

Journée d'études organisée par l'UFR LAM, IDHES Evry, les départements d'Histoire, de LEA et de Musicologie

modéliser, saisir, explorer

retour sur la mise en œuvre du cidoc crm dans des bdd en musicologie

https://github.com/Amleth/communications/blob/main/out/2025.01.17_evry_hn/index.pdf

thomas.bottiniacnrs.fr

Institut de Recherche en Musicologie - IReMus, UMR 8223 CNRS

Consortium Musica2 IR* Huma-Num

plan

- 1. Modéliser
- 2. Saisir
- 3. Explorer



modéliser

Le CIDOC CRM au service de la musicologie



La question : comment faire tenir les données de la recherche dans le temps (!) ?
 technologies + traditions interprétatives + moyens de dev limités



- La question : comment faire tenir les données de la recherche dans le temps (!) ?
 technologies + traditions interprétatives + moyens de dev limités
- 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, chacune étant identifiée par une URI.



- La question : comment faire tenir les données de la recherche dans le temps (!) ?
 technologies + traditions interprétatives + moyens de dev limités
- 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, chacune étant identifiée par une URI.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description qui est le RDF.



- La question : comment faire tenir les données de la recherche dans le temps (!) ?
 technologies + traditions interprétatives + moyens de dev limités
- 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, chacune étant identifiée par une URI.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description qui est le RDF.
- La connexion de ces triplets RDF forme un graphe.



- La question : comment faire tenir les données de la recherche dans le temps (!) ? technologies + traditions interprétatives + moyens de dev limités
- 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, chacune étant identifiée par une URI.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description qui est le RDF.
- La connexion de ces triplets RDF forme un graphe.
- Chaque prédicat est également identifié par une URL.

https://data.bnf.fr/13962206/morton feldman for philip guston/

http://purl.org/dc/terms/creator https://data.bnf.fr/fr/13928795/morton_feldman/



- La question : comment faire tenir les données de la recherche dans le temps (!) ?
 technologies + traditions interprétatives + moyens de dev limités
- 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, chacune étant identifiée par une URI.
- Toute information s'exprime sous la forme d'un triplet (sujet/prédicat/objet) dans un langage de description qui est le RDF.
- La connexion de ces triplets RDF forme un graphe.
- Chaque prédicat est également identifié par une URL.

```
https://data.bnf.fr/13962206/morton_feldman_for_philip_guston/ http://purl.org/dc/terms/creator https://data.bnf.fr/fr/13928795/morton_feldman/
```

• C'est le milieu technique idéal pour des données FAIR, pour l'expression et la diffusion des données de la recherche (publication + nouveaux usages).



• Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.

- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).



- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).
- Permet de capitaliser des connaissances de modélisation d'un projet à l'autre (démarche KM).



- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).
- Permet de capitaliser des connaissances de modélisation d'un projet à l'autre (démarche KM).
- Le CIDOC CRM en bref :
 - Ontologie qui documente le patrimoine matériel et immatériel ainsi que les processus de production de connaissances à son propos (sources, connaissances, faits sociaux, concepts, objets matériels, objets sémiotiques dénotés ou connotés, contexte de production des connaissances, etc.).



- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).
- Permet de capitaliser des connaissances de modélisation d'un projet à l'autre (démarche KM).
- Le CIDOC CRM en bref:
 - Ontologie qui documente le patrimoine matériel et immatériel ainsi que les processus de production de connaissances à son propos (sources, connaissances, faits sociaux, concepts, objets matériels, objets sémiotiques dénotés ou connotés, contexte de production des connaissances, etc.).
 - Communauté large et ancienne. Venant du monde des musées, elle est désormais utilisée dans tous les domaines des HN.



- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).
- Permet de capitaliser des connaissances de modélisation d'un projet à l'autre (démarche KM).
- Le CIDOC CRM en bref :
 - Ontologie qui documente le patrimoine matériel et immatériel ainsi que les processus de production de connaissances à son propos (sources, connaissances, faits sociaux, concepts, objets matériels, objets sémiotiques dénotés ou connotés, contexte de production des connaissances, etc.).
 - Communauté large et ancienne. Venant du monde des musées, elle est désormais utilisée dans tous les domaines des HN.
 - Elle est extrêmement abstraite et générique.



- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).
- Permet de capitaliser des connaissances de modélisation d'un projet à l'autre (démarche KM).
- Le CIDOC CRM en bref :
 - Ontologie qui documente le patrimoine matériel et immatériel ainsi que les processus de production de connaissances à son propos (sources, connaissances, faits sociaux, concepts, objets matériels, objets sémiotiques dénotés ou connotés, contexte de production des connaissances, etc.).
 - Communauté large et ancienne. Venant du monde des musées, elle est désormais utilisée dans tous les domaines des HN.
 - Elle est extrêmement abstraite et générique.
 - Ontologie centrée événement.



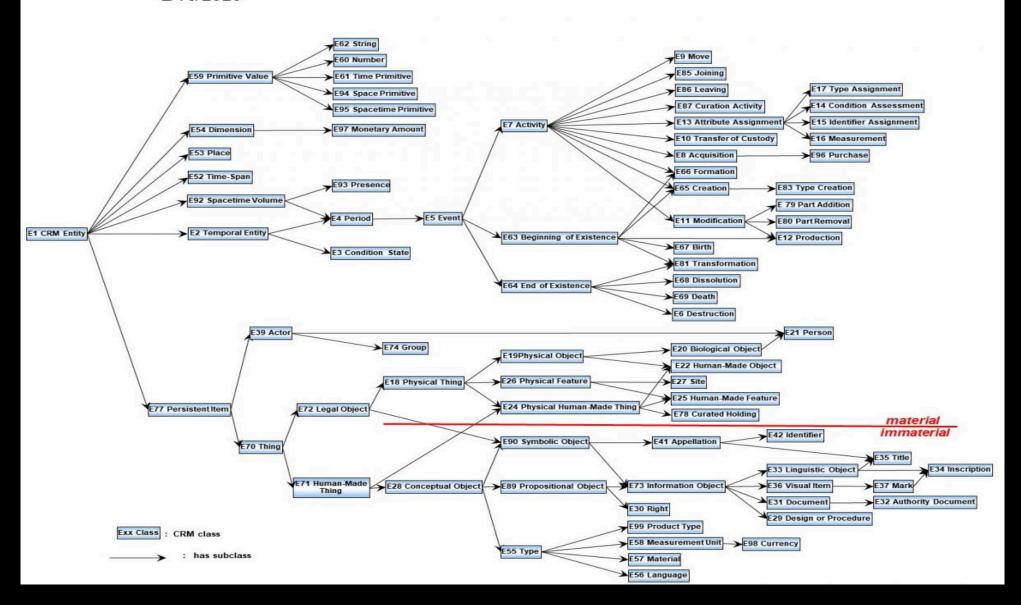
- Formalisation d'un modèle conceptuel pour un domaine donné, contenant des classes et des propriétés.
- Utiliser les classes et les propriétés d'une ontologie confère ainsi une sémantique partagée aux données (partage d'individus, partage de prédicats).
- Permet de capitaliser des connaissances de modélisation d'un projet à l'autre (démarche KM).
- Le CIDOC CRM en bref :
 - Ontologie qui documente le patrimoine matériel et immatériel ainsi que les processus de production de connaissances à son propos (sources, connaissances, faits sociaux, concepts, objets matériels, objets sémiotiques dénotés ou connotés, contexte de production des connaissances, etc.).
 - Communauté large et ancienne. Venant du monde des musées, elle est désormais utilisée dans tous les domaines des HN.
 - Elle est extrêmement abstraite et générique.
 - Ontologie centrée événement.
 - Informations bibliographiques avec LRMoo (œuvres, expressions, manifestation, item).



hiérarchie des classes crm

CRM Class Hierarchy

Martin Doerr 2/16/2020





cas d'usage à l'iremus

- Référentiel des personnes, lieux et institution de l'Ancien Régime (charges).
- Édition critique du Mercure Galant (1672-1724), structuration logique et matérielle des corpus original et TEI, indexation sémantique des articles avec des thésaurus et référentiels.
- Indexation sémantique (avec des thésaurus) et description analytique de collections iconographiques.
- Annotation collaborative de partitions MEI avec structuration logique des annotations (cadences et leur structure interne) et articulation fine à la partition.
- Divers entreprises de catalogage de sources anciennes.
- Modèle Antony: sauvegarde et partage du patrimoine musical utilisant des technologies numériques.



saisir

Ça se complique



 Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).

- Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).
- Le CRM est expressif, mais :

- Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).
- Le CRM est expressif, mais :
 - Il existe parfois plusieurs manières de modéliser une situation avec les classes de base.

- Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).
- Le CRM est expressif, mais :
 - Il existe parfois plusieurs manières de modéliser une situation avec les classes de base.
 - Ses *patterns* fondamentaux (pour nommer, type, dater, annoter...) induisent beaucoup de des sous-entités.



- Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).
- Le CRM est expressif, mais :
 - Il existe parfois plusieurs manières de modéliser une situation avec les classes de base.
 - Ses *patterns* fondamentaux (pour nommer, type, dater, annoter...) induisent beaucoup de des sous-entités.
 - Son caractère abstrait et générique fait écran avec la compréhension naturelle que le chercheur pourrait avoir de ses données.



- Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).
- Le CRM est expressif, mais :
 - Il existe parfois plusieurs manières de modéliser une situation avec les classes de base.
 - Ses *patterns* fondamentaux (pour nommer, type, dater, annoter...) induisent beaucoup de des sous-entités.
 - Son caractère abstrait et générique fait écran avec la compréhension naturelle que le chercheur pourrait avoir de ses données.
- En conséquence, une interface d'édition générique de données CRM n'a pas de sens, car chaque collectif construit sa manière d'utiliser l'ontologie.



- Un graphe de données ouvert est plus difficile à éditer que des données relationnelles (données tabulaires s'éditant naturellement avec des formulaires contraints).
- Le CRM est expressif, mais :
 - Il existe parfois plusieurs manières de modéliser une situation avec les classes de base.
 - Ses *patterns* fondamentaux (pour nommer, type, dater, annoter...) induisent beaucoup de des sous-entités.
 - Son caractère abstrait et générique fait écran avec la compréhension naturelle que le chercheur pourrait avoir de ses données.
- En conséquence, une interface d'édition générique de données CRM n'a pas de sens, car chaque collectif construit sa manière d'utiliser l'ontologie.
- Je préconise le recours à un outil de saisie de données existant, libre et ergonomique. Par exemple, un candidat de la constellation « No-code ».



nocodb





du relationnel au rdf

• Le modèle relationnel doit être créé pour répondre aux attendus ergonomiques du projet. Sa structure doit permettre de générer des données CIDOC CRM par la suite, mais il n'est qu'un modèle de saisie. Il représente la manière dont un collectif se saisit du CRM dans un contexte précis (classes et propriétés utilisées + idiomes de modélisation).

du relationnel au rdf

- Le modèle relationnel doit être créé pour répondre aux attendus ergonomiques du projet. Sa structure doit permettre de générer des données CIDOC CRM par la suite, mais il n'est qu'un modèle de saisie. Il représente la manière dont un collectif se saisit du CRM dans un contexte précis (classes et propriétés utilisées + idiomes de modélisation).
- Du code doit être écrit pour récupérer les données via l'API offerte par le système et les convertir en données RDF modélisées avec le CIDOC CRM (un couple efficace pour ce genre de tâches : python/rdflib).



du relationnel au rdf

- Le modèle relationnel doit être créé pour répondre aux attendus ergonomiques du projet. Sa structure doit permettre de générer des données CIDOC CRM par la suite, mais il n'est qu'un modèle de saisie. Il représente la manière dont un collectif se saisit du CRM dans un contexte précis (classes et propriétés utilisées + idiomes de modélisation).
- Du code doit être écrit pour récupérer les données via l'API offerte par le système et les convertir en données RDF modélisées avec le CIDOC CRM (un couple efficace pour ce genre de tâches : python/rdflib).
- Tous les moyens sont bons pour obtenir des données en CRM.



explorer

De l'intérêt d'exposer des données RDF/CRM avec SPARQL

SHERLOCK

explorateur de données rdf/cidoc crm



contexte

Contexte:

- Programme de recherche SHERLOCK à l'IReMus :
 - « Comment et pourquoi modéliser les données musicologiques avec le CIDOC CRM ?»
 - « Comment faire interagir les données sémantiques et les sources ? »
 - « Comment publier et manipuler les données sémantiques ? »
- Pas ou peu d'apport financier.
- Développeur (presque) unique.
- Recourir à un modèle unique dans les différents projets permet de ne concevoir, développer et maintenir qu'une unique application pour présenter et exploiter les données.

Technologies :

- Front : TypeScript, React, Next UI, Tailwind CSS
- Back : Apache Jena Fuseki



objectifs fonctionnels (1/2)

• Une interface de navigation hypertexte générique portant sur la totalité des graphes RDF d'un *triplestore* accessible via un SPARQL *endpoint*.

objectifs fonctionnels (1/2)

- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, sans être exposé à la technicité inhérente aux triplets RDF et aux noms abstraits des classes et des propriétés des ontologies convoquées...



objectifs fonctionnels (1/2)

- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, 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. Toutes les requêtes SPARQL utilisées devraient être exposées.



- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, 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. Toutes les requêtes SPARQL utilisées devraient être exposées.
- Exploitation des patterns spécifiques du CRM ou de LRMoo pour proposer des interfaces spécifiques :
 - Triplets décrivant l'<u>identité d'une ressource</u>



- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, 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. Toutes les requêtes SPARQL utilisées devraient être exposées.
- Exploitation des patterns spécifiques du CRM ou de LRMoo pour proposer des interfaces spécifiques :
 - Triplets décrivant l'<u>identité d'une ressource</u>
 - Dates



- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, 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. Toutes les requêtes SPARQL utilisées devraient être exposées.
- Exploitation des patterns spécifiques du CRM ou de LRMoo pour proposer des interfaces spécifiques :
 - Triplets décrivant l'<u>identité d'une ressource</u>
 - Dates
 - Place dans une structure : sociale, bibliographique (<u>lrmoo:F18</u>), logique, matérielle...



- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, 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. Toutes les requêtes SPARQL utilisées devraient être exposées.
- Exploitation des patterns spécifiques du CRM ou de LRMoo pour proposer des interfaces spécifiques :
 - Triplets décrivant l'<u>identité d'une ressource</u>
 - Dates
 - Place dans une structure : sociale, bibliographique (Irmoo:F18), logique, matérielle...
 - Annotations (qui ? qui ? quand ? pourquoi ?)



- Une interface de navigation hypertexte générique portant sur 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 et l'affichage des métadonnées seraient clairs, 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. Toutes les requêtes SPARQL utilisées devraient être exposées.
- Exploitation des patterns spécifiques du CRM ou de LRMoo pour proposer des interfaces spécifiques :
 - Triplets décrivant l'<u>identité d'une ressource</u>
 - Dates
 - Place dans une structure : sociale, bibliographique (<u>lrmoo:F18</u>), logique, matérielle...
 - Annotations (qui ? qui ? quand ? pourquoi ?)
 - Annotations (multiplicité des points de vue)



identité d'une ressource

SHERLOCK

explorateur de données rdf/cidoc crm

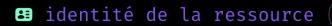
ressource consultée

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











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	/mercure-galant/1681-06e_363
<pre>(crm:P1_is_identified_by)</pre>	est un crm: E42_Identifier: « Identifiant projet »
a pour type (crm: P2_has_type)	Article 🌐
	est un crm: E55_Type
<pre>a pour type (crm:P2_has_type)</pre>	Fichier TEI @fr 🌐
	est un crm: E55_Type
a pour classe (rdf:type)	lrmoo:F2_Expression

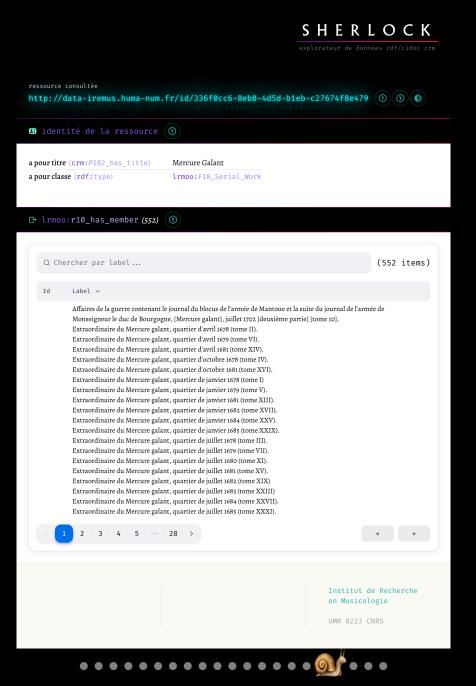


affichage des ressources liées

→ ressources pointées lrmoo:R3_is_realised_in iremus-data:5bc605b2-fe9f-431a-84e3-847ac7935d4a est identifiée par (crm:P1_is_identified_by) /mercure-galant/1681-12e 🌐 est un crm: E42_Identifier : « Identifiant projet » Contenu TEI @fr 🌐 a pour type (crm:P2_has_type) est un crm: E55 Type a pour classe (rdf:type) lrmoo:F2_Expression lrmoo:R3_is_realised_in iremus-data:912c46b6-3ece-47e8-8d58-5d665e6538e1 Livraison (a pour type (crm:P2_has_type) est un crm: E55_Type Texte original 🌐 a pour type (crm:P2_has_type) est un crm: E55 Type a pour classe (rdf:type) lrmoo:F2_Expression



liste des articles d'un périodique



Proposer l'éventuel contenu associé à la ressource en cours de consultation (<u>rendu HTML d'un fragment TEI</u>, image IIIF, fichier MEI s'ouvrant dans l'interface d'annotation collaborative <u>Tonalities...</u>).



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



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



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



rendu html d'un fragment tei associé à la ressource

ressource consultée

http://data-iremus.huma-num.fr/id/2d300dd8-7e10-4de1-ac12-95602a0e78fe

① ① ①

🖪 identité de la ressource 🕜



√ rendu du contenu tei ⊕

[Madrigaux sur la premiere Enigme du Mois de May, dont le Mot estoit la Lanterne]

Extraordinaire du Mercure galant, quartier d'avril 1681 (tome XIV), p. 363-366. Je viens aux deux en Vers que je vous ay envoyées dans ma Lettre du Mois de May. Le vray Mot de la premiere, qui estoit une Lanterne, a donné lieu à ces cinq Madrigaux.

T.

Dans les tristes horreurs d'une profonde nuit,

Qui couvroit de ses sombres voiles

Le Ciel, la Lune, & les Etoiles,

Je marchois pas à pas, & sans faire de bruit.

Mercure qui me vit sortir d'une Taverne;

Cher Amy, me dit-il, où vas-tu? viens chez moy.

.

conclusion

Démarche d'ingénierie : rendre techniquement indépendants les phases de modélisation, saisie et exploration (il manque le quatrième volet : l'exploitation).

Le CRM comme ciment permettant cette indépendance.



