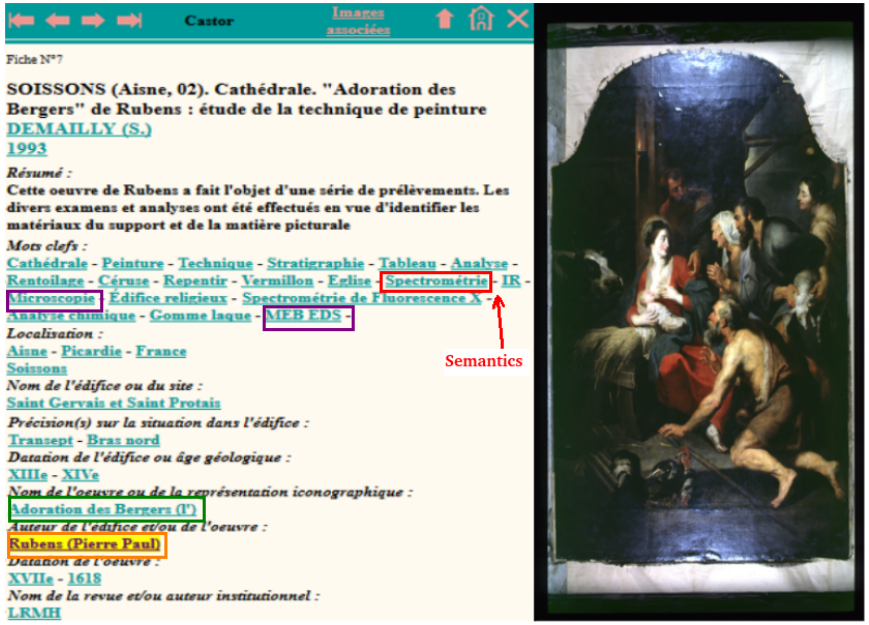
- **Distributed data:** isolated data due to restrictions
- **Databases heterogeneity:** relational, NoSql, bibliographic records...
- **Lack of semantic power:** data with implicit semantics
- **Terminology heterogeneity and ambiguity:** databases do not share the same meaning about the terms they use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

Motivation

- **C2RMF:** EROS: relational database

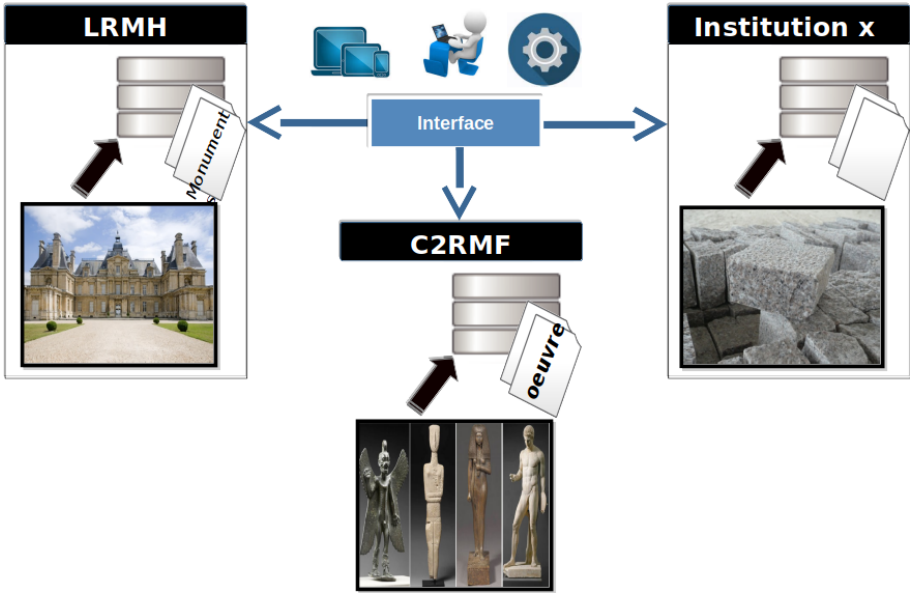


- **LRMH:** CASTOR: bibliographical NoSql database



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

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

Lack of semantics in the database

Bibliographical base CASTOR

Issue:

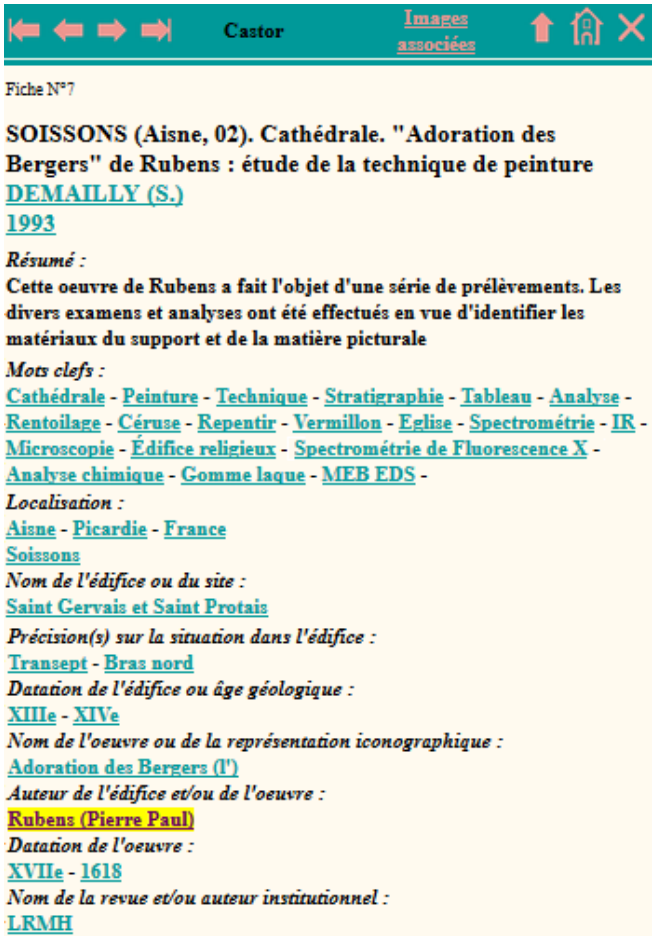
- The field *keywords* contains the knowledge concerning the conservation and restoration
- Impossible to know which keyword corresponds to: analyses, restorations, alterations, etc.

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



Castor | Images associées

Fiche N°7

SOISSONS (Aisne, 02). Cathédrale. "Adoration des Bergers" de Rubens : étude de la technique de peinture
DEMAILLY (S.)
1993

Résumé :
Cette oeuvre de Rubens a fait l'objet d'une série de prélèvements. Les divers examens et analyses ont été effectués en vue d'identifier les matériaux du support et de la matière picturale

Mots clefs :
Cathédrale - Peinture - Technique - Stratigraphie - Tableau - Analyse - Rentoilage - Céruse - Repentir - Vermillon - Eglise - Spectrométrie - IR - Microscopie - Édifice religieux - Spectrométrie de Fluorescence X - Analyse chimique - Gomme laque - MEB EDS -

Localisation :
Aisne - Picardie - France
Soissons

Nom de l'édifice ou du site :
Saint Gervais et Saint Protais

Précision(s) sur la situation dans l'édifice :
Transept - Bras nord

Datation de l'édifice ou âge géologique :
XIIIe - XIVe

Nom de l'oeuvre ou de la représentation iconographique :
Adoration des Bergers (l')

Auteur de l'édifice et/ou de l'oeuvre :
Rubens (Pierre Paul)

Datation de l'oeuvre :
XVIIe - 1618

Nom de la revue et/ou auteur institutionnel :
LRMH

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

Terminological heterogeneity and ambiguity

Issues:

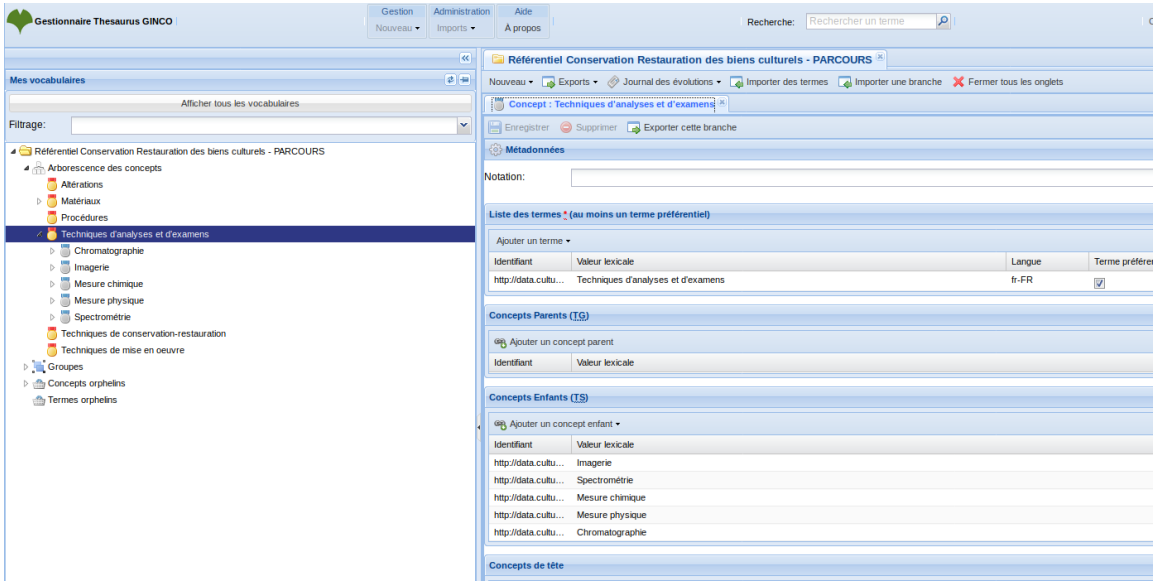
- Synonym terms and abbreviations of the terms: spectrometry/spectroscopy, etc.
- Polysemic terms: artificial aging (analysis technique) / natural aging (alteration of natural origin)
- Rich vocabulary: presence of generic terms and specific terms: MEB / MEB EDS, etc.

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

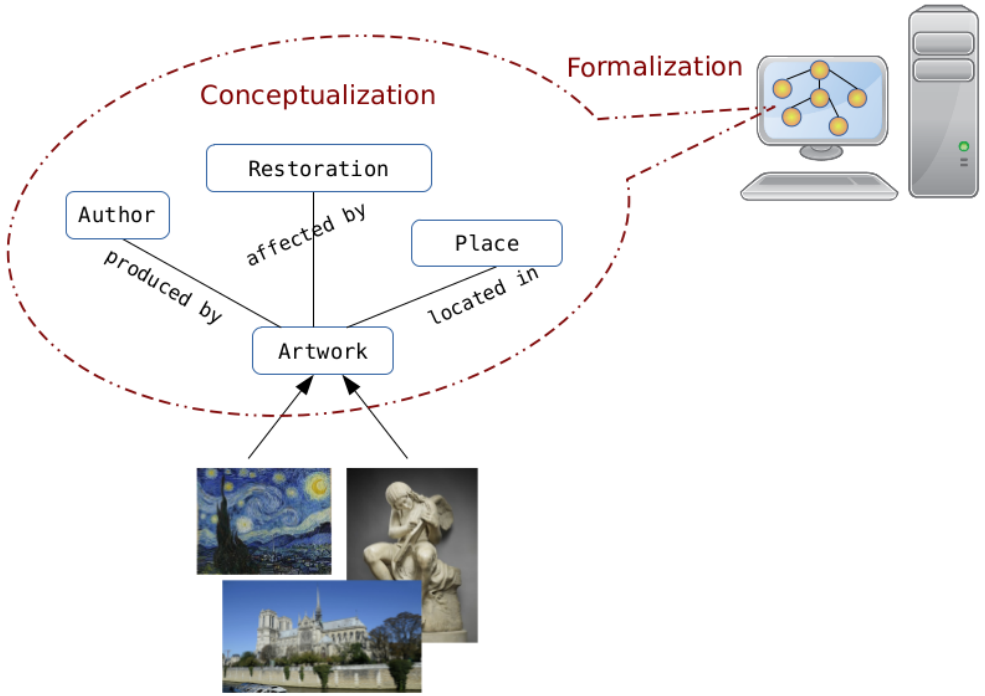
What constraints and steps to produce the **shared conceptual model**?

Detailed table of contents

Definitions

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



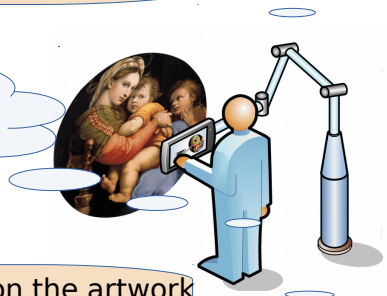
Requirements of the cultural institutions

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...*



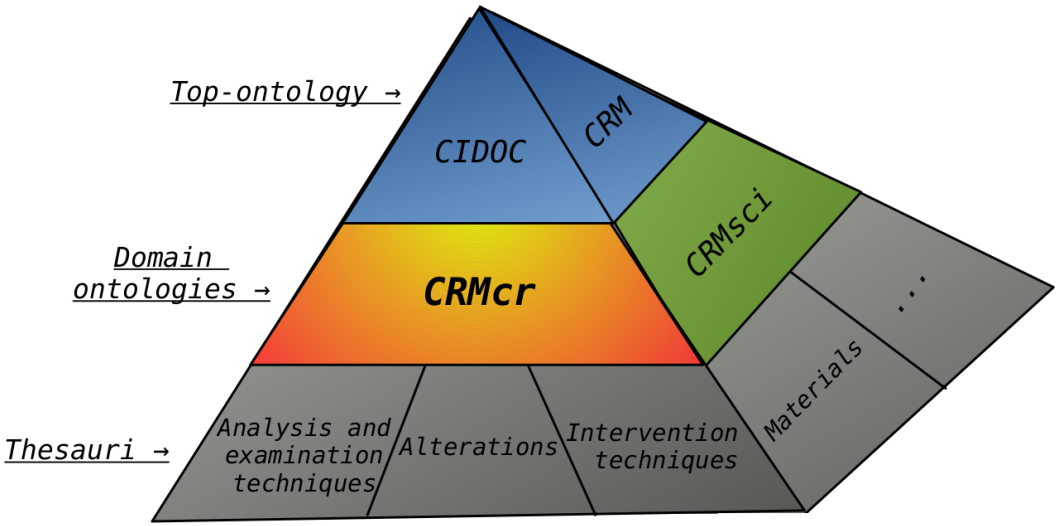
What kind **scientific studies** can be conducted on the artwork and what kind of **data** and **results** can be collected ?

**Diagnosis, studies, sample taking, imagery...*
**Mesaurements, spectrums, interpretations, images...*

What kind of **interventions** can be undertaken on an artwork after possible analysis

restoration, preventive conservation, remedial Conservation...

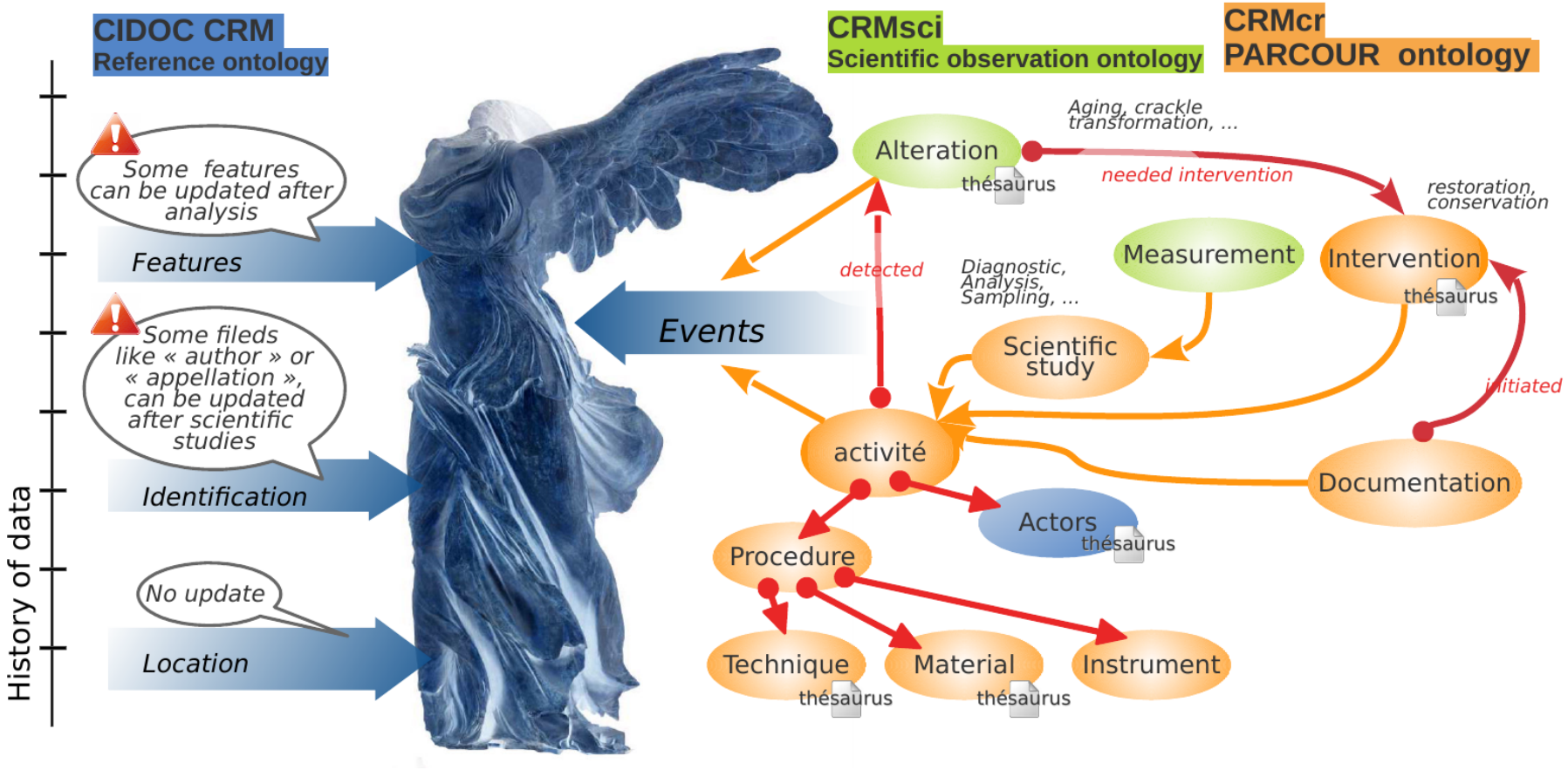
Conceptual architecture



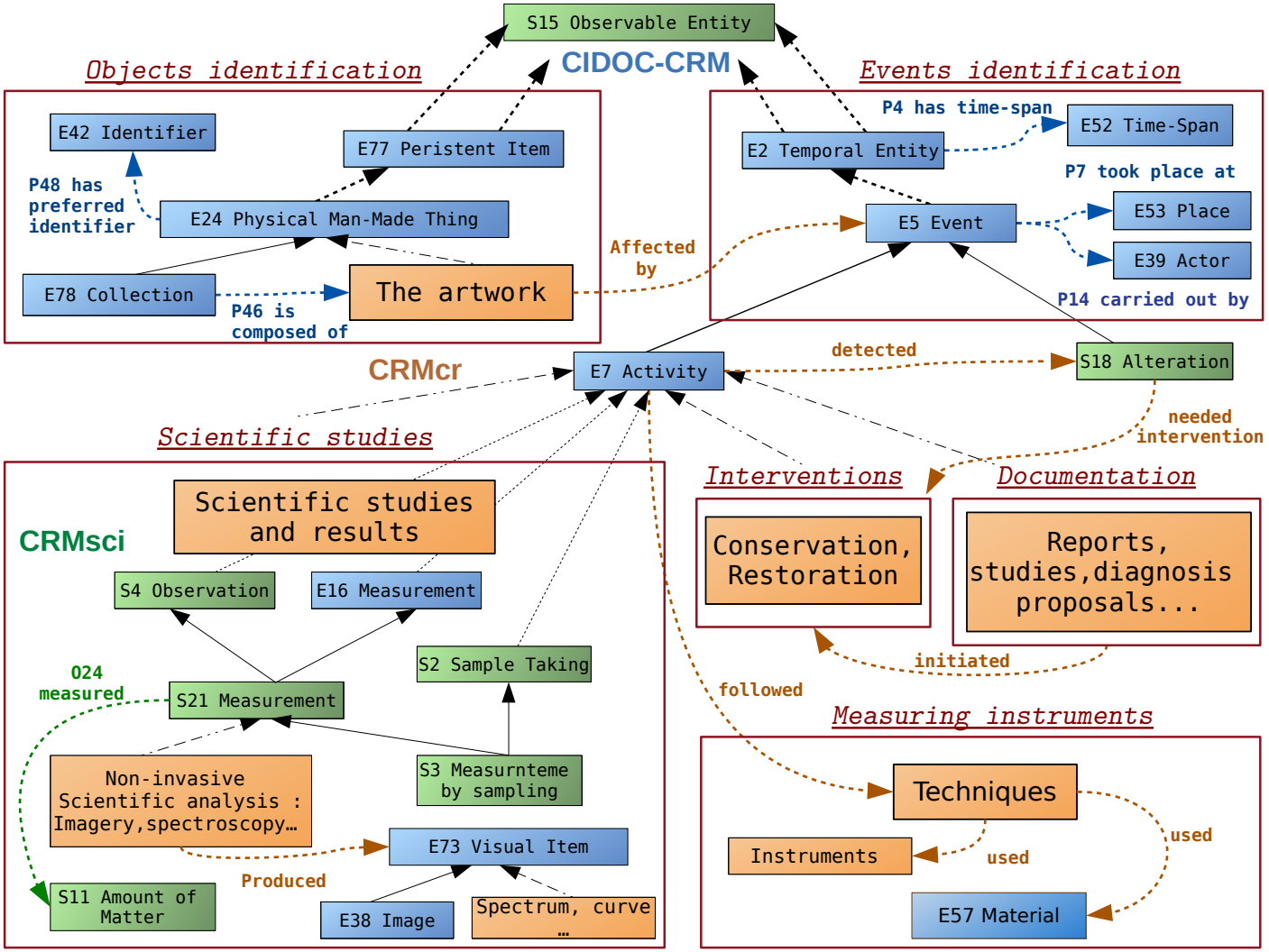
- **CIDOC-CRM (Conceptual Reference Model)**[1]**:** reference conceptual model for information related to cultural heritage [Class (**E**), Property (**P**), Blue]
- **CRMsci**[2]**:** extension of CIDOC-CRM for scientific observation [Class (**S**), Property (**O**), Green]
- **CRMcr:** proposed extension of CIDOC **CRM** for the **C**onservation-**R**estoration of cultural objects [Class (**C**), Property (**R**), Orange]

[1] http://www.cidoc-crm.org/
[2] http://www.cidoc-crm.org/crmsci/

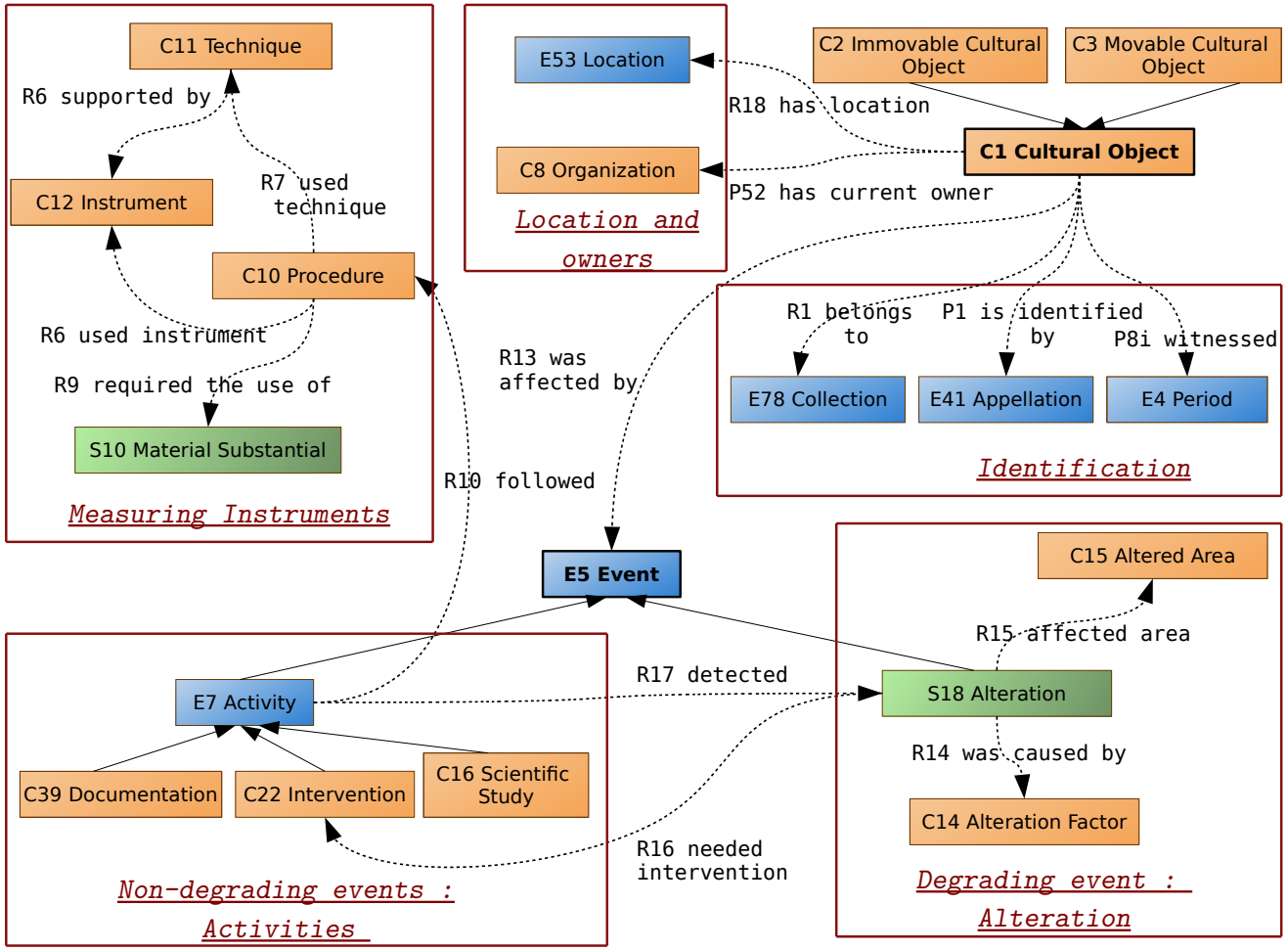
Requirements' validation



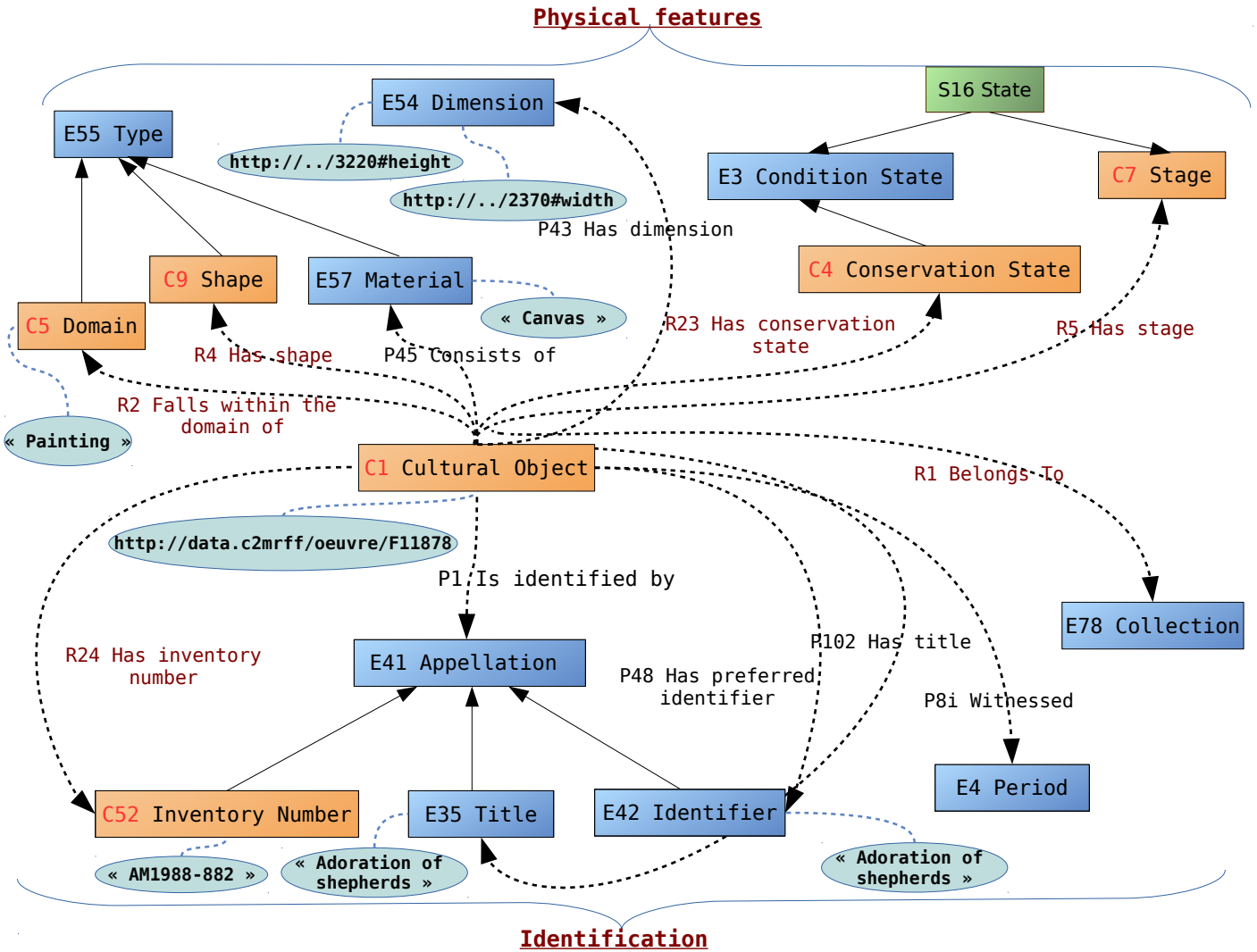
Join institutions' requirements



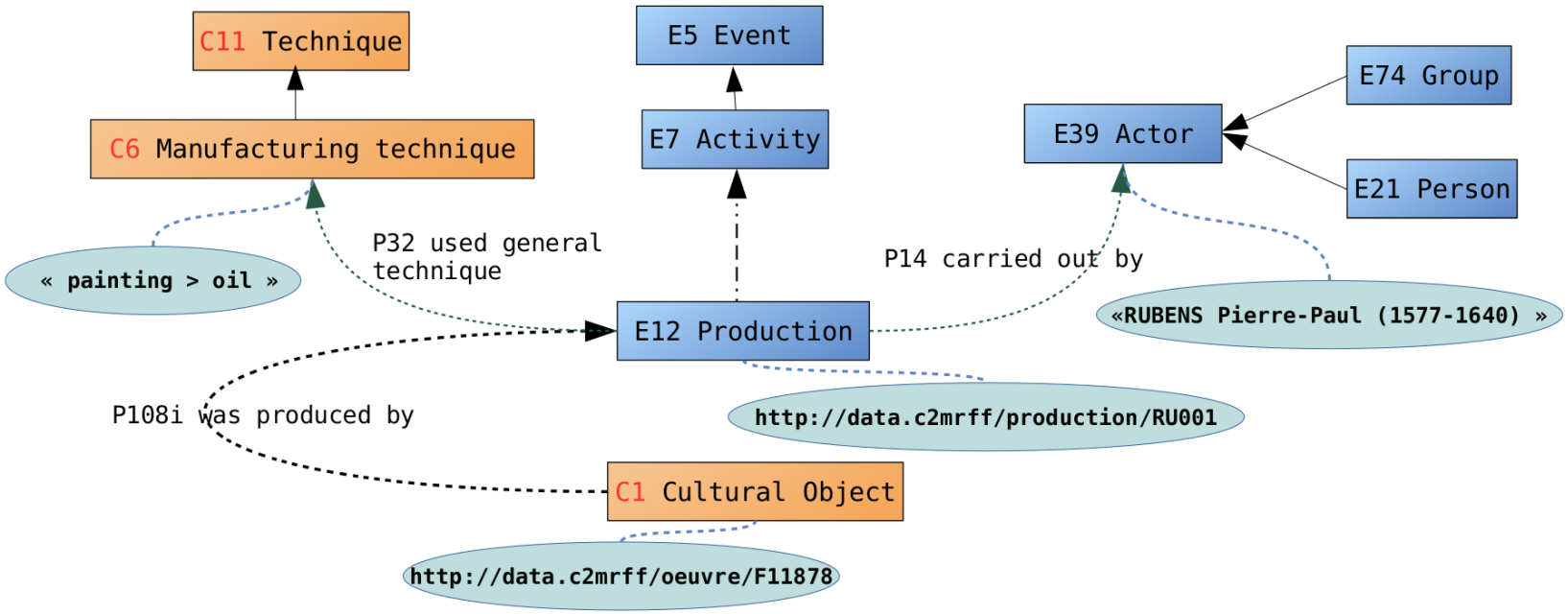
CRMcr: overview of the main concepts



Identification and physical features: Simple paths

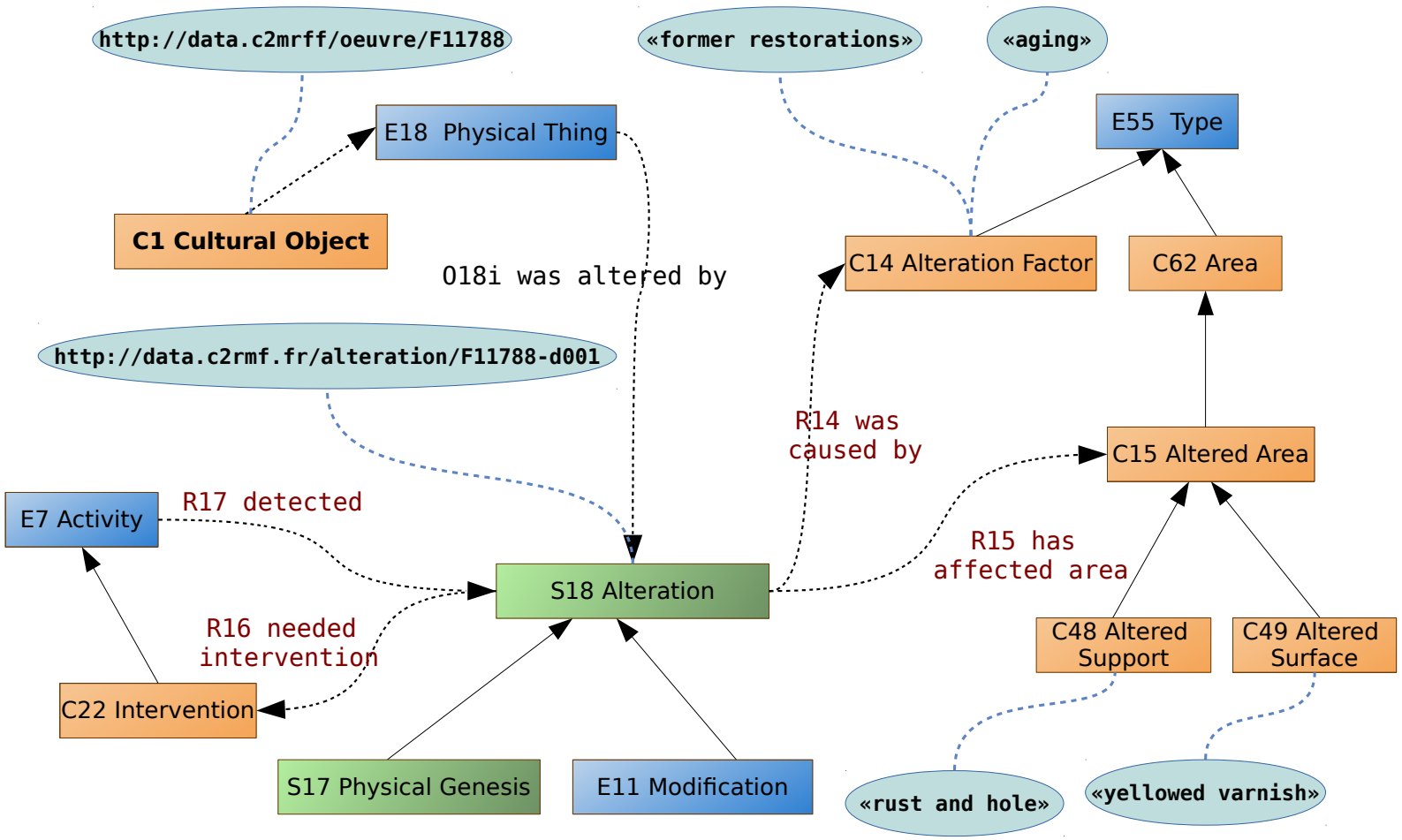


Identification and physical features: Complex paths

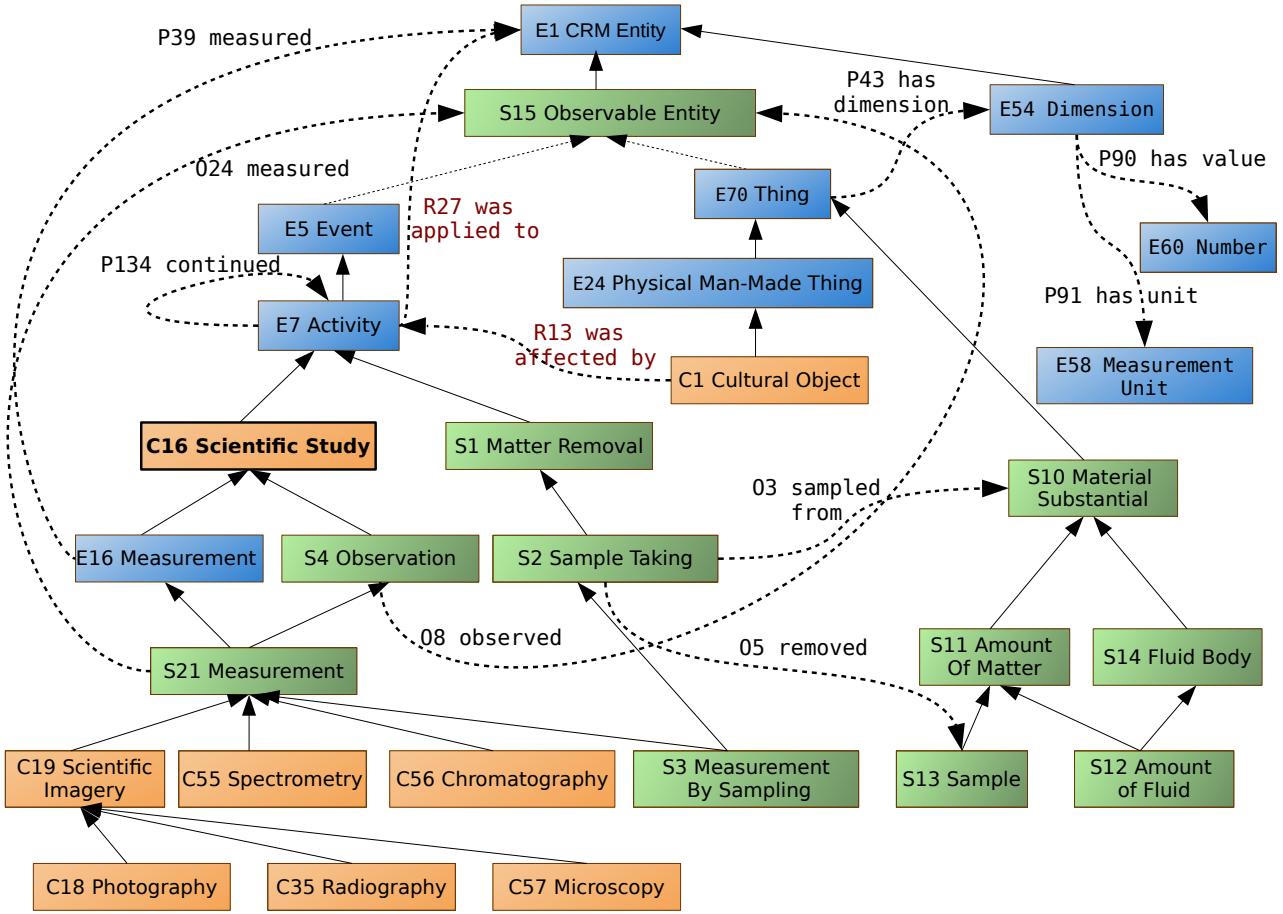


- **The author :**
  C1 Cultural Object → P108i was produced by → E12 Production
  E12 Production → P14 Carried Out By → E39 Actor
- **The manufacturing technique:**
  C1 Cultural Object → P108i Was Produced By → E12 Production
  E12 Production → P32 Used General Technique → C7 Manufacturing Technique

Degrading events: alterations

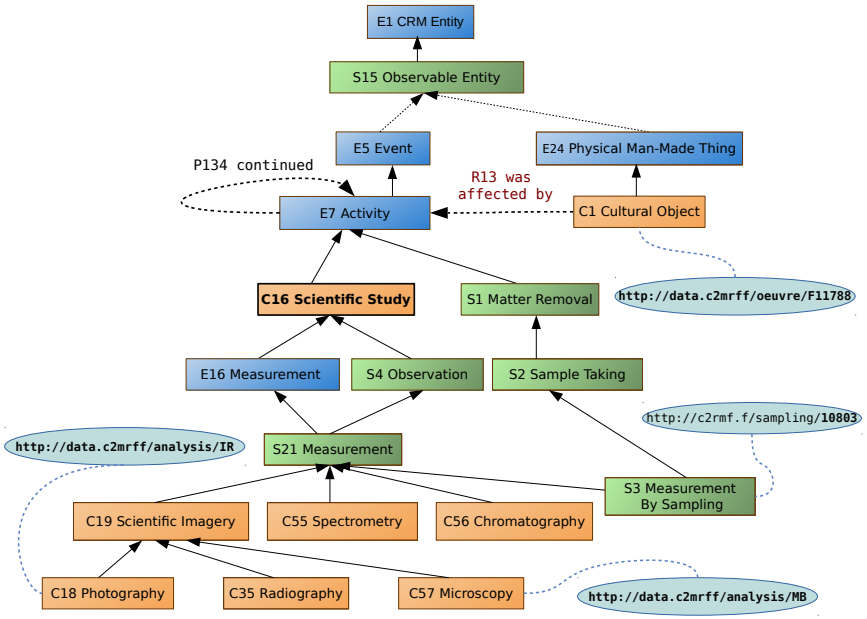


Non-degrading events: activities

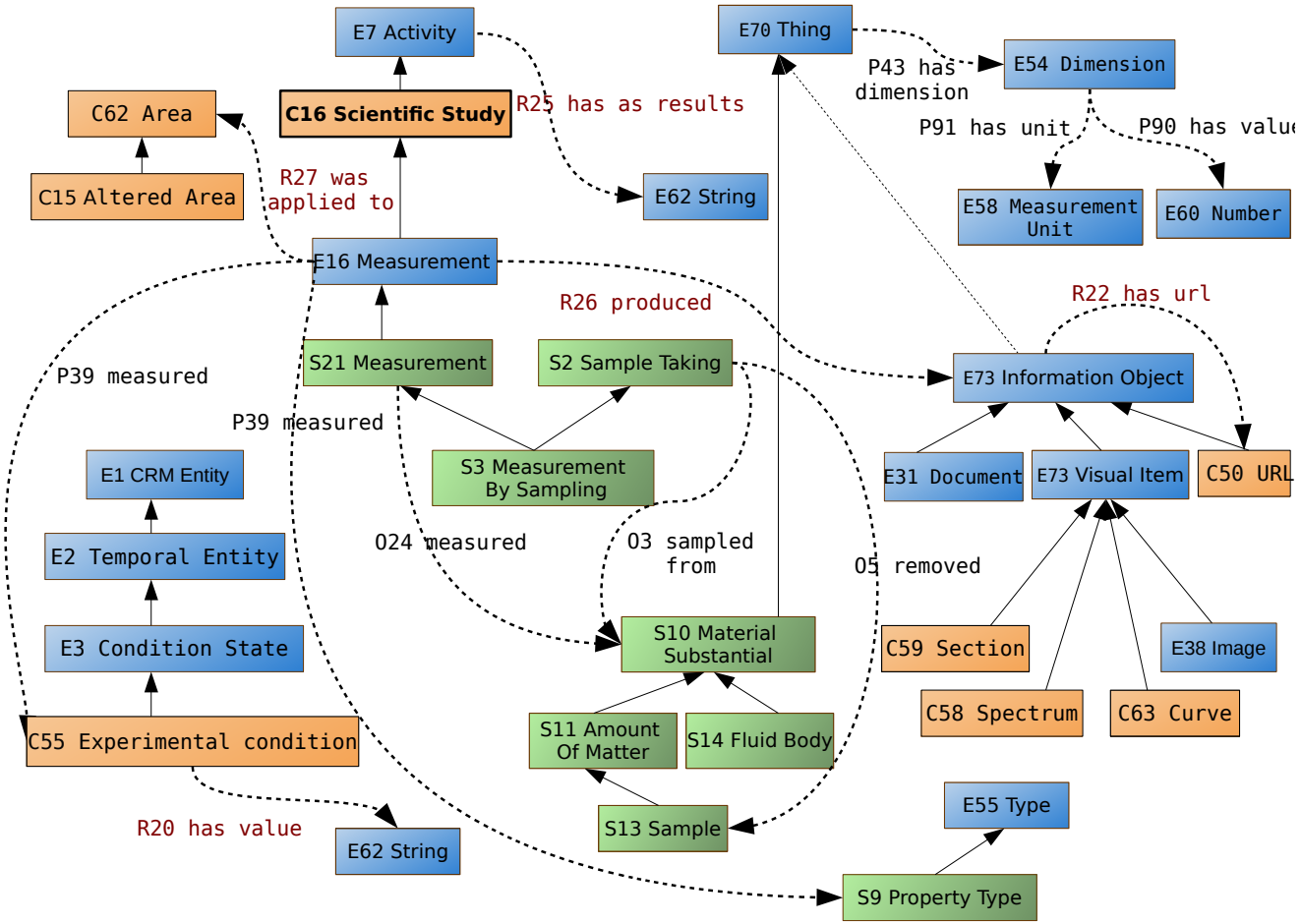


Non-degrading events: activities

- *(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

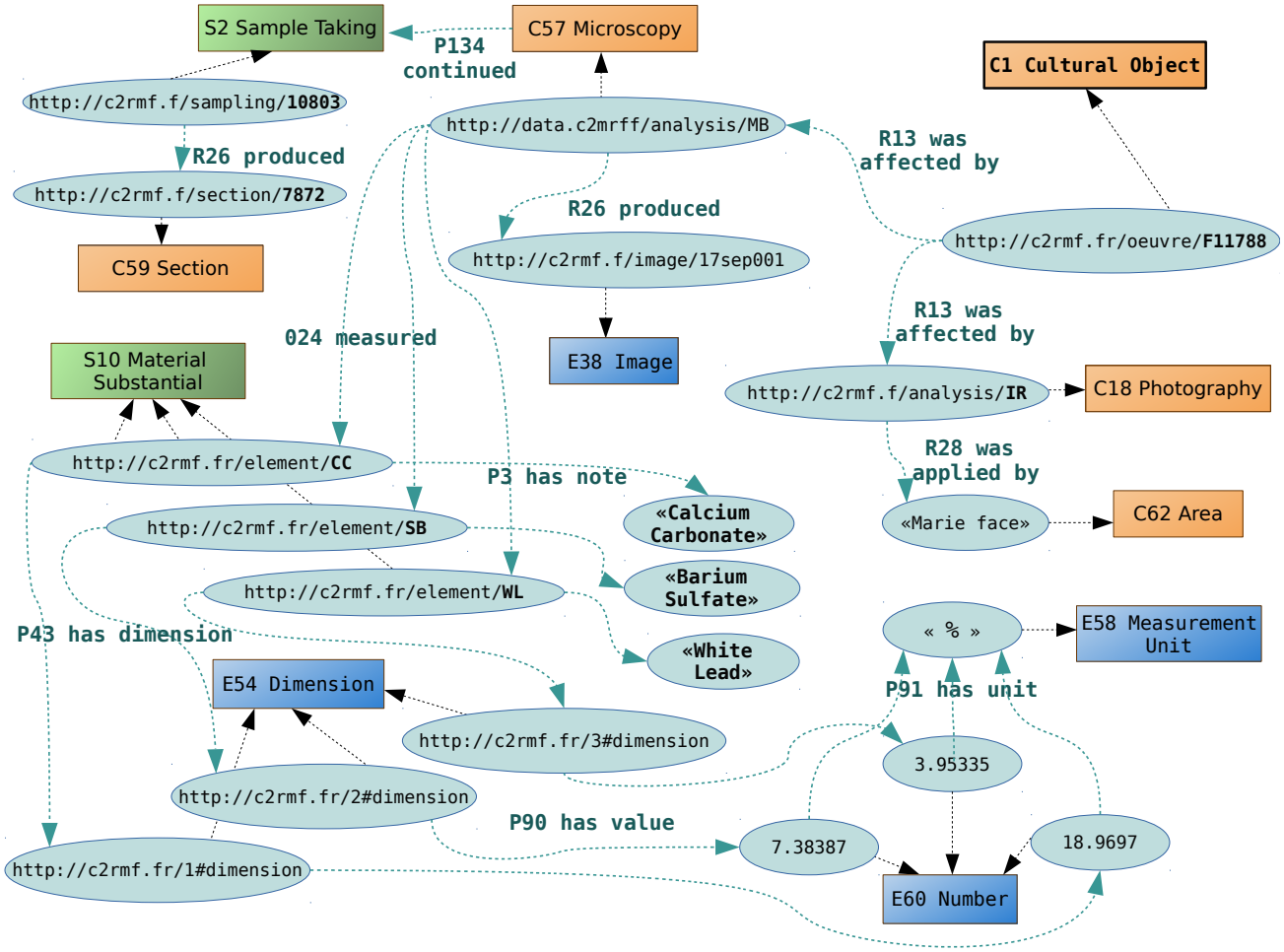


Non-degrading events: activities' results



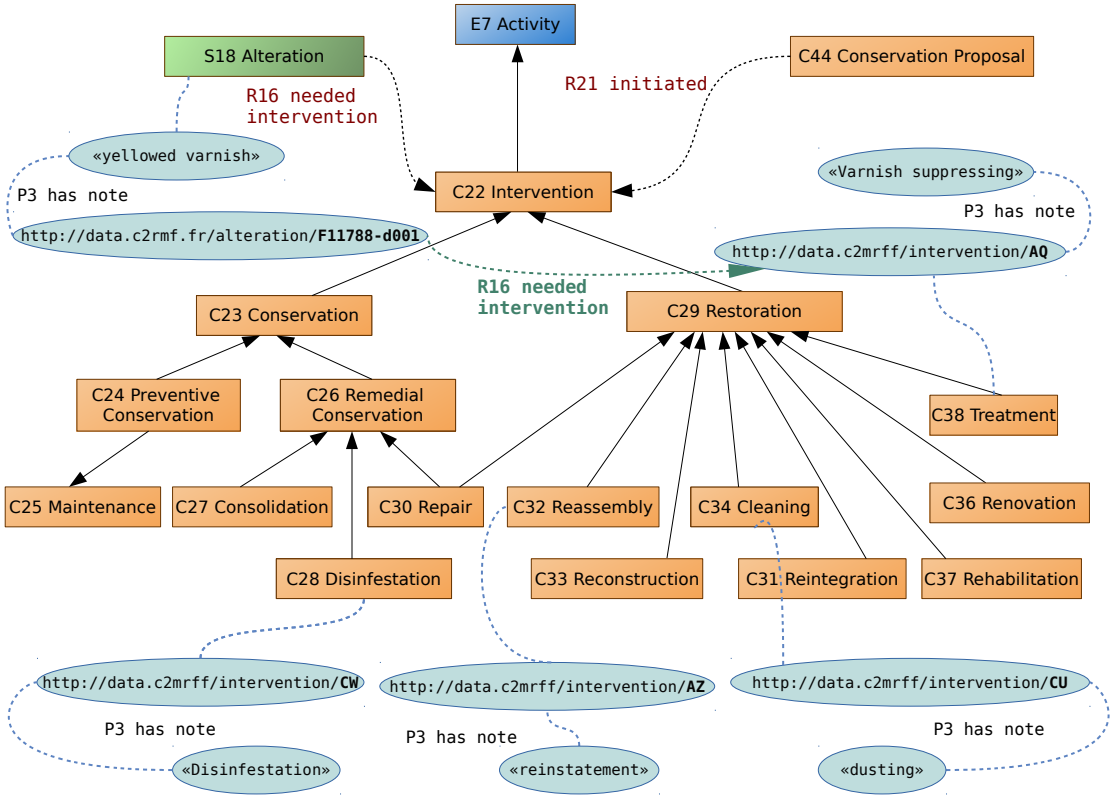
Non-degrading events: activities' results

- *(1) C16 Scientific Study*: an example of analysis results

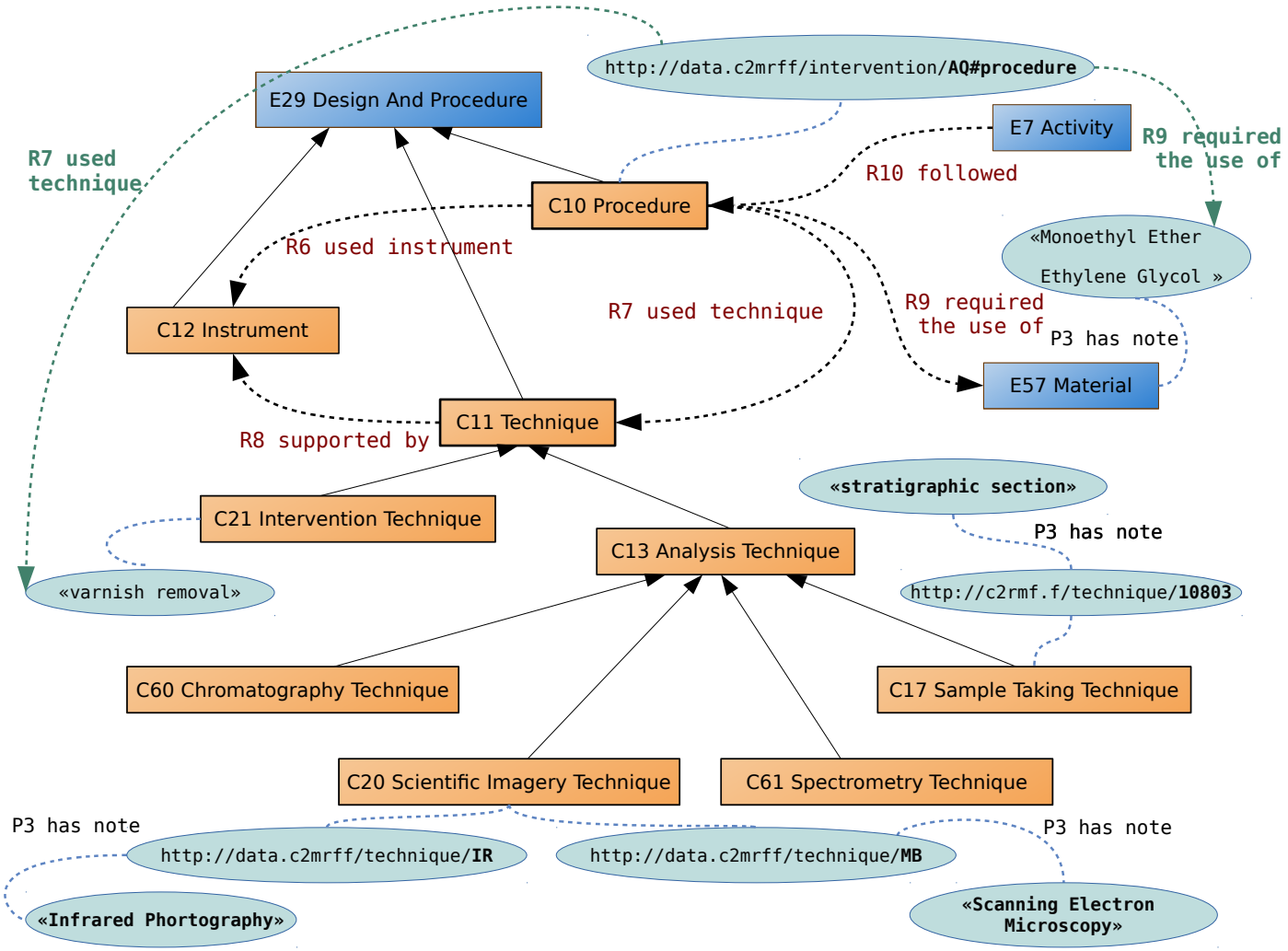


Non-degrading events: activities

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



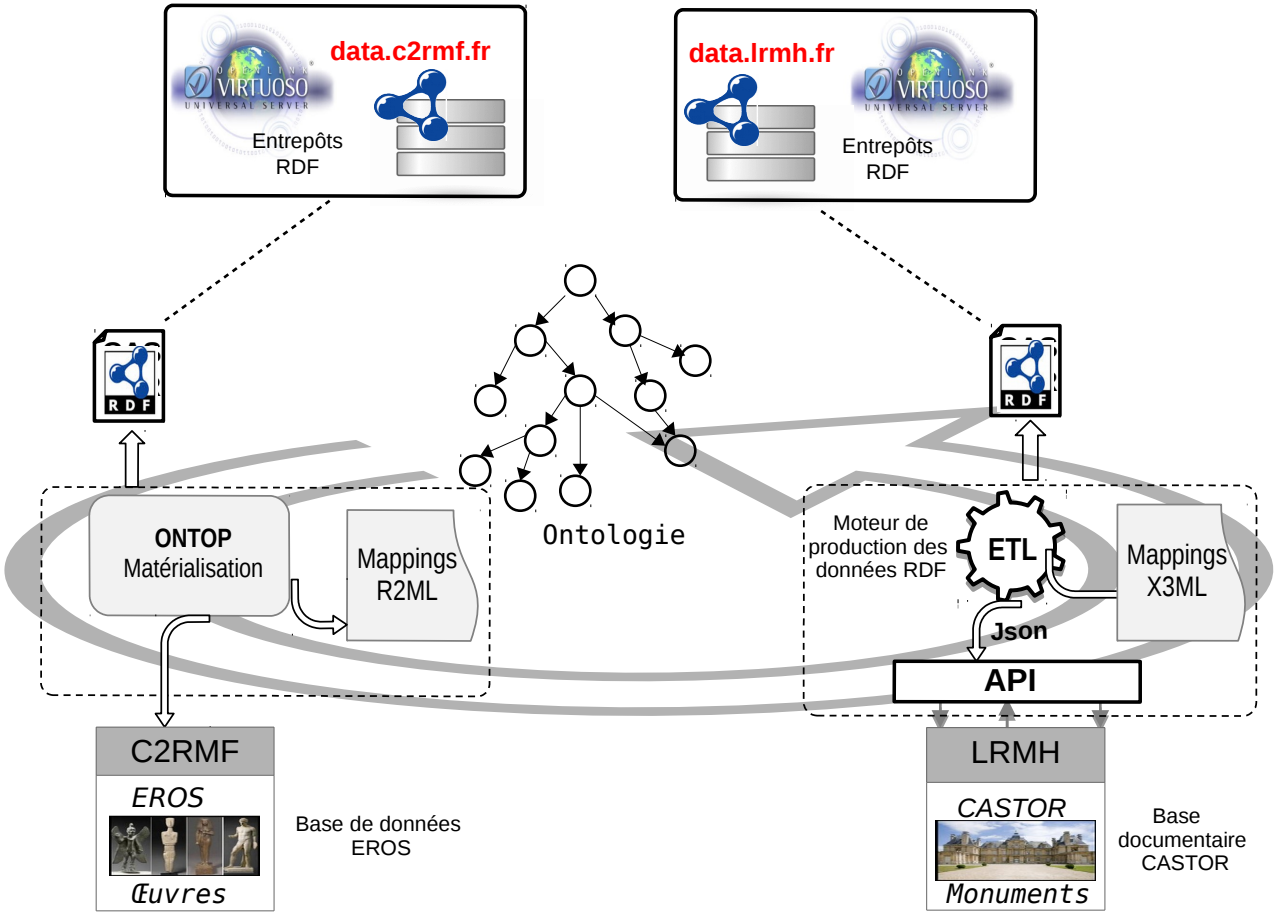
Analysis techniques



Detailed table of contents

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

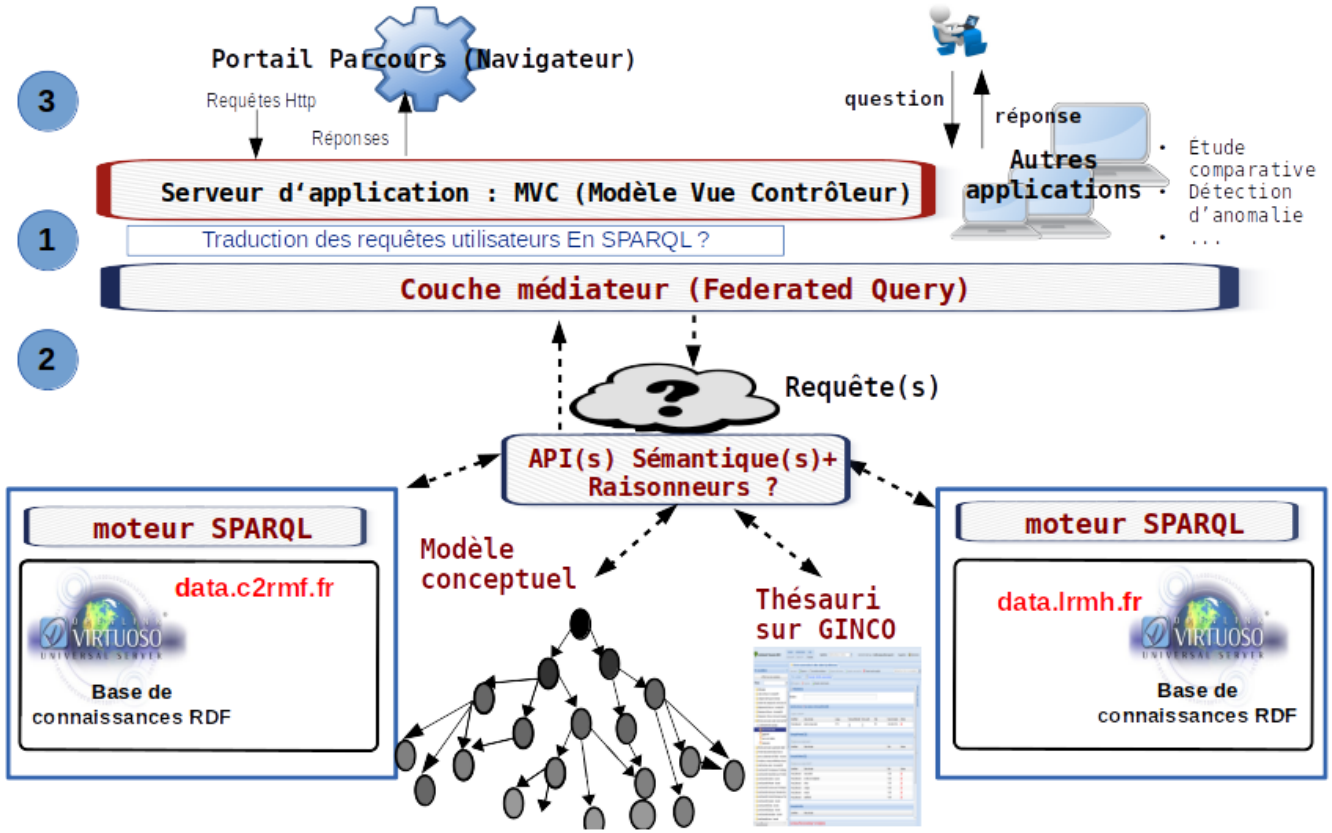
Knowledge base development



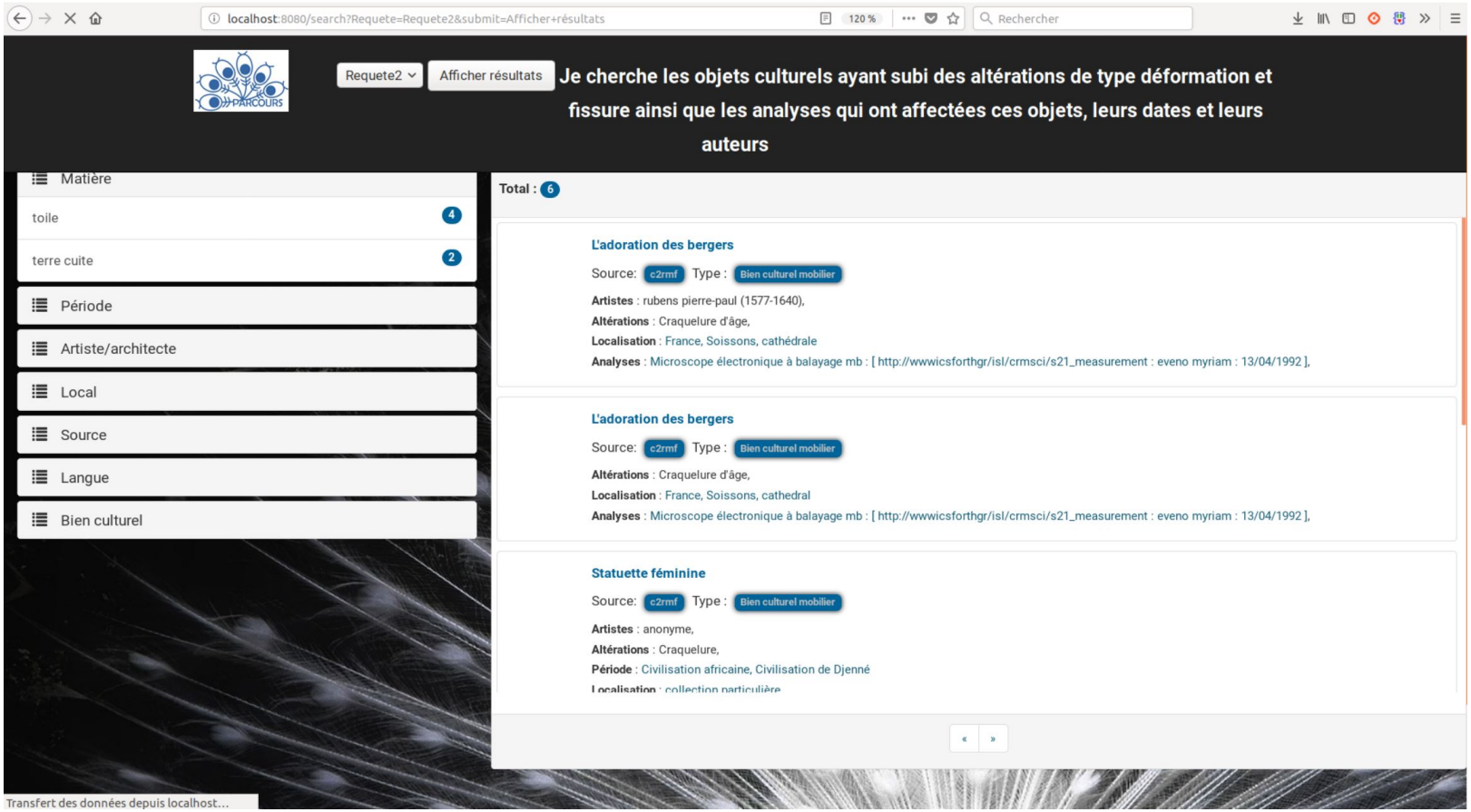
How to query **distributed** databases and which architecture for the query system?

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



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 knwoledge 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 : $\cong$ 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 : $\cong$ 7000 mots
- Mots-clés non classés : $\cong$ 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