



Éthique liée à la gestion des données de recherche

Atelier d'initiation

Dominic Boisvert
Responsable de la gouvernance des données d'entreprise

17 juin 2022

Plan de l'atelier

1. Rappel du thème de l'atelier
2. Les données et l'éthique
3. Les principes FAIR

Qui suis-je

Dominic Boisvert

1. Formation en histoire (UdS) et en archivistique (UdeM)
2. Associé-principal pendant 22 ans chez HB archivistes, s.e.n.c.
3. Responsable de la gouvernance des données opérationnelles, Université de Montréal, depuis octobre 2021
 - Responsable de la qualité et des mécanismes de partages des données
4. Chargé de cours à l'EBSI, depuis 2013 :
 - Aspects juridiques de l'information et des archives
 - Système d'information organisationnels
 - *Archives numériques, classification, évaluation*

Université 
de Montréal
et du monde.



Votre rôle

L'intendance des données de recherche

« Les données de recherche et les productions savantes qui en découlent sont des actifs de recherche précieux. Bien gérés, ces actifs peuvent avoir, grâce au partage et à la réutilisation, un fort impact sur l'avancement du savoir et les capacités d'innovation. »



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada

Financé par le
gouvernement
du Canada

Canada 

L'intendance des données

1. Le profil de l'intendant.e

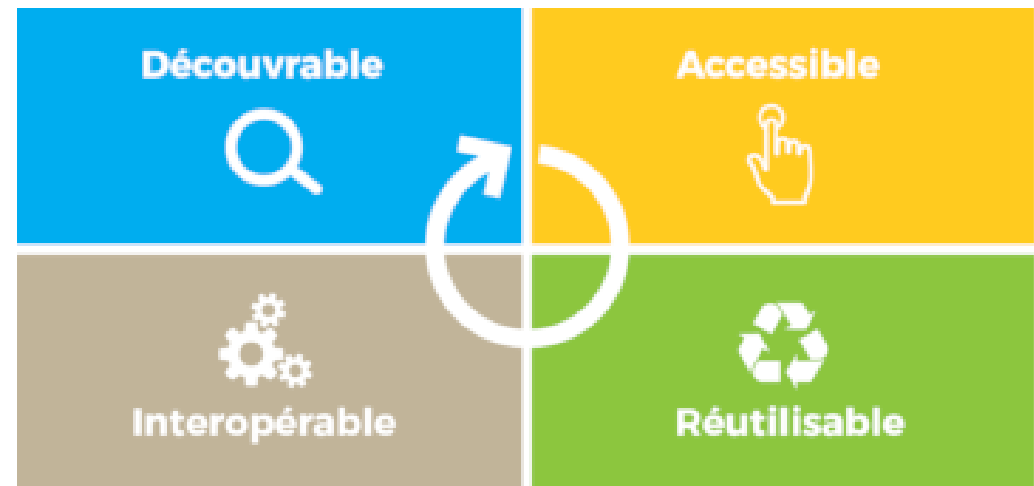
1. Qualité des données
 - Gère, contrôle et surveille la qualité des données
2. Conformité avec la politique
 - Gère et protège les données
3. Processus
 - Définit les règles de validation des données

2. Les ressources

1. Stratégie institutionnelle de gestion des données de recherche
 - Les politiques et ressources de votre institution
2. Plan de gestion des données de recherche
 - Collecte, métadonnées, stockage, conservation, partage et réutilisation, responsabilité

FAIR

Des principes qui guident vos réflexions en gestion des données.



Fair principles. (s. d.). GO FAIR. Consulté 15 juin 2022, à l'adresse <https://www.go-fair.org/fair-principles/>

Découvrable

Être découvrable signifie avoir un identifiant persistant unique (ex. DOI ou URL), être décrit par de riches métadonnées **respectant des normes reconnues** et enfin, assurer une présence dans les portails de découverte de données autant locaux que nationaux.

- F1. Les (méta)données se voient attribuer un identifiant global unique et persistant
- F2. Les données sont décrites avec des métadonnées riches (définies par R1 ci-dessous)
- F3. Les métadonnées incluent clairement et explicitement l'identifiant des données qu'elles décrivent
- F4. Les (méta)données sont enregistrées ou indexées dans une ressource consultable

L'identifiant persistant unique

Pour vous

- ORCID
- ResearcherID
- Scopus author ID



Pour les données et les publications

- Handle
 - <http://hdl.handle.net/1866/23102>
- DOI
 - <https://doi.org/10.1108/09565690210454770>
 - 10.1108/09565690210454770



Métadonnées riches

Schémas

- Schéma général : Dublin Core
- Humanités : Text Encoding Initiative (TEI)

```
<msDesc>
  <msIdentifier>
    <settlement>Oxford</settlement>
    <repository>Bodleian Library</repository>
    <idno>MS. Add. A. 61</idno>
    <altIdentifier type="SC">
      <idno>28843</idno>
    </altIdentifier>
  </msIdentifier>
  <p>In Latin, on parchment: written in more than one hand of the 13th cent. in England: 7¼ x 5½ in., i + 55 leaves, in double columns: with a few coloured capitals.</p>
  <p>'Hic incipit Bruitus Anglie,' the De origine et gestis Regum Angliae of Geoffrey of Monmouth (Galfridus Monumetensis: beg. 'Cum mecum multa & de multis.'</p>
  <p>On fol. 54v very faint is 'Iste liber est fratris guillelmi de buria de ... Roberti ordinis fratrum Pred[icatorum],' 14th cent. (?): 'hanauilla' is written at the foot of the page (15th cent.). Bought from the rev. W. D. Macray on March 17, 1863, for £1 10s.</p>
</msDesc>
```

Cadres

- Data Catalog Vocabulary (DCAT)
- Exemple de la licence

```
<div property="dct:license"
  resource="{license URL}">
  <a href="{license URL}">
    <span property="dct:title">{license name}</span>
  </a>
</div>
```

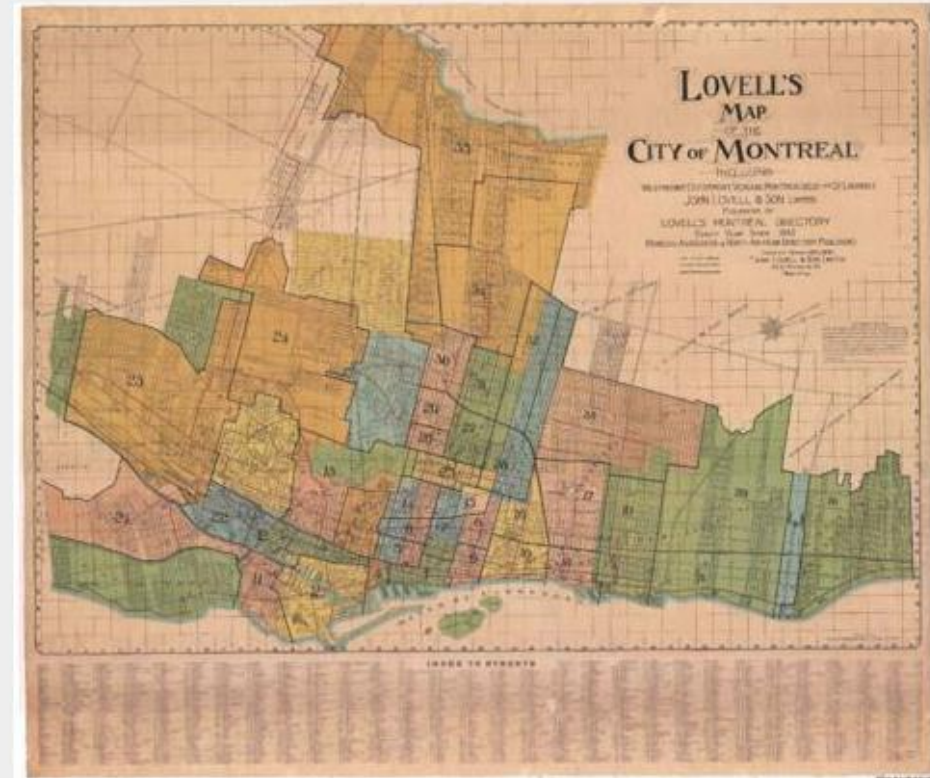
- [ISA tools | Standardizing metadata for scientific experiments \(isa-tools.org\)](https://isa-tools.org/)

Suite - TEI

```

<msDesc>
  <msIdentifier>
    <settlement>Montréal</settlement>
    <repository>Ville de Montréal. Section des
archives</repository>
    <idno>CA M001 VM066-5-P148</idno>
  </msIdentifier>
  <msContents>
    <p>920-3: Lovell's map of the city of Montreal including
Westmount, Outremont, Verdun, Montreal-West and St-Laurent / John
Lovell & Son Limited publishers of Lovell's Montreal directory every
year since 1842. - 1920</p>
  </msContents>
  <physDesc>
    <p>
      <material function="support">Parchemin</material>: plan :
coul., entoilé ; 101 x 119 cm, rouleau 101 x 8 cm de diam..</p>
  </physDesc>
  <history>
    <p>Ce plan présente les quartiers de la cité de Montréal et les
municipalités avoisinantes (avec le port au sud, le quartier Saint-
Gabriel à l'ouest, le quartier Ahuntsic-Bordeaux au nord et le
quartier Mercier-Maisonneuve à l'est). Les couleurs délimitent.</p>
  </history>
</msDesc>

```



Titre propre	1920-3: Lovell's map of the city of Montreal including Westmount, Outremont, Verdun, Montreal-West and St-Laurent / John Lovell & Son Limited publishers of Lovell's Montreal directory every year since 1842. - 1920
Niveau de description	Pièce
Dépôt	Ville de Montréal. Section des archives
Cote	CA M001 VM066-5-P148
Date(s)	• 1920 (Production)
Description matérielle	1 plan : coul., entoilé ; 101 x 119 cm, rouleau 101 x 8 cm de diam.

Accessible

Être accessible signifie que les données peuvent être facilement obtenues autant par des humains que par les machines, selon des protocoles définis et idéalement standardisés. Les conditions d'accès doivent être clairement établies (licence, droits de réutilisation, etc.). Même quand les données sont inaccessibles, les métadonnées le demeurent accessibles afin que le jeu de données reste découvrable. L'auteur du jeu de données pourra alors être contacté pour obtenir davantage d'information, ou une demande d'accès aux données (par exemple sous certaines conditions) pourra être effectuée.

- A1. Les (méta)données sont récupérables par leur identifiant à l'aide d'un protocole de communication standardisé
 - A1.1 Le protocole est ouvert, gratuit et universellement implémentable
 - A1.2 Le protocole permet une procédure d'authentification et d'autorisation, si nécessaire
- A2. Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles

Interopérable

L'interopérabilité rend possible l'intégration de différents jeux de données sur un portail unique, par exemple au sein de la même application web. Cela peut être assuré de différentes manières, incluant : utilisation d'un langage de programmation commun, utilisation d'un vocabulaire contrôlé, ainsi que de formats non-propriétaires. Les mêmes efforts peuvent être appliqués aux métadonnées ainsi que les références vers d'autres (méta)données.

- I1. Les (méta)données utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.
- I2. Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR
- I3. Les (méta)données incluent des références qualifiées à d'autres (méta)données

Langage formel et accessible

RDF : Resource Description Framework

OWL/XML : Web Ontology Language

JSON-LD : JavaScript Object Notation for Linked Data

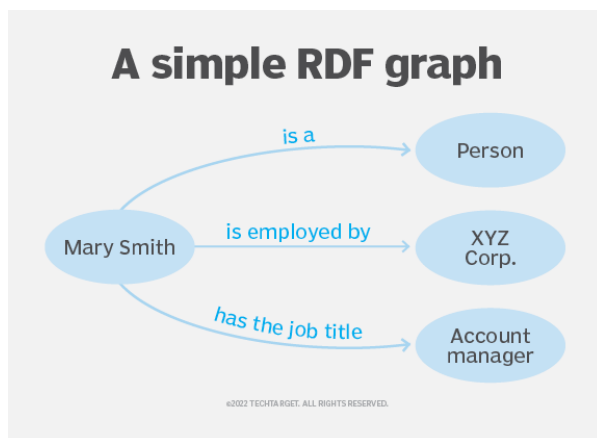
Exemple

RDF

Le terme « Utilisateur »

RDF

Exemple simple



Gestion de données | Home

Classes | Properties | Individuals | Comments | Changes by Entity | History

Class Hierarchy

- owl:Thing (1)
 - Gouvernance
 - OAIS
 - Acteurs de l'archivage
 - Archive (OAIS)
 - Management
 - Producteur
 - Utilisateur**
 - Fonctions de l'archivage
 - Types d'information
 - Renseignement
 - Statut légal au Canada
 - Système d'information
 - Sécurité informatique* (1)
 - Traitement de données* (1)
 - Utilisation des données sensibles à des fins

Class: Utilisateur

IRI
http://webprotege.stanford.edu/RCIxOLGCxu9c5zAhUHweC9b

Annotations

- rdfs:label Utilisateur
- skos:prefLabel Utilisateur

Enter property Enter value

Parents

- Acteurs de l'archivage

Enter a class name

Relationships

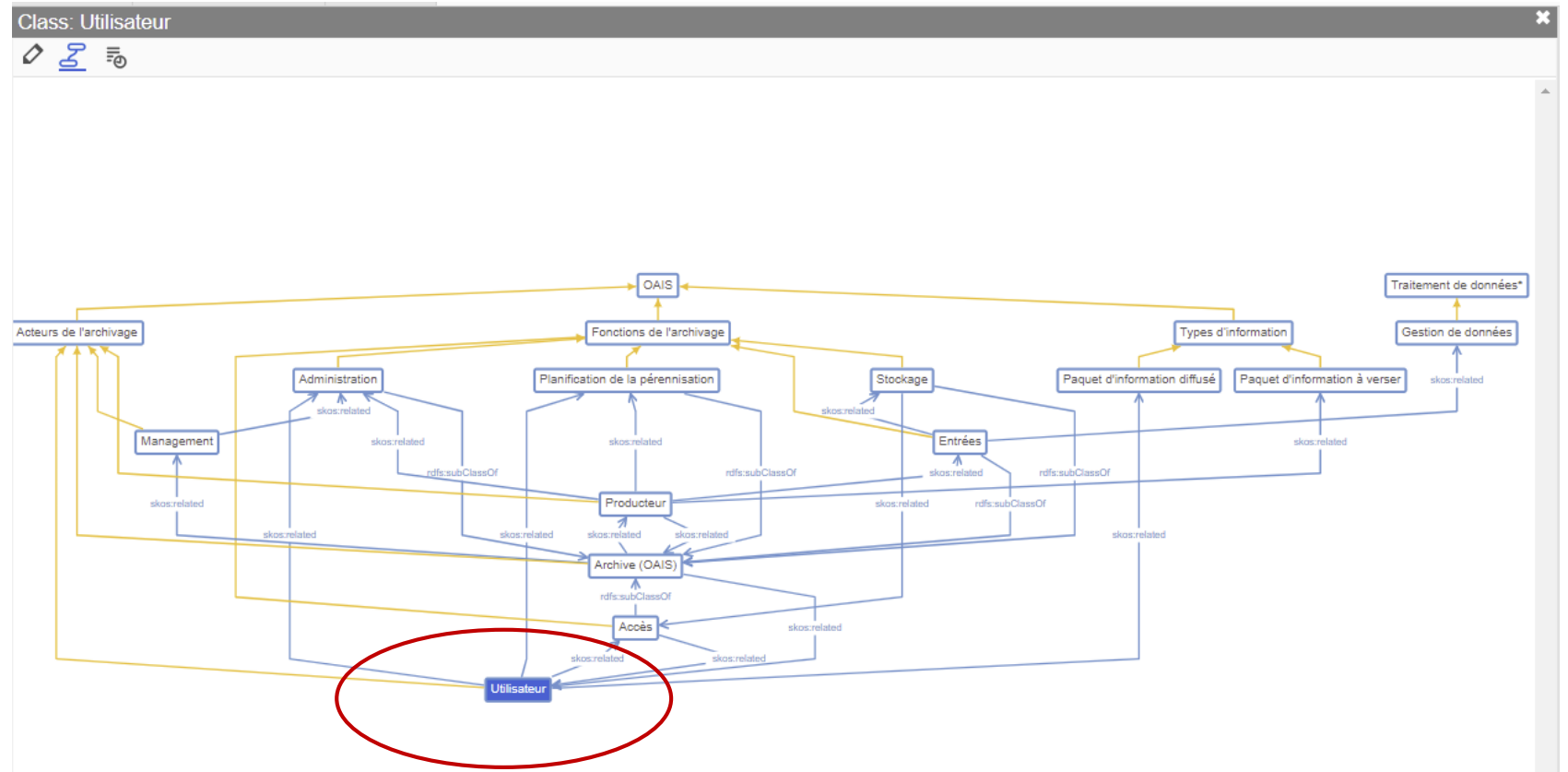
- skos:related Accès
- skos:related Administration
- skos:related Paquet d'information diffusé
- skos:related Planification de la pérennisation

Enter property Enter value

Exemple

RDF

Le terme « Utilisateur »



Exemple

OWL/XML

Le terme « Utilisateur »

```
<!-- http://webprotege.stanford.edu/RCIxOLGCxu9c5zAhUHweC9b -->

<owl:Class
rdf:about="http://webprotege.stanford.edu/RCIxOLGCxu9c5zAhUHweC9b">
  <rdfs:subClassOf
rdf:resource="http://webprotege.stanford.edu/RCIYWxaX3Y0YcjZ9Lxoz7I0"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty
rdf:resource="http://www.w3.org/2004/02/skos/core#related"/>
          <owl:someValuesFrom
rdf:resource="http://webprotege.stanford.edu/R7tnIILy1elk9h1qyPVt5Rn"/>
            </owl:Restriction>
          </rdfs:subClassOf>
        <rdfs:subClassOf>
          <owl:Restriction>
            <owl:onProperty
rdf:resource="http://www.w3.org/2004/02/skos/core#related"/>
              <owl:someValuesFrom
rdf:resource="http://webprotege.stanford.edu/R9Za9zUz3xWAgfP6PeEmA4b"/>
                </owl:Restriction>
            </rdfs:subClassOf>
          <rdfs:subClassOf>
            <owl:Restriction>
              <owl:onProperty
rdf:resource="http://www.w3.org/2004/02/skos/core#related"/>
                <owl:someValuesFrom
rdf:resource="http://webprotege.stanford.edu/RChGLN6x1TLzCyTtaGpRpBg"/>
                  </owl:Restriction>
            </rdfs:subClassOf>
          <rdfs:subClassOf>
            <owl:Restriction>
              <owl:onProperty
rdf:resource="http://www.w3.org/2004/02/skos/core#related"/>
                <owl:someValuesFrom
rdf:resource="http://webprotege.stanford.edu/RDvXxymFZ000j8N8TqWbN7d"/>
                  </owl:Restriction>
            </rdfs:subClassOf>
          <rdfs:label xml:lang="fr-ca">Utilisateur</rdfs:label>
          <skos:prefLabel xml:lang="fr-ca">Utilisateur</skos:prefLabel>
        </owl:Class>
```

Exemple

JSON-LD

Un exemple générique

```
{
  "@context": {
    "name": "http://xmlns.com/foaf/0.1/name",
    "homepage": {
      "@id": "http://xmlns.com/foaf/0.1/workplaceHomepage",
      "@type": "@id"
    },
    "Person": "http://xmlns.com/foaf/0.1/Person"
  },
  "@id": "http://me.example.com",
  "@type": "Person",
  "name": "John Smith",
  "homepage": "http://www.example.com/"
}
```

Structure

Des fichiers

Bonne pratique

Séparation des données, de la documentation et des scripts.

Exemple pour R

```
nom_du_projet
|--donnees_brutes
|  |--jeu_donnees.xlsx
|  |--1968rapport.csv
|  |--1969rapport.pdf
|--donnees_derivees
|  |--sommaire1968_1969.csv
|--rmd
|  |--01-analyse.Rmd
|--docs
|  |--01-analyse.html
|  |--01-analyse.pdf
|--scripts
|  |--analyse_exploratoire.R
|  |--pdf_scraper.R
|--nom_du_projet.Rproj
|--run_all.R
```

Structure

Des fichiers

Oxford Common File Layout

Complétude : afin qu'un référentiel puisse être reconstruit à partir des fichiers qu'il stocke.

Lisible : à la fois par les humains et les machines, pour s'assurer que le contenu peut être compris en l'absence de logiciel original

Robustesse contre les erreurs : la corruption et la migration entre les technologies de stockage

Gestion des versions : afin que les référentiels puissent apporter des modifications aux objets permettant à leur historique de persister

Diversité de stockage : pour garantir que le contenu peut être stocké sur diverses infrastructures de stockage, y compris les systèmes de fichiers conventionnels et les magasins d'objets cloud



Réutilisable

En ayant des données et des métadonnées suivant les meilleures pratiques, il est possible les optimiser afin qu'elles soient réutilisées. Les données comme les métadonnées doivent alors être richement détaillées, ainsi que les licences d'utilisation, et toute information spécifique au domaine d'intérêt des données en question.

- R1. Les (méta)données sont richement décrites avec une pluralité d'attributs précis et pertinents
 - R1.1. Les (méta)données sont publiées avec une licence d'utilisation des données claire et accessible
 - R1.2. Les (méta)données sont associées à une provenance détaillée
 - R1.3. Les (méta)données répondent aux normes communautaires applicables au domaine

Sélection d'une licence libre

Creative Commons

- Six déclinaisons qui permettent le partage et la distribution et la réutilisation des données ou de la base de données
- CC0 : données ou base de données directement au domaine public






















<https://creativecommons.org/>

Open Data Commons

- Public Domain Dedication and Licence (PDDL) : La base de données entre dans le domaine public et peut être utilisée sans aucune restriction.
- Attribution Licence (ODC-By) : La base de données peut être utilisée sans restriction, à la condition que l'auteur de la base de données soit crédité.
- Open Database Licence (ODbl) : Le produit de l'utilisation des données doit être partagé sous cette même licence ODbl. La source des données devra également être citée.

<https://opendatacommons.org/>

Les licences Creative Commons (CC)

		Utilisation Partage	Adaptation Modification	Utilisation commerciale	Modification de licence	
TRÈS LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
LIBRE						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
NON LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix NON permis
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix NON permis



BY (attribution to the creator)

ATTRIBUTION

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

L'auteur doit être cité



SA (Share Alike)

PARTAGE DANS LES MÊMES CONDITIONS

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Partage sous licence compatible



NC (Non Commercial)

POUR USAGE NON COMMERCIAL

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Pour usage non commercial



ND (No Derivative works)

PAS DE MODIFICATION

Vous pouvez retenir, réutiliser et redistribuer.

Aucune modification permise

Dictionnaire de données

Définir ses données

« Un dictionnaire de données est un fichier ou un ensemble de fichiers qui contient les métadonnées d'une base de données. Le dictionnaire de données contient des enregistrements sur d'autres objets de la base de données, tels que la propriété des données, les relations de données avec d'autres objets et d'autres données. »

Qu'est-ce qu'un dictionnaire de données? - Définition de techopedia - Bases de données - 2022. (s. d.). Icy Science. Consulté 15 juin 2022, à l'adresse <https://fr.theastrologypage.com/data-dictionary>

Université 
de Montréal
et du monde.

Ressources

Ressources

Diapo 5 : <https://alliancecan.ca/sites/default/files/2022-03/Priorite%CC%81s-du-Conseil-des-chercheurs-28-septembre-2021.pdf>

Diapo 6 : <https://www.canada.ca/fr/commission-fonction-publique/services/publications/roles-responsabilite-a-legard-donnees-a-la-cfp.html>

Diapo 7 : <https://ogsl.ca/fr/principes-fair/>

Diapo 7 : <https://www.go-fair.org/fair-principles/>

Diapo 9 : <http://valorisation-partage-de-la-recherche.etsmtl.net/identifiants-uniquees>

Diapo 10 : <https://theodi.org/article/marking-up-your-dataset-with-dcat/>

Diapo 11 : <https://archivesdemontreal.ica-atom.org/1920-3-lovelles-map-of-city-of-montreal-including-westmount-outremont-verdun-montreal-west-and-st-laurent-john-lovell-son-limited-publishers-of-lovelles-montreal-directory-every-year-since-1842-1920>

Ressources

Ressources

Diapo 20 : https://www.researchgate.net/figure/An-example-of-a-flat-Oxford-Common-File-Layout-OCFL-Storage-Hierarchy_fig1_333639228

Diapo 20 : <https://ocfl.io/>

Diapo 21 : <http://schema.datacite.org/>

Diapo 22 : <https://recherche.umontreal.ca/services-aux-chercheurs/gestion-des-donnees-de-recherche/>

Diapo 23 : https://fabriquerel.org/wp-content/uploads/CC-BY_fabriqueREL2.pdf

Diapo 24 : https://fr.wikipedia.org/wiki/Dictionnaire_des_donn%C3%A9es

Diapo 24 : <https://fr.theastrologypage.com/data-dictionary>

Me joindre

Dominic Boisvert

Responsable de la gouvernance des
données d'entreprise

dominic.boisvert.1@umontreal.ca