



biblissima

patrimoine écrit du Moyen Âge et de la Renaissance

Introduction au CIDOC CRM et au Linked Data

Formation BnF et Biblissima : Modélisation des données et humanités numériques

Pauline Charbonnier (Biblissima)

Bibliothèque nationale de France

1er décembre 2014



Biblissima bénéficie d'une aide de l'Etat gérée par l'ANR au titre du programme Investissements d'avenir (ANR--11--EQPX--0007)

Introduction au CIDOC CRM



Le CIDOC CRM très rapidement

- Développé par le **Comité International pour la Documentation (CIDOC)** de l'**International Council Of Museum (ICOM)**
- **CRM : Conceptual Reference Model**
- Réflexion qui a commencé en 1996
- Norme ISO ([ISO 21127:2006](#)) publiée en 2006
- Standard international pour l'échange contrôlé de données relatives au patrimoine culturel
- 2008 : publication de la version 5
- Décembre 2011 : version 5.0.4, publiée en décembre 2014 comme norme ISO ([ISO 21127:2014](#))
- Janvier 2014 : version 5.1.2 (draft)
- Maintenu par le CRM Special Interest Group

Documents de référence

Site Web du CIDOC CRM : <http://www.cidoc-crm.org/>

Official release : http://www.cidoc-crm.org/official_release_cidoc.html

Introduction :

- [*CRM Primer*](#), Dominic Oldman et CRM Labs
- Liste de tutoriels : <http://www.cidoc-crm.org/tutorials.html>
- [Représentations graphiques](#) du CIDOC CRM

Implémentation (Université d'Erlangen):

<http://erlangen-crm.org/> et <https://github.com/erlangen-crm/ecrm>

Définition du CIDOC CRM

- Le CIDOC CRM est une ontologie qui représente la sémantique sous-jacente des structures de la documentation du patrimoine culturel
- Une ontologie est une formalisation des connaissances
 - concepts fondamentaux d'un domaine d'information et relations qui les unissent
- une ontologie empirique
- modèle conceptuel de référence
- modèle orienté objet
- Conceptual **Reference** Model : indépendant de toute implémentation technique
 - non prescriptif
 - ne se préoccupe pas des terminologies

Objectifs du CIDOC CRM

“The primary role of the CRM is to enable exchange and integration between heterogeneous sources of cultural heritage information.”

- fournir des définitions sémantiques
- faciliter l'échange et la recherche d'informations
- cadre pour l'harmonisation des données

Structure du CIDOC CRM - classes et propriétés

- Entités (classes) et relations (propriétés)
 - Une classe = une catégorie d'items qui partagent une ou plusieurs caractéristique(s) commune(s)
 - Un item appartenant à cette classe est appelé une instance de cette classe
 - Les classes sont connectées entre elles via des propriétés
 - Une classe peut être le domaine ou le co-domaine de propriétés (0, 1 ou n)



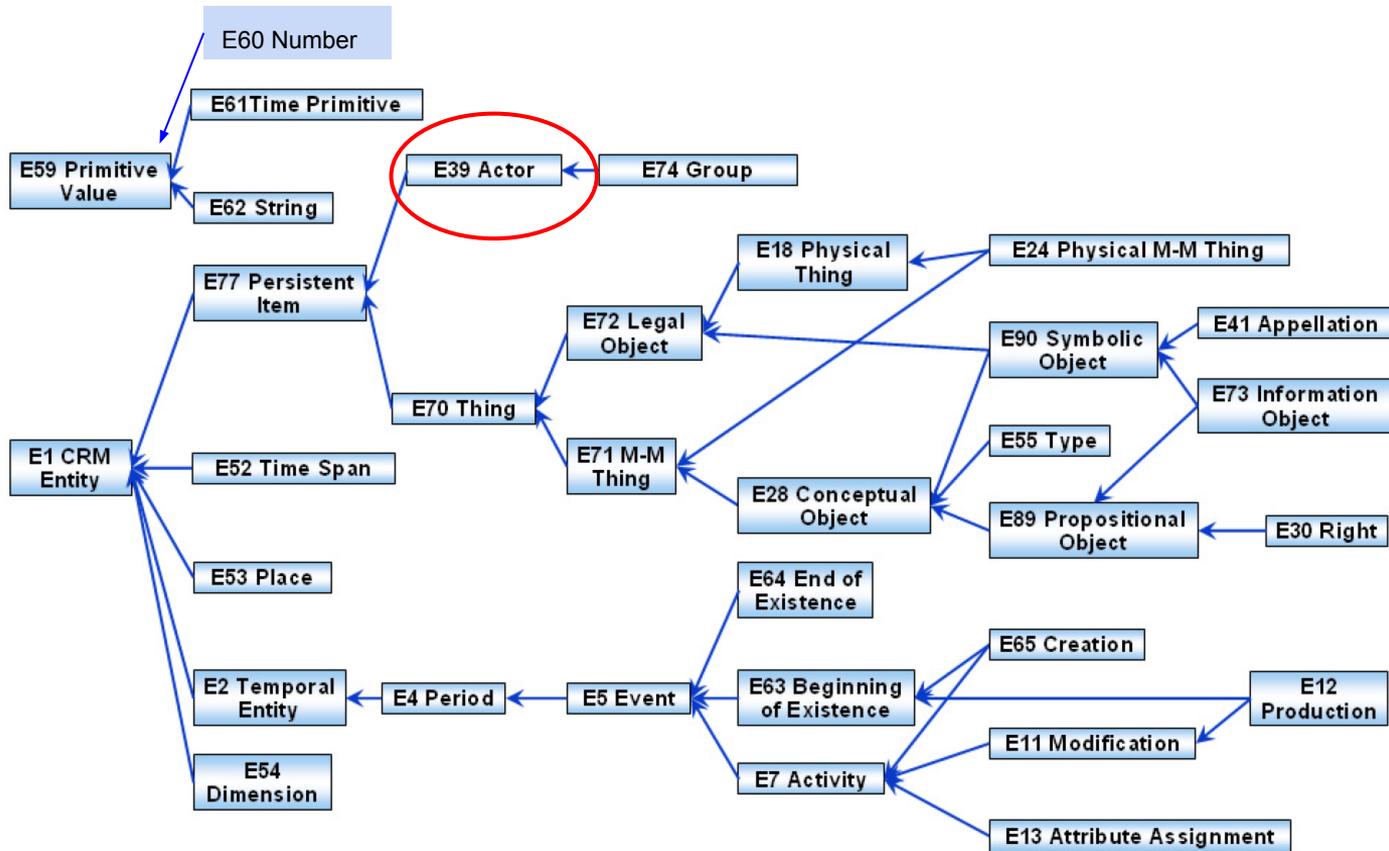
Structure du CIDOC CRM - classes et propriétés

- Classes : E (pour Entity) + un numéro + un nom
 - e.g. *E39 Actor*
- Propriété : P (pour Property) + un numéro + un nom
 - e.g. *P11 had participant*
- environ 85 classes et 283 propriétés dans le CIDOC CRM (source : [ECRM](#))

Structure du CIDOC CRM - classes et sous-classes

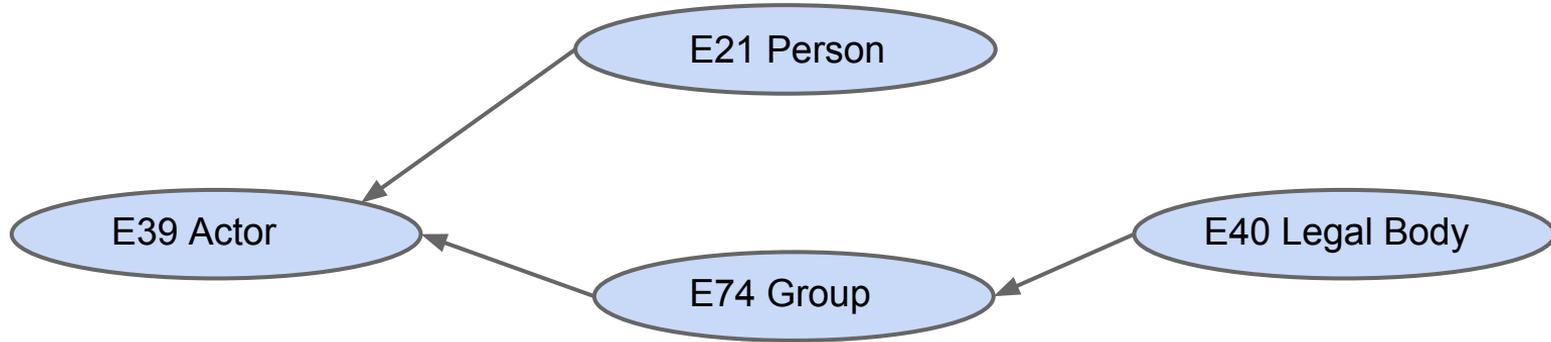
- Les classes ont des sous-classes
- Classe E1 CRM Entity : sommet de la hiérarchie des classes
 - *“This class comprises all things in the universe of discourse of the CIDOC Conceptual Reference Model.”*
- Une sous-classe est une spécialisation d’une autre classe (sa super-classe)
 - instances de cette sous-classe sont aussi des instances de la super-classe
 - caractéristiques plus restrictives
 - hérite de la définition et des propriétés de la super-classe (+ propriétés spécifiques)
- Hiérarchie de classes permet d’utiliser différents niveaux de généralisation et de spécialisation.

Hierarchy of core classes



Source : [CIDOC-CRM graphical representation, Hierarchy of core classes](#)

Structure du CIDOC CRM - exemple de sous-classe (E39 Actor)



 : classe

 : est sous-classe de

Structure du CIDOC CRM - documentation d'une classe

(source : http://cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf)

E39 Actor

Subclass of: E77 Persistent Item

Superclass of: E21 Person
E74 Group

Scope note: This class comprises people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible.

The CRM does not attempt to model the inadvertent actions of such actors. Individual people should be documented as instances of E21 Person, whereas groups should be documented as instances of either E74 Group or its subclass E40 Legal Body.

Examples:

- London and Continental Railways (E40)
- the Governor of the Bank of England in 1975 (E21)
- Sir Ian McKellan (E21)

Properties:

P74 has current or former residence (is current or former residence of): E53 Place

P75 possesses (is possessed by): E30 Right

P76 has contact point (provides access to): E51 Contact Point

P131 is identified by (identifies): E82 Actor Appellation

Structure du CIDOC CRM - les propriétés d'objet

- Une propriété définit une relation particulière entre deux classes
- Une propriété se caractérise par sa signification qui est définie dans une note d'application.

- Une propriété a une classe domaine (classe de départ) et une classe co-domaine (classe d'arrivée)
 - domaine (*domain*) : classe pour laquelle une propriété est définie
 - co-domaine (*range*) : classe qui comprend toutes les valeurs possibles d'une propriété.
 - deux sens de lecture

Structure du CIDOC CRM - documentation d'une propriété

P11 had participant (participated in)

Domain: E5 Event

Range: E39 Actor

Subproperty of: E5 Event. P12 occurred in the presence of (was present at): E77 Persistent Item

Superproperty of: E7 Activity. P14 carried out by (performed): E39 Actor

E67 Birth. P96 by mother (gave birth): E21 Person

E68 Dissolution. P99 dissolved (was dissolved by): E74 Group

E85 Joining. P143 joined (was joined by): E39 Actor

E85 Joining. P144 joined with (gained member by): E74 Group

E86 Leaving. P145 separated (left by): E39 Actor

E86 Leaving. P146 separated from (lost member by): E74 Group

P151 was formed from: E74 Group

Quantification: many to many (0,n:0,n)

Scope note: This property describes the **active or passive** participation of instances of E39 Actors in an E5 Event.

It connects the life-line of the related E39 Actor with the E53 Place and E50 Date of the event. The property implies that the Actor was involved in the event but does not imply any causal relationship. The subject of a portrait can be said to have participated in the creation of the portrait.

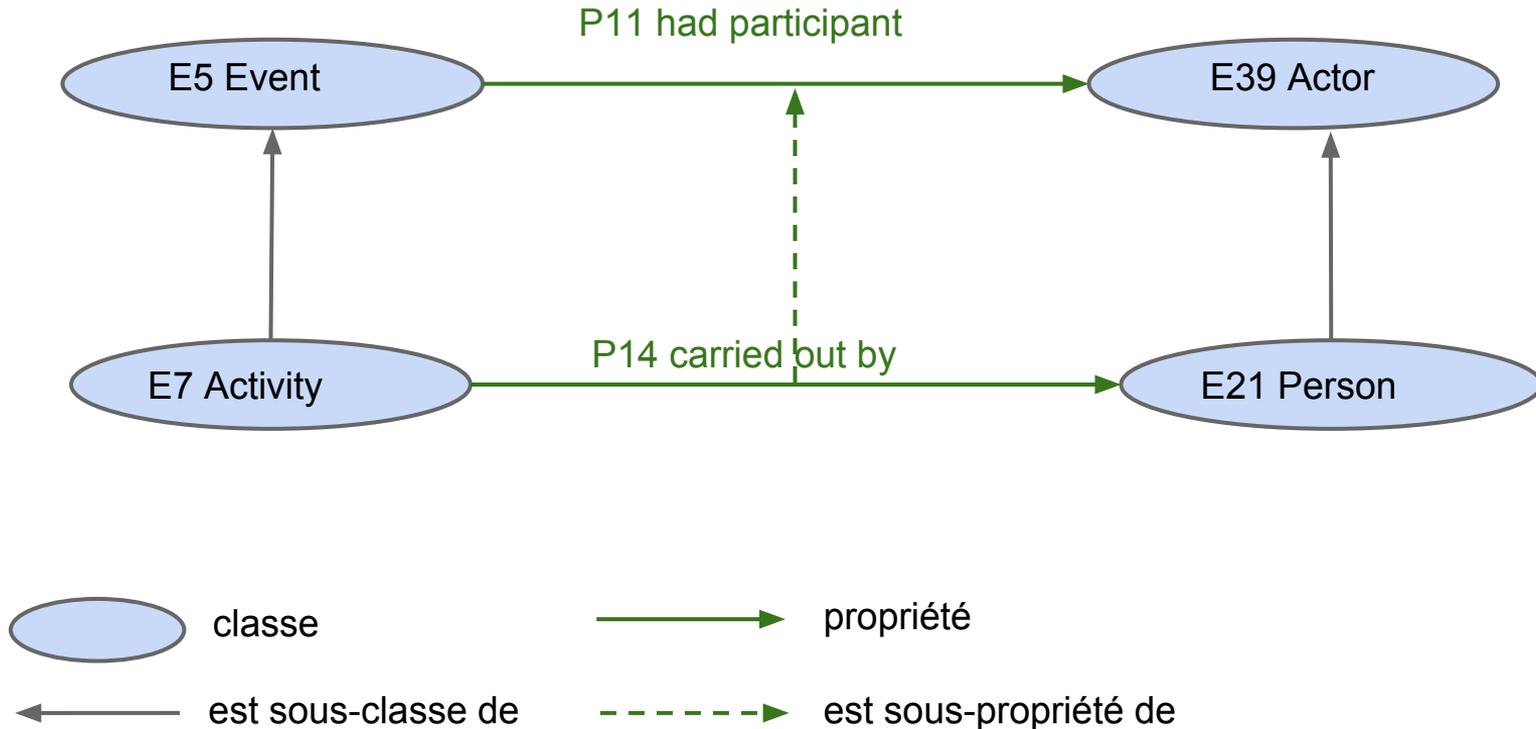
Examples:

- Napoleon (E21) *participated in* The Battle of Waterloo (E7)
- Maria (E21) *participated in* Photographing of Maria (E7)

Structure du CIDOC CRM - hiérarchie des propriétés

- Une sous-propriété est une spécialisation d'une autre propriété (sa super-propriété)
 - instances de cette sous-propriété sont aussi des instances de la super-propriété
 - caractéristiques plus restrictives
 - le domaine de la sous-propriété est le même que celui de sa super-propriété ou une sous-classe de ce domaine.
 - le co-domaine de la sous-propriété est le même que celui de sa super-propriété ou une sous-classe de ce co-domaine
 - hérite de la définition de la super-propriété
- Hiérarchie des propriétés permet d'utiliser différents niveaux de généralisation et de spécialisation.

Structure du CIDOC CRM - exemple de sous-propriétés



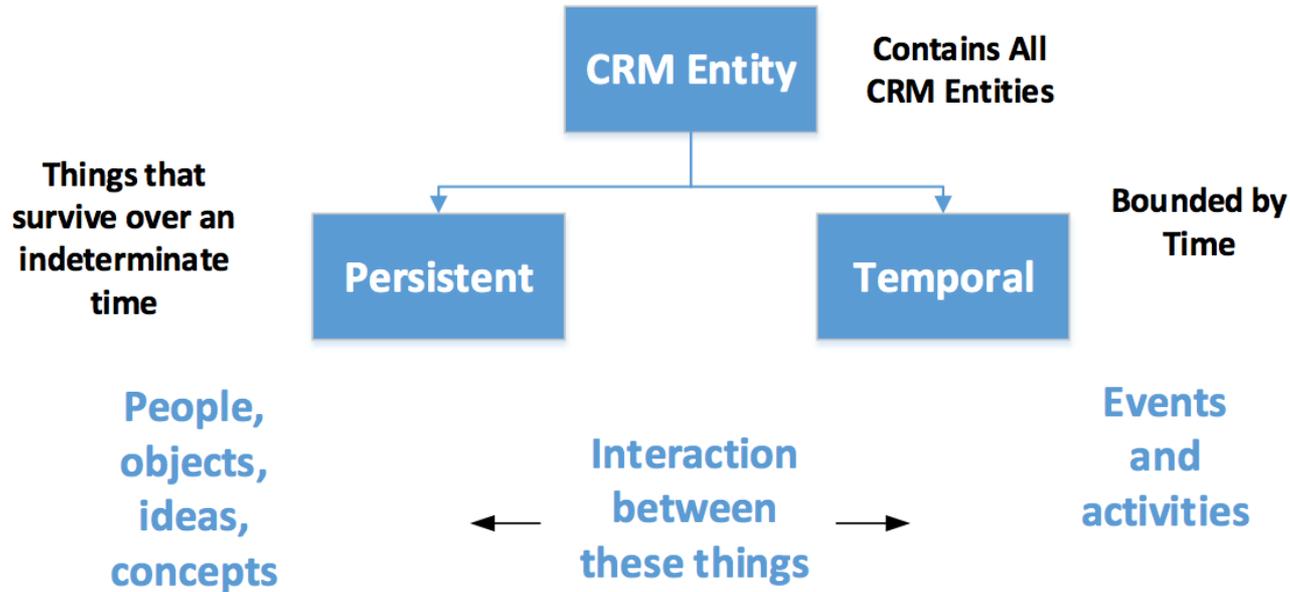
Structure du CIDOC CRM - valeurs primitives et propriétés de type de données

- Les valeurs primitives (E59 Primitive Value) ne font pas partie de la hiérarchie générale du CIDOC CRM
- 3 sous-classes de E59 Primitive Value :
 - Primitif temporel (E61 Time Primitive) (durée/date)
 - Chaîne de caractères (E62 String)
 - Nombre (E60 Number)
- Système de valeurs primitives : propre à l'implémentation
(e.g. XML Schema datatypes)

Structure du CIDOC CRM - valeurs primitives et propriétés de type de données

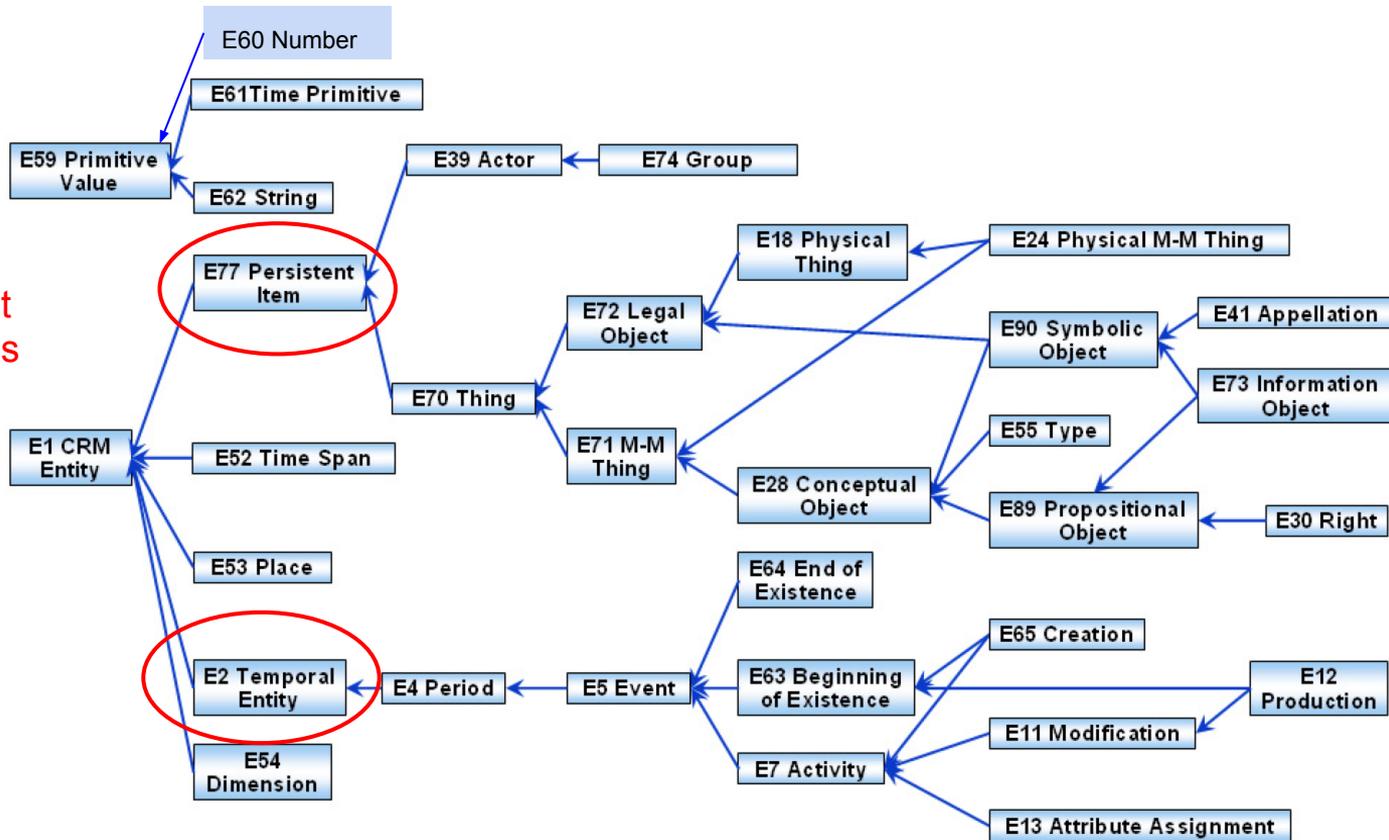
- Les propriétés de type de données relient des classes à des valeurs de données
- Distinction propriété d'objet VS propriété de type de données
- exemples de *datatype properties* :
 - E1 CRM Entity -> P3 has note -> E62 String
 - E19 Physical Object -> P57 has number of parts -> E60 Number
 - E52 Time-Span -> P81 ongoing throughout -> E61 Time Primitive
 - E52 Time-Span -> P82 at some time within -> E61 Time Primitive
- à la différence des propriétés d'objet, les propriétés de type de données ne sont exprimées que dans un sens (classe -> valeur)

Structure du CIDOC CRM - entités persistantes et entités temporelles



Hierarchy of core classes

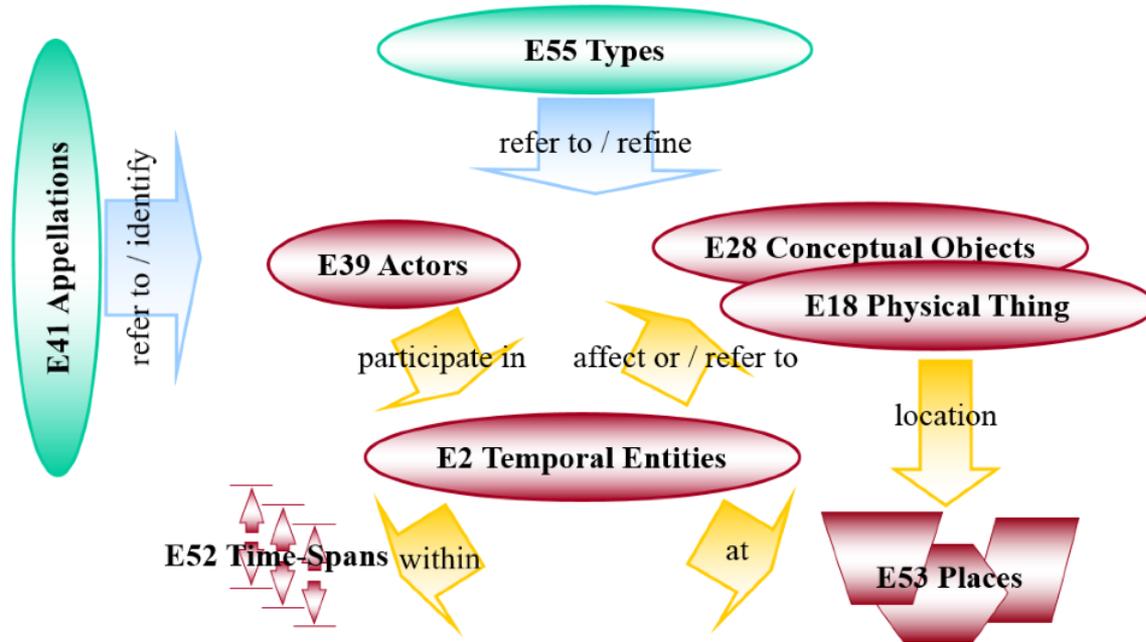
disjoint classes



Source : [CIDOC-CRM graphical representation, Hierarchy of core classes](#)

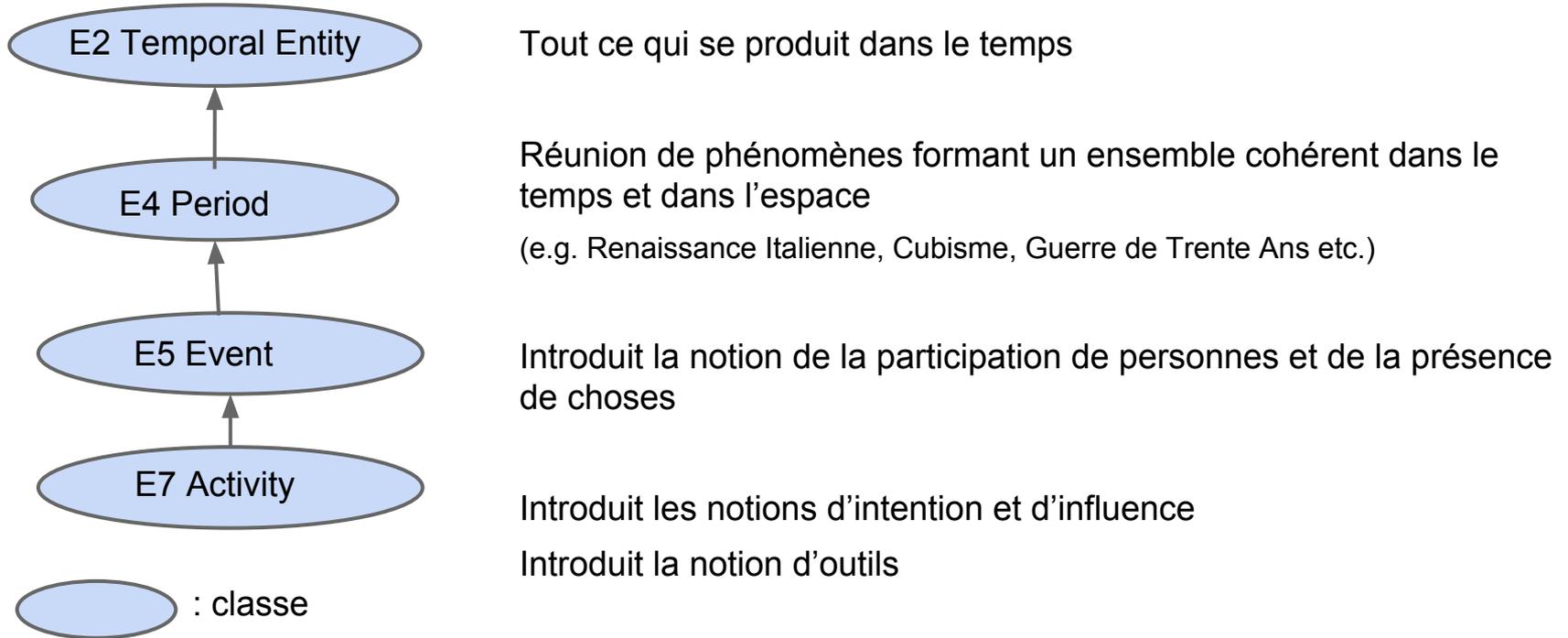
The CIDOC CRM

Top-level classes useful for integration



Source : [CRM Primer](#), Dominic Oldman, CRM Labs

Structure du CIDOC CRM - entités temporelles



← : est sous-classe de

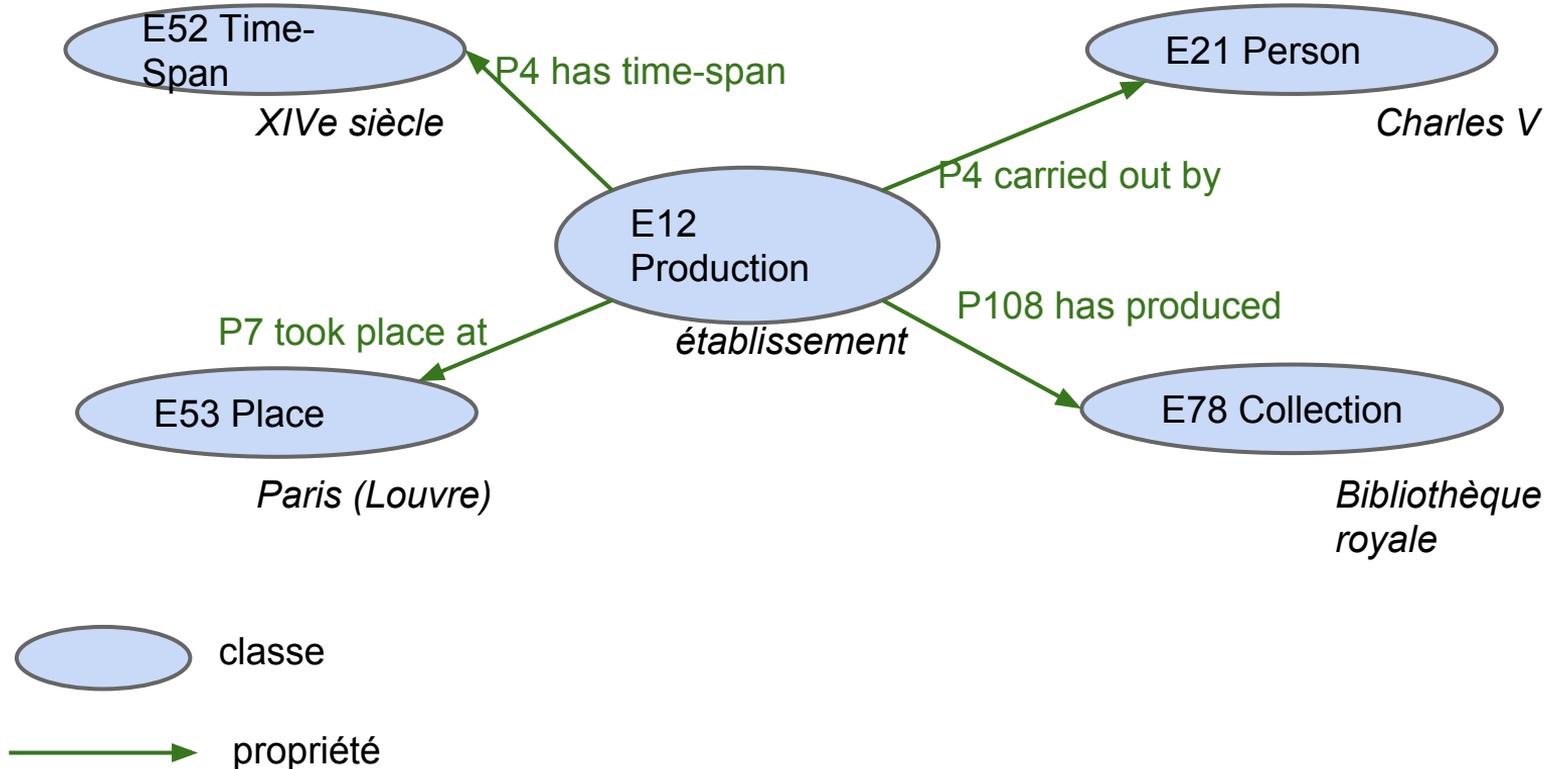
Source : [Présentation du CIDOC CRM](#), Patrick Le Boeuf

CIDOC CRM - notions centrales

- Durée (E52) : seules les entités temporelles (événements) peuvent être liées au temps
- Lieu (E53) : géographique ou relatif, lié aux entités temporelles mais aussi aux objets, aux acteurs
- Agent (E39):
 - sont des entités qui effectuent intentionnellement des actions dont ils peuvent être tenus responsables
 - interagissent avec des objets conceptuels ou des choses matérielles
- Choses matérielles (E18) VS Objets conceptuels (E28)

Structure du CIDOC CRM - exemple d'événement

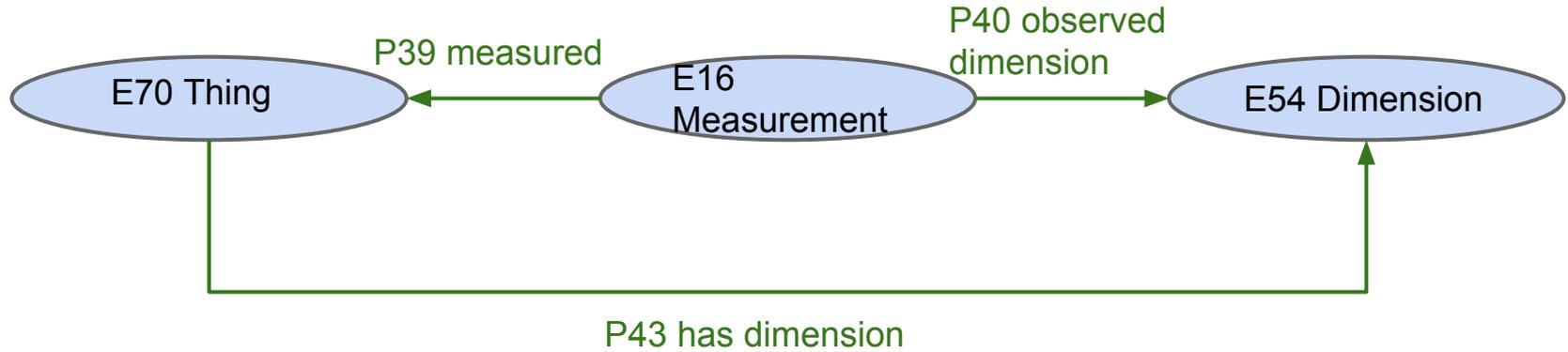
L'établissement de la bibliothèque royale au Louvre



CIDOC CRM - notions centrales

- Appellation : une chose et son nom sont deux entités distinctes dans le CIDOC CRM
 - E39 Actor -> P131 is identified by -> E82 Actor Appellation
 - E53 Place -> P87 is identified by-> E44 Place Appellation
 - E52 Time-Span -> P78 is identified by -> E49 Time Appellation
 - E71 Man-Made Thing -> P102 has title -> E35 Title
- Type (E55):
 - P2_has type
 - gestion terminologique externalisée (thésaurus, etc.)

Structure du CIDOC CRM - raccourci

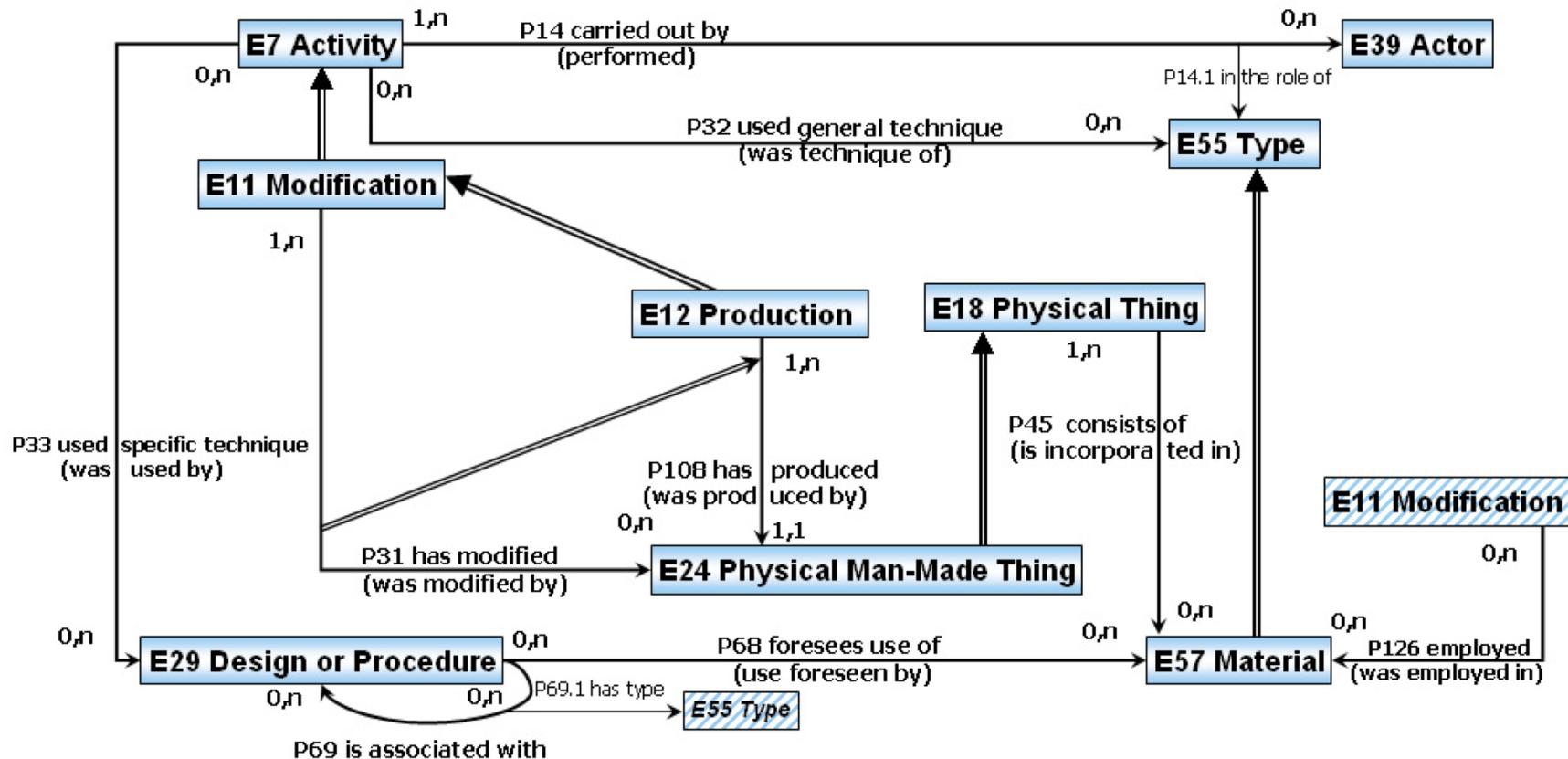


 classe

 propriété

P43_has dimension : "It offers no information about how and when an E54 Dimension was established, nor by whom."

MATERIAL AND TECHNIQUE INFORMATION



Source : [CIDOC-CRM graphical representation, Material and technique information](#)

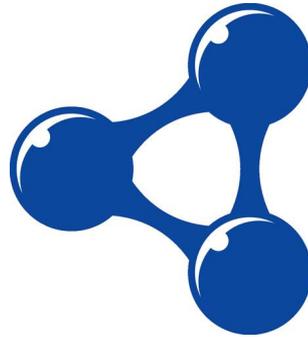
Comment utiliser le CIDOC CRM ?

- identifier les types classes et propriétés nécessaires
- mapping des données vers ces classes et propriétés
- il n'est pas nécessaire d'implémenter l'ensemble du CIDOC CRM
- extensions
 - [CRMEH](#) : Conceptual Reference Model English Heritage CRMEH - English Heritage (domaine archéologique)
 - [CRM geo](#) : intégration de l'information spatiale
 - [CRM SCI](#) : axée sur l'observation scientifique
 - [CRM dig](#) : axée sur la numérisation et les représentations numériques
 - [FRBRoo](#) : harmonisation CIDOC CRM et FRBR (Functional Requirements for Bibliographic Records)

Quelques projets utilisant le CIDOC CRM :

- [British Museum](#)
- [Projet Claros](#)
- [Yale, Center for British Art](#)
- [Sharing Ancient Wisdoms](#)
- [Biblissima](#) (work in progress)

Introduction au Linked Data



Linked Data / Web de données

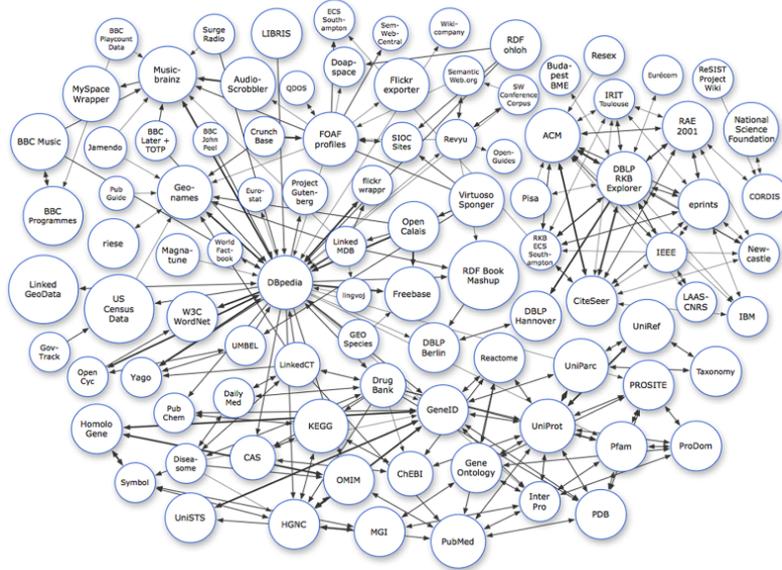
- Lier les données sur le Web à des entités du monde réel dans un format compréhensible par les machines
- 4 principes :
 - URI (*Uniform Resource Identifier*)
 - HTTP (*HyperText Transfert Protocol*)
 - fournir via ces URI des renseignements lisibles par les humains et les machines (via RDF, SPARQL, etc.)
 - liens vers autres jeux de données du LOD (Linked Open Data)

Exemples de jeux de données dans le LOD :

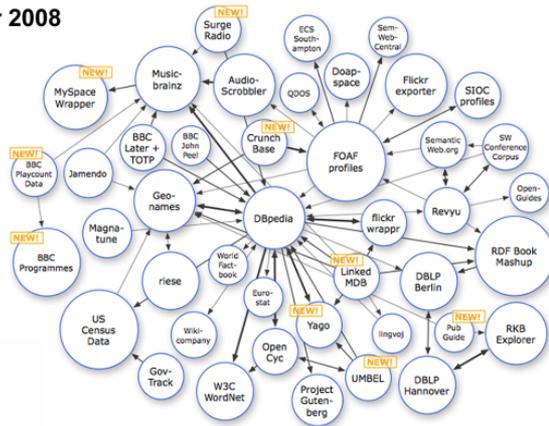
- [DBpedia](#)
- [GeoNames](#)
- [data.bnf.fr](#)

The Linked Open Data Cloud (2007-2009)

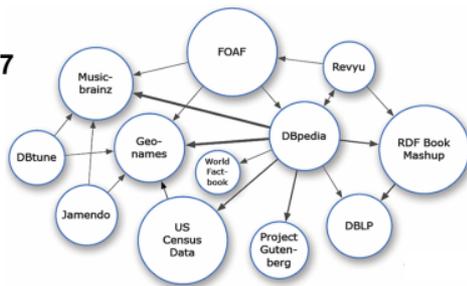
July 2009



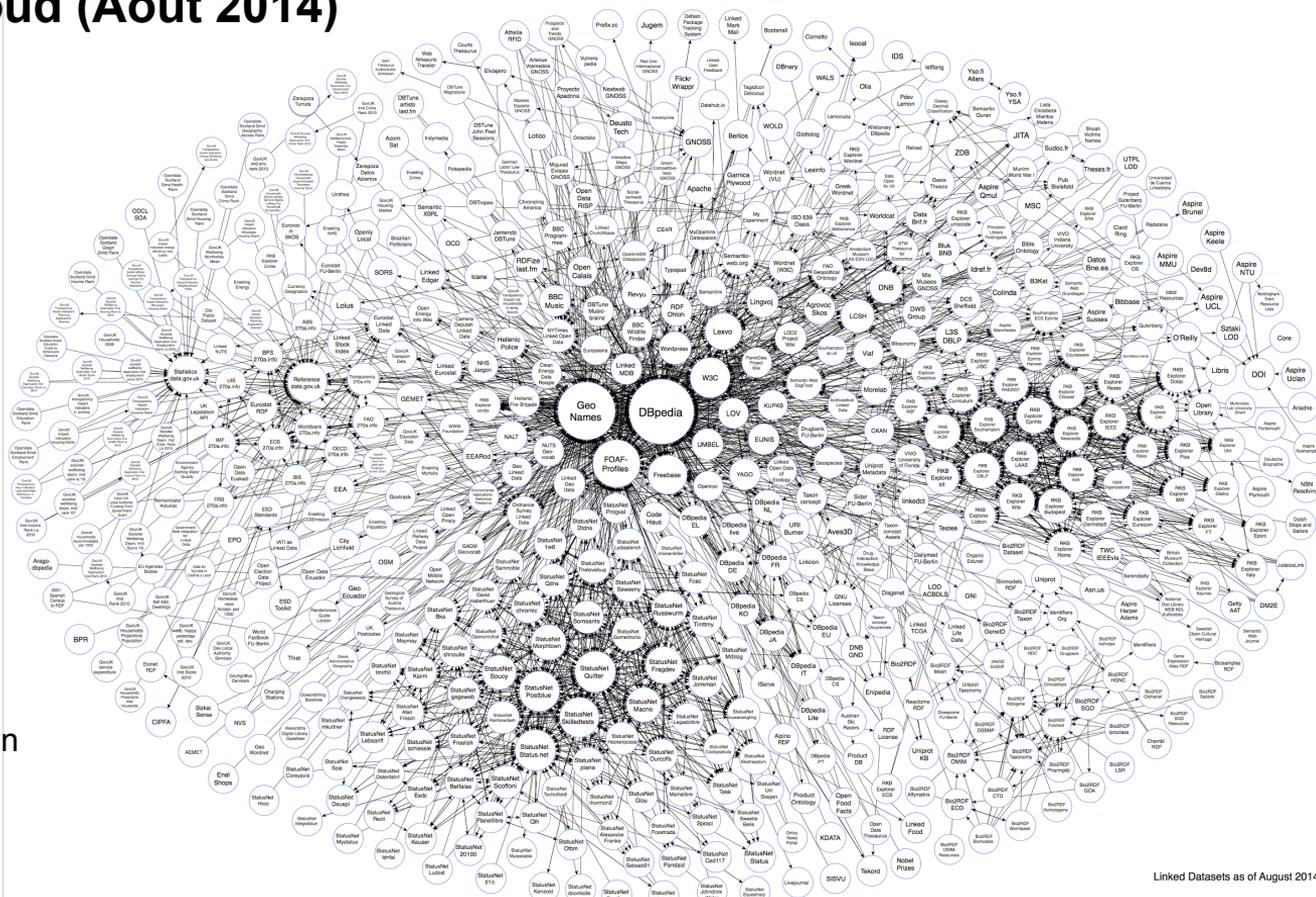
September 2008



May 2007

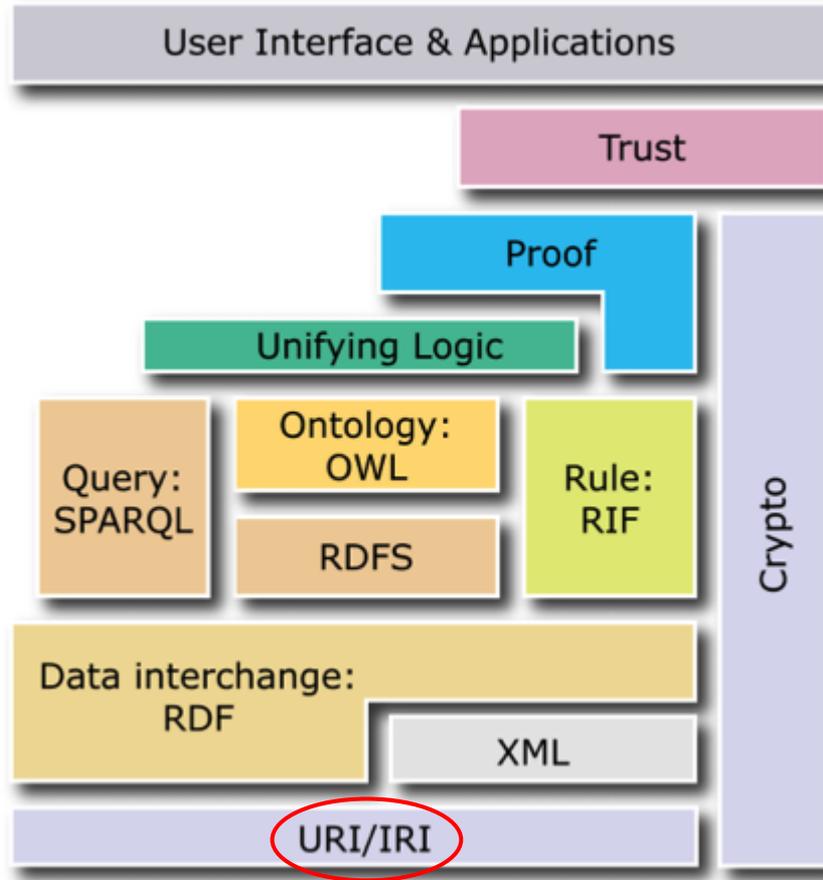


The LOD Cloud (Août 2014)



Source : The Linking Open Data cloud diagram <http://lod-cloud.net>

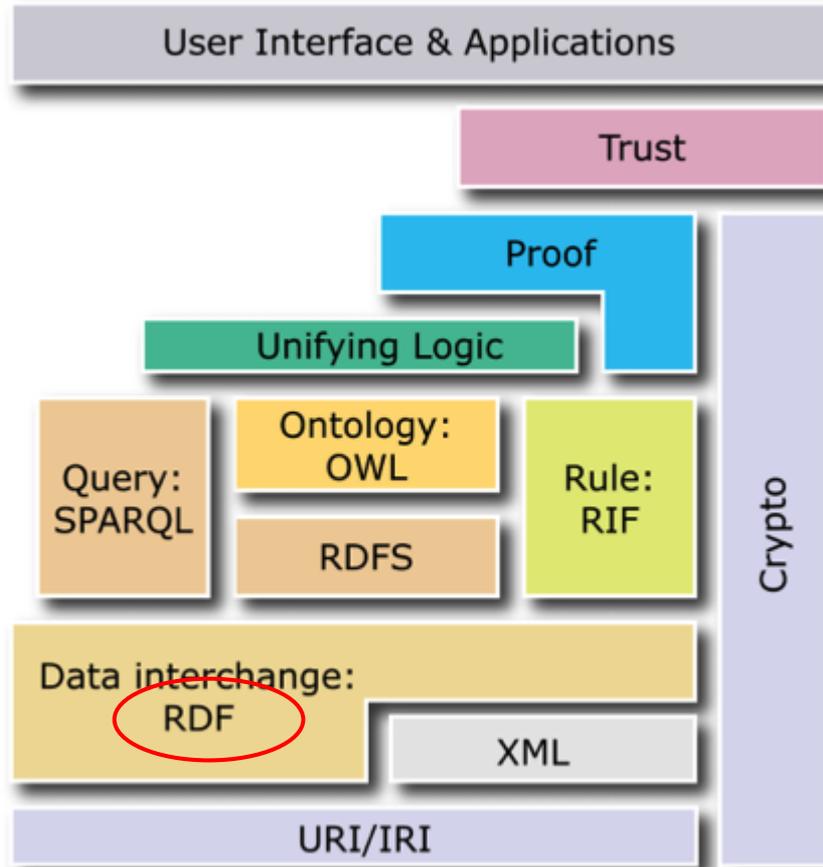
Semantic Web layer cake



URI

- **URI** (*Uniform Resource Identifier*)
- identifiant sur le Web des concepts/entités du monde réel
- e.g. <http://data.bnf.fr/ark:/12148/cb11902007r> :
= URI identifiant Eusèbe de Césarée

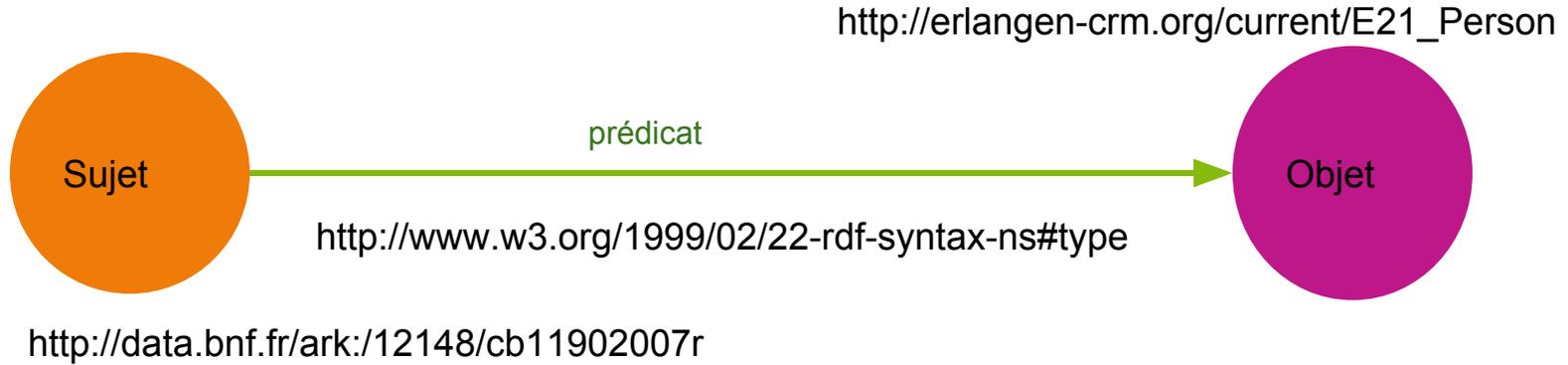
Semantic Web layer cake



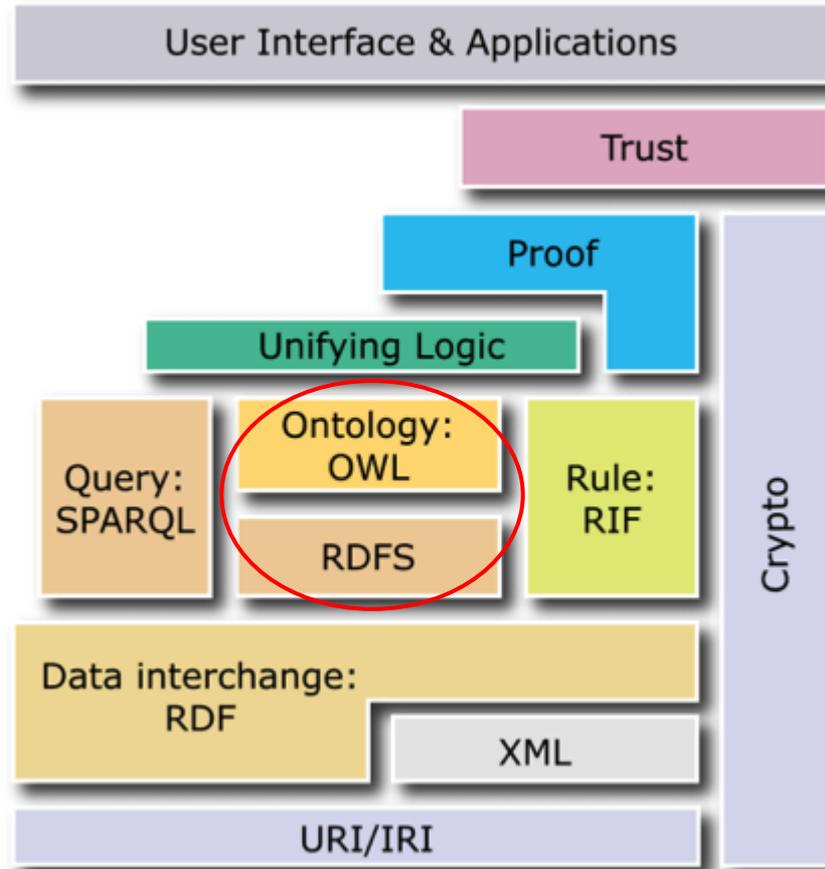
Triplet RDF (Resource Description Framework)

- Standard W3C pour le Linked Data
- Identification sur le Web des concepts/entités du monde réel avec des URI
- Expression des relations entre ces ressources sous la forme de triplets
- Chaque élément du triplet peut être identifié par une URI
- Une même ressource peut être sujet, prédicat ou objet dans plusieurs triplets.
- Modèle de graphe

Triplet RDF (Resource Description Framework)



Semantic Web layer cake



OWL & RDFS

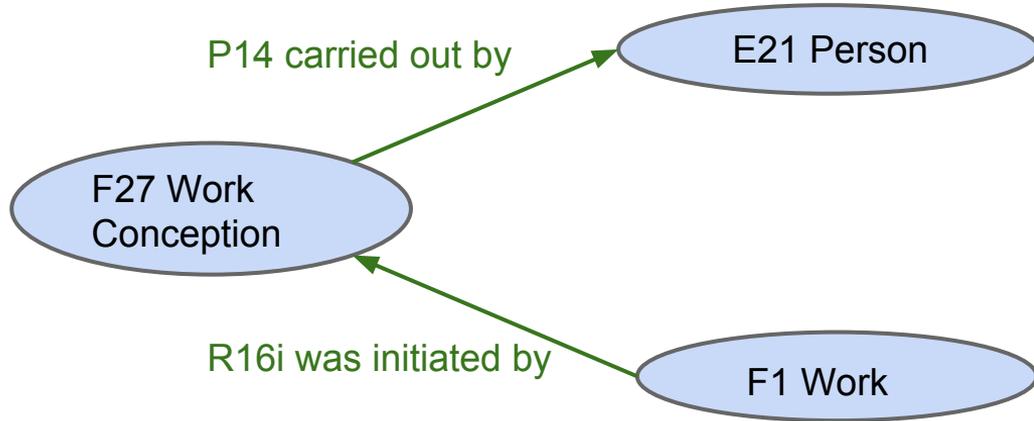
- Ontologies exprimées en RDFS ou OWL
 - RDFS : Resource Description Framework Schema
 - OWL : Web Ontology Language
- “vocabulaires RDF”
- support à l’expression des descriptions RDF

- descriptions formelles employées par des moteurs d’inférence
 - déduire nouveaux faits
 - détecter des incohérences

- CIDOC CRM et FRBRoo sont encodés en RDFS et en OWL (université d’Erlangen. cf. <http://erlangen-crm.org/>)

Graphe

L'histoire ecclésiastique est une oeuvre d'Eusèbe de Césarée.

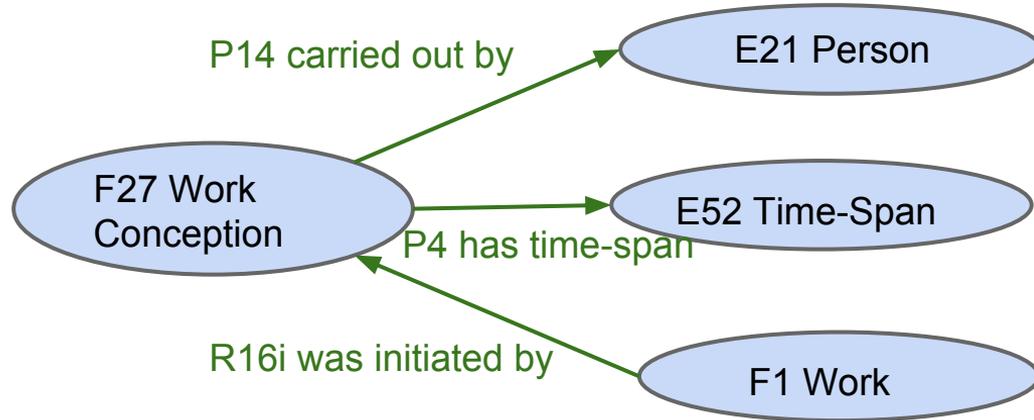


 : classe

 : propriété

Graphe

L'histoire ecclésiastique est une oeuvre d'Eusèbe de Césarée.

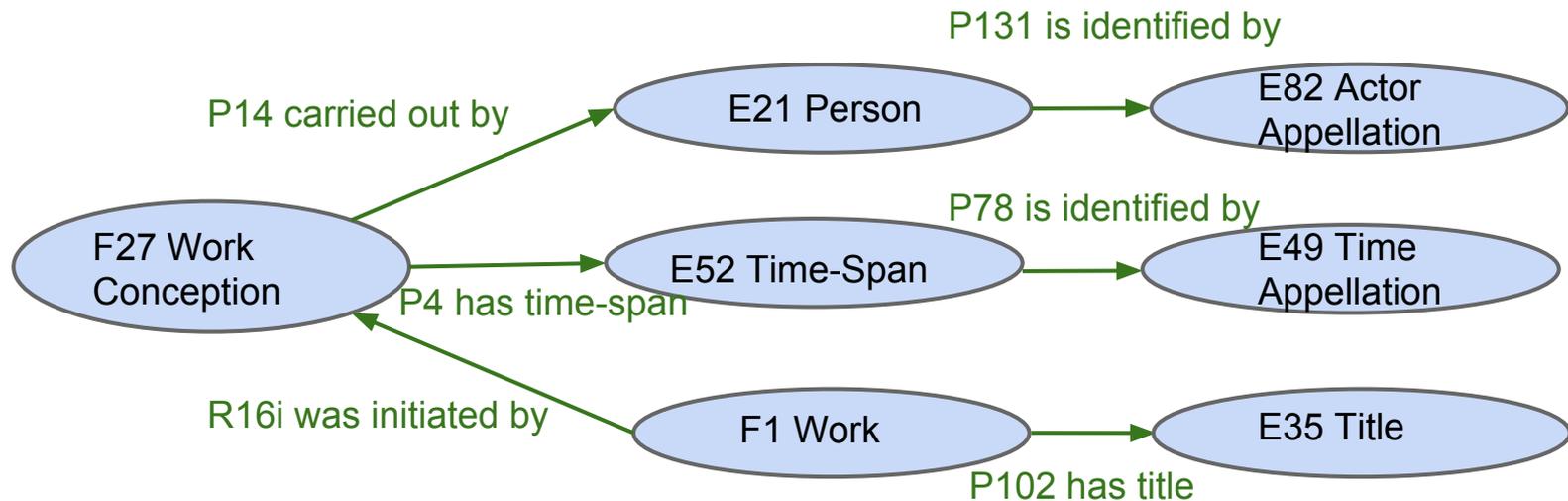


 : classe

 : propriété

Graphe

L'histoire ecclésiastique est une oeuvre d'Eusèbe de Césarée.

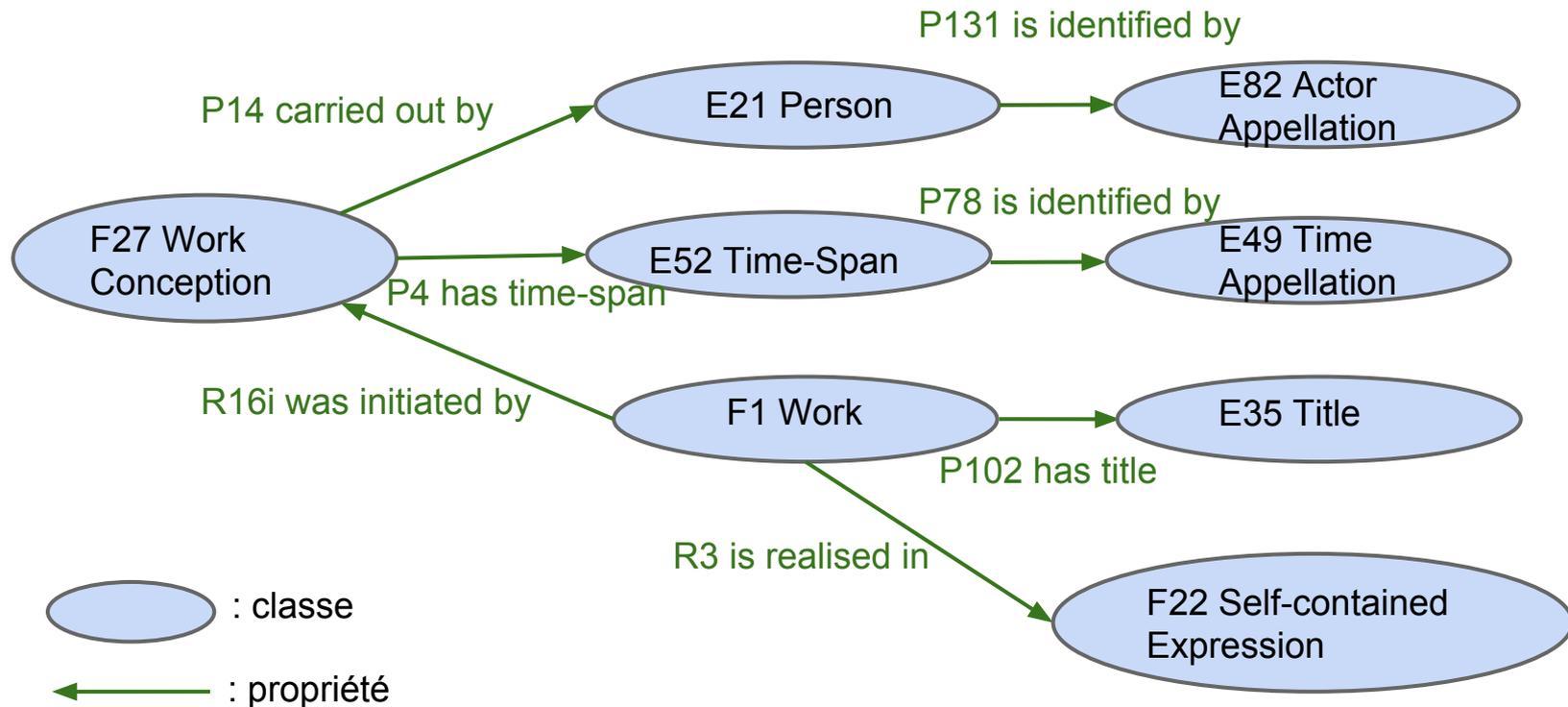


○ : classe

← : propriété

Graphe

L'histoire ecclésiastique est une oeuvre d'Eusèbe de Césarée.



Triplet RDF & ontologie

sujet	prédicat	objet
http://data.bnf.fr/ark:/12148/cb11902007r	<code>rdf:type</code>	E21 Person
Eusèbe de Césarée	est de type	E21 Personne

En RDF :

```
<http://data.bnf.fr/ark:/12148/cb11902007r> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://erlangen-crm.org/current/E21\_Person> .
```

Triplet RDF & ontologie

sujet	prédicat	objet
http://data.bnf.fr/ark:/12148/cb12012392f	<code>rdf:type</code>	F1 Work
l'oeuvre <i>Histoire ecclésiastique</i>	est de type	F1 Oeuvre

En RDF :

`<http://data.bnf.fr/ark:/12148/cb12012392f> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://erlangen-crm.org/efrbroo/F1_Work> .`

Triplet RDF & ontologie

sujet	prédicat	objet
http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f	<code>rdf:type</code>	F27 Work Conception
la conception de l'oeuvre <i>Histoire ecclésiastique</i>	est de type	F27 Conception de l'oeuvre

En RDF :

`<http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f > <http://www.w3.org/1999/02/22-rdf-syntax-ns#type > <http://erlangen-crm.org/efrbroo/F27_Work_Conception > .`

Triplet RDF & ontologie

Sujet	prédicat	objet
F1 Work	r16i was initiated by	F27 Work Conception
F27 Work Conception	P14 carried out by	E21 Person

<http://data.bnf.fr/ark:/12148/cb12012392f> <http://erlangen-crm.org/efrbroo/R16i_was_initiated_by> <<http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f>> .

<<http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f>> <http://erlangen-crm.org/current/P14_carried_out_by> <<http://data.bnf.fr/ark:/12148/cb11902007r>> .

Graphe RDF

<<http://data.bnf.fr/ark:/12148/cb12012392f>> <http://erlangen-crm.org/efrbroo/R16i_was_initiated_by>
<<http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f>> .

<<http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f>> <http://erlangen-crm.org/current/P14_carried_out_by> <<http://data.bnf.fr/ark:/12148/cb11902007r>> .

<<http://data.bnf.fr/ark:/12148/cb11902007r>> <<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>>
<http://erlangen-crm.org/current/E21_Person> .

<<http://data.bnf.fr/ark:/12148/cb12012392f>> <<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>>
<http://erlangen-crm.org/efrbroo/F1_Work> .

<<http://biblissima-condorcet.fr/ark:/43093/WorkConception/12148/cb12012392f>> <<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>> <http://erlangen-crm.org/efrbroo/F27_Work_Conception> .

SPARQL Query

Query:

```
1 PREFIX ecrm: <http://erlangen-crm.org/current/>
2 SELECT *{?s a ecrm:E39_Actor.}
```

Include inferred

Expand results over equivalent URIs

Submit

Sample queries:

Append predefined namespaces:

Requête sur l'ensemble des instances de la classe E39_Actor (<http://collection.britishmuseum.org/sparql>)

Sudan Archaeological Research Society RDF Rank

RDF Search and Explore

Source: <http://collection.britishmuseum.org/id/person-institution/195997>

Subject (6) [Predicate](#) [Object](#) [All](#)

Download in: [JSON](#) | [RDF](#) | [N3/Turtle](#) | [N-Triples](#)

Named Graph [All](#) Language [English](#) Inference [Explicit only](#)

Statements in which the resource exists as a subject.

Predicate	Object
rdf:type	ecrm:E39_Actor , skos:Concept
ecrm:P131_is_identified_by	http://collection.britishmuseum.org/id/person-institution/195997/appellation/1 , http://collection.britishmuseum.org/id/person-institution/195997/appellation/2
skos:inScheme	thesIdentifier:person-institution
skos:prefLabel	Sudan Archaeological Research Society

Exemple d'une instance de la classe E39_Actor

Dr William Yewdale Adams RDF Rank

RDF Search and Explore

Source: <http://collection.britishmuseum.org/id/person-institution/150827>

[Subject \(11\)](#) [Predicate](#) [Object](#) [All](#)

Download in: [JSON](#) | [RDF](#) | [N3/Turtle](#) | [N-Triples](#)

Named Graph Language Inference

Statements in which the resource exists as a subject.

Predicate	Object
rdf:type	ecrm:E21_Person , skos:Concept
ecrm:P3_has_note	Anthropologist and archaeologist.
bmo:PX_gender	http://collection.britishmuseum.org/id/thesauri/gender/male
bmo:PX_nationality	http://collection.britishmuseum.org/id/thesauri/nationality/American-USA

Exemple d'une instance de la classe E21_Person

BM Department Ancient Egypt and Sudan RDF Rank

RDF Search and Explore

Source: <http://collection.britishmuseum.org/id/thesauri/department/Y>

[Subject \(5\)](#) [Predicate](#) [Object](#) [All](#)

Download in: [JSON](#) | [RDF](#) | [N3/Turtle](#) | [N-Triples](#)

Named Graph Language Inference

Statements in which the resource exists as a subject.

Predicate	Object
rdf:type	ecrm:E74_Group , skos:Concept
skos:scopeNote	Department Ancient Egypt and Sudan at The British Museum
skos:inScheme	thes:department
skos:prefLabel	BM Department Ancient Egypt and Sudan

Exemple d'une instance de la classe E74_Group