

modéliser, saisir, publier, exploiter

apport des ontologies sémantiques à la conception
des systèmes d'information pour la recherche

https://github.com/Amleth/communications/tree/main/out/2025.01.17_evry_hn

PLAN

I. Problèmes récurrents lors de la conception d'un SIR & CIDOC CRM



PROBLÈMES RÉCURRENTS LORS DE LA CONCEPTION D'UN SIR & CIDOC CRM



le web sémantique, en une slide

- Promesse d'une base de données à l'échelle du Web. Le Web initial (Tim Berners Lee, 1991) était un Web de documents liés (hypertexte), le Web sémantique est un Web de données liées.



le web sémantique, en une slide

- Promesse d'une base de données à l'échelle du Web. Le Web initial (Tim Berners Lee, 1991) était un Web de documents liés (hypertexte), le Web sémantique est un Web de données liées.
- Chaque donnée est identifiée par une URL.



le web sémantique, en une slide

- Promesse d'une base de données à l'échelle du Web. Le Web initial (Tim Berners Lee, 1991) était un Web de documents liés (hypertexte), le Web sémantique est un Web de données liées.
- Chaque donnée est identifiée par une URL.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description, le RDF.



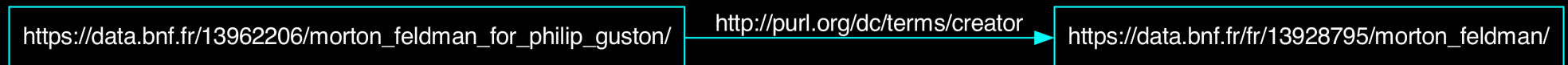
le web sémantique, en une slide

- Promesse d'une base de données à l'échelle du Web. Le Web initial (Tim Berners Lee, 1991) était un Web de documents liés (hypertexte), le Web sémantique est un Web de données liées.
- Chaque donnée est identifiée par une URL.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description, le RDF.
- La connexion de ces triplets RDF forme un graphe.



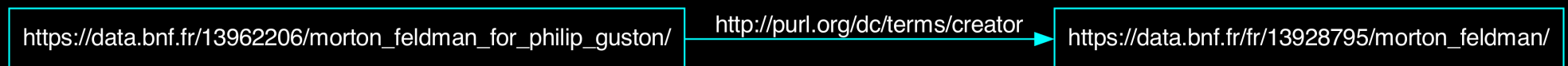
le web sémantique, en une slide

- Promesse d'une base de données à l'échelle du Web. Le Web initial (Tim Berners Lee, 1991) était un Web de documents liés (hypertexte), le Web sémantique est un Web de données liées.
- Chaque donnée est identifiée par une URL.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description, le RDF.
- La connexion de ces triplets RDF forme un graphe.
- Chaque prédicat est également identifié par une URL (sens partagé des propriétés).



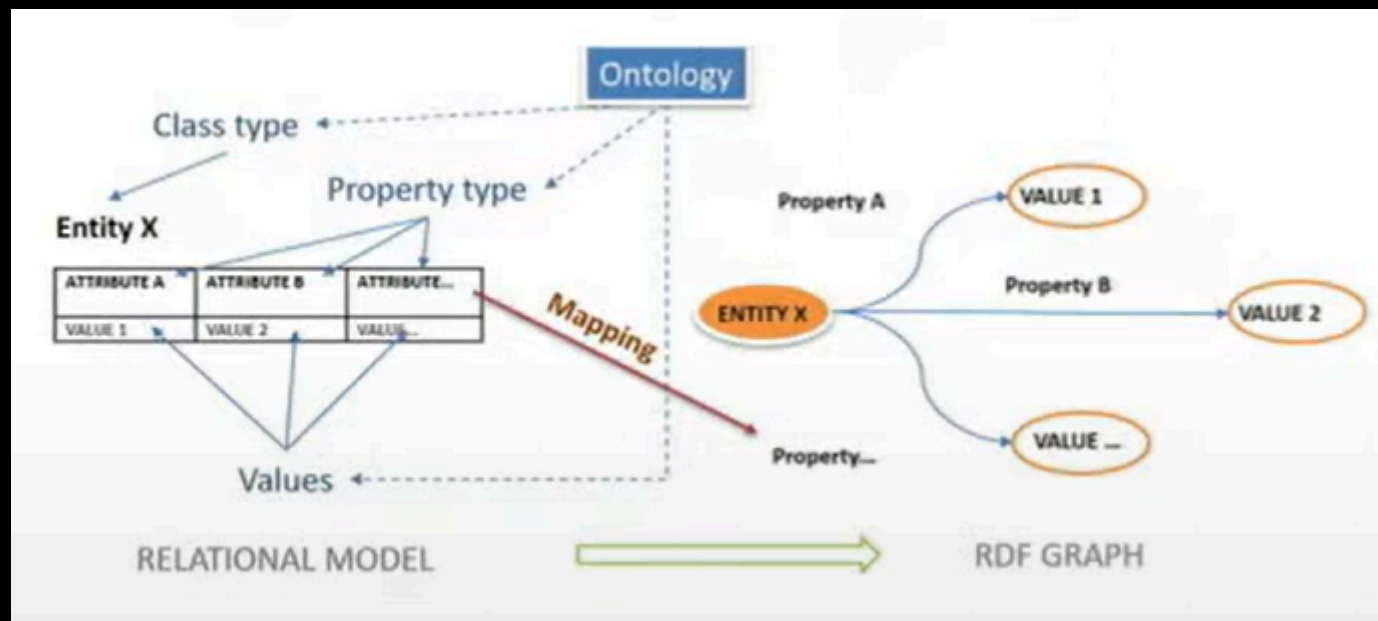
le web sémantique, en une slide

- Promesse d'une base de données à l'échelle du Web. Le Web initial (Tim Berners Lee, 1991) était un Web de documents liés (hypertexte), le Web sémantique est un Web de données liées.
- Chaque donnée est identifiée par une URL.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description, le RDF.
- La connexion de ces triplets RDF forme un graphe.
- Chaque prédicat est également identifié par une URL (sens partagé des propriétés).



- Milieu technique idéal pour des données scientifiques FAIR.

DONNÉES RELATIONNELLES VS GRAPHE RDF



Corago in LOD - Seminar by Angelo Pompilio and Paolo Bonora, Digital Humanities and Digital Knowledge, Università di Bologna, 2017.

QU'EST CE QU'UNE ONTOLOGIE ?

- Formalisation d'un modèle conceptuel pour un domaine identifié proposant des :
 - **Classes** : types d'entités peuplant le domaine, possiblement organisées selon des relations d'héritage (spécificité). On appelle **individu** une ressource qui est du type d'une classe.
 - **Propriétés** : aspects, caractéristiques, attributs possibles de ces classes, qui peuvent soit pointer vers une valeur, soit vers un individu.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données RDF (les individus identifiés par des URL seront des sujets ou des objets, les propriétés des classes seront des prédicats).
- Vous connaissez peut-être déjà une ontologie : SKOS (pour construire des thésauri).



III. INTRODUCTION À L'ONTOLOGIE CIDOC-CRM



LE CIDOC-CRM EN BREF

- Le CIDOC-CRM est une ontologie qui documente le patrimoine matériel et immatériel ainsi que les processus de production de connaissances à son propos.
- <https://www.cidoc-crm.org/>
- Venant du monde des musées, elle est désormais utilisée dans tous les domaines des HN.
- Elle est extrêmement abstraite et générique.
- Ontologie centrée événement (nous y reviendrons dans les exemples...)
- Classes et propriétés : https://cidoc-crm.org/html/cidoc_crm_v7.1.2.html



Martin Doerr
2/16/2020



Opinion : En dépit du nombre de classes centrées sur les usages de musées, le CIDOC-CRM propose des classes génériques permettant de rendre compte de l'ensemble des problématiques de modélisation de la structure et du contenu des sources, ainsi que des processus analytiques qui les prennent pour cible.



LE TEMPS DANS LE CRM

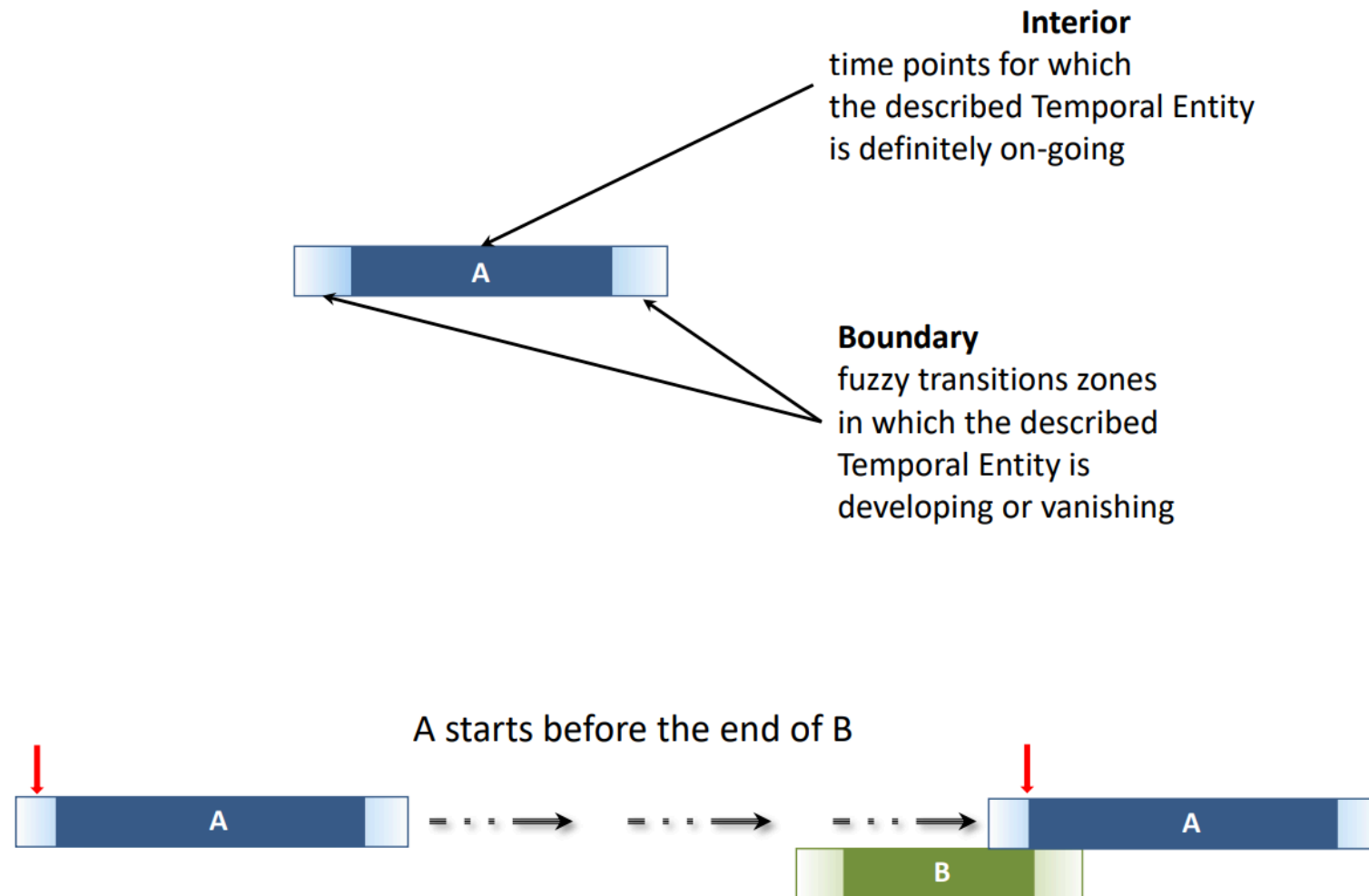


Figure 7: Explanation of Interior and Boundary and an Example of Use from P174 starts before the end of (ends after the start of).

DE QUELQUES ONTOLOGIES FILLES BIEN UTILES

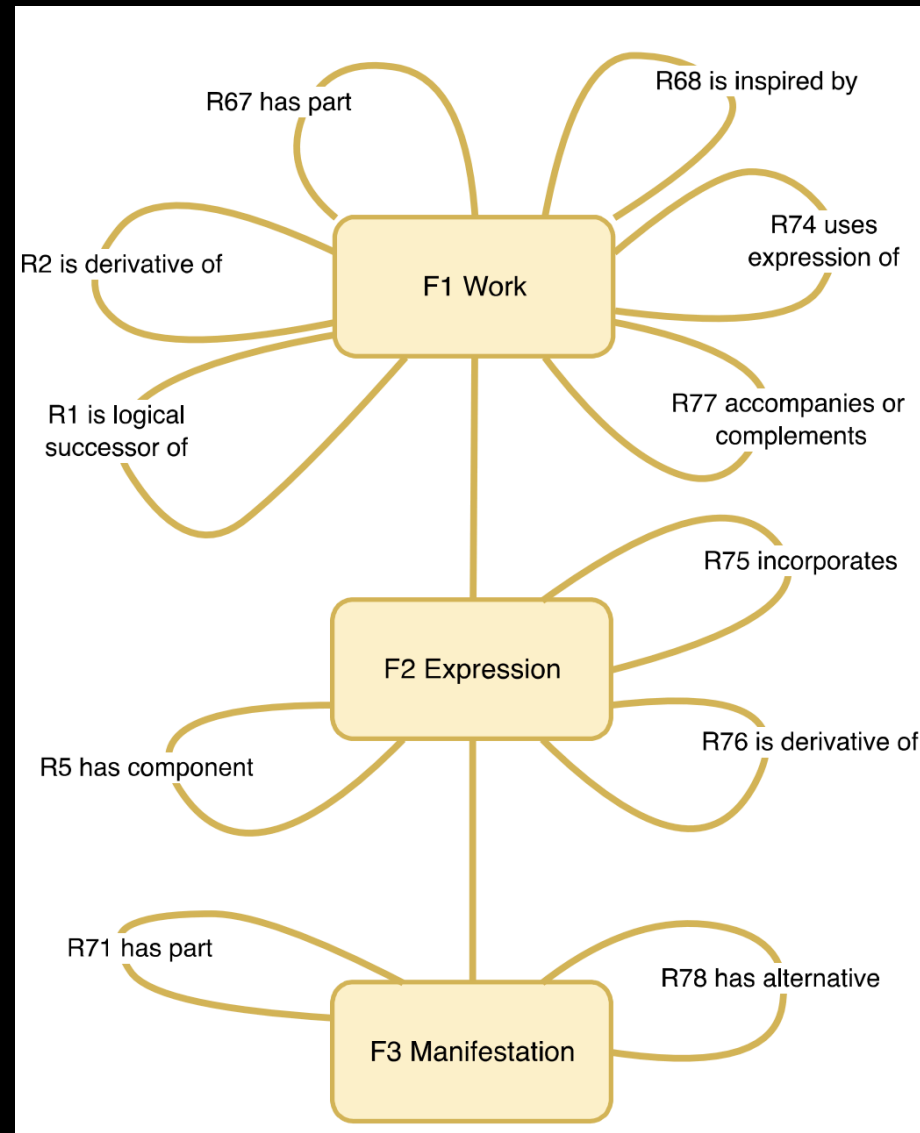


lrmoo

- Convergence du monde des musées (CIDOC-CRM) et du monde des bibliothèques (IFLA LRM, anciennement FRBR) : l'ontologie LRMoo spécialise certaines classes du CIDOC-CRM (qui sert donc d'ontologie de base).
- Quatre entités fondamentales — Work, Expression, Manifestation, Item (WEMI) — représentant les quatre modes d'existence des sources (conceptuel, sémiotique, éditorial, matériel).



lrmoo : des relations riches entre les entités



doremus

- Des vocabulaires utiles pour la musique : data.doremus.org/vocabularies + github.com/DOREMUS-ANR/knowledge-base/tree/master/vocabularies
- Une ontologie sans équivalent pour la modélisation des effectifs prévus dans l'œuvre, des concerts, et des effectifs réalisés durant les performances : <https://data.doremus.org/ontology/>
- Un parti pris (désormais techniquement obsolète) sur la manière dont on peut modéliser une situation de composition : data.doremus.org/ontology
- Un énorme dataset (BNF + Radio France + Philharmonie) d'œuvres, partitions, concerts... data.doremus.org



UNE APPLICATION WEB POUR EXPLORER DES DONNÉES CRM

Motivation :

Recourir à un unique modèles dans les différents projets permet de ne concevoir, développer et maintenir qu'une unique application pour présenter et exploiter les données.

Contexte :

Pas ou peu d'apport financier. Programme de recherche SHERLOCK à l'IReMus (Comment et pourquoi modéliser les données musicologiques avec le CIDOC CRM ? Comment les publier et les manipuler ?).



objectifs fonctionnels de la « sherlock app » (i/ii)

- Une interface de navigation hypertexte générique dans la totalité des graphes RDF d'un Triplestore accessible via un SPARQL Endpoint.



objectifs fonctionnels de la « sherlock app » (i/ii)

- Une interface de navigation hypertexte générique dans la totalité des graphes RDF d'un Triplestore accessible via un SPARQL Endpoint.
- L'utilisateur devrait avoir le sentiment de naviguer dans des fiches, dont la structure serait claire et où les métadonnées y seraient affichées clairement, sans être exposé à la technicité inhérente aux triplets RDF et aux noms abstraits des classes et des propriétés des ontologies convoquées...



objectifs fonctionnels de la « sherlock app » (i/ii)

- Une interface de navigation hypertexte générique dans la totalité des graphes RDF d'un Triplestore accessible via un SPARQL Endpoint.
- L'utilisateur devrait avoir le sentiment de naviguer dans des fiches, dont la structure serait claire et où les métadonnées y seraient affichées clairement, sans être exposé à la technicité inhérente aux triplets RDF et aux noms abstraits des classes et des propriétés des ontologies convoquées...
- ... mais la teneur des sujets/prédicats/objets RDF devrait toujours être clairement indiquée, pour raisons pédagogique et technique.



objectifs fonctionnels de la « sherlock app » (i/ii)

- Une interface de navigation hypertexte générique dans la totalité des graphes RDF d'un Triplestore accessible via un SPARQL Endpoint.
- L'utilisateur devrait avoir le sentiment de naviguer dans des fiches, dont la structure serait claire et où les métadonnées y seraient affichées clairement, sans être exposé à la technicité inhérente aux triplets RDF et aux noms abstraits des classes et des propriétés des ontologies convoquées...
- ... mais la teneur des sujets/prédicats/objets RDF devrait toujours être clairement indiquée, pour raisons pédagogique et technique.
- Exploitation des *patterns* spécifiques du CRM ou de LRMoo pour proposer des interfaces spécifiques ([lrmoo:F18](#)).
 - Triplets décrivant l'identité d'une ressource
 - Dates
 - Place dans une structure (sociale, bibliothécaire, logique, matérielle...)
 - Annotations (qui ? qui ? quand ? pourquoi ?)
 - Annotations (multiplicité des points de vue)



SHERLOCK

explorateur de données rdf/cidoc crm

ressource consultée

<http://data-iremum.huma-num.fr/id/82a7b7d8-394a-4e47-a83e-ce34b5640b68>



identité de la ressource



a pour titre (crm:P102_has_title)	[Madrigaux sur la premiere Enigme du Mois de May, dont le Mot estoit la Lanterne]
est identifiée par (crm:P1_is_identified_by)	/mercure-galant/1681-06e_363
a pour type (crm:P2_has_type)	Article est un crm:E55_Type
a pour type (crm:P2_has_type)	Fichier TEI @fr est un crm:E55_Type
a pour classe (rdf:type)	lrmoo:F2_Expression



objectifs fonctionnels de la « sherlock app » (ii/ii)

- Proposer l'éventuel contenu associé à la ressource en cours de consultation (rendu HTML d'un fragment TEI, image, fichier MEI s'ouvrant dans l'interface d'annotation collaborative Tonalities...)
- Moteur de recherche plein texte dans l'ensemble des valeurs littérales textuelles.
- Moteur de recherche par concepts issus de thésaurus.