

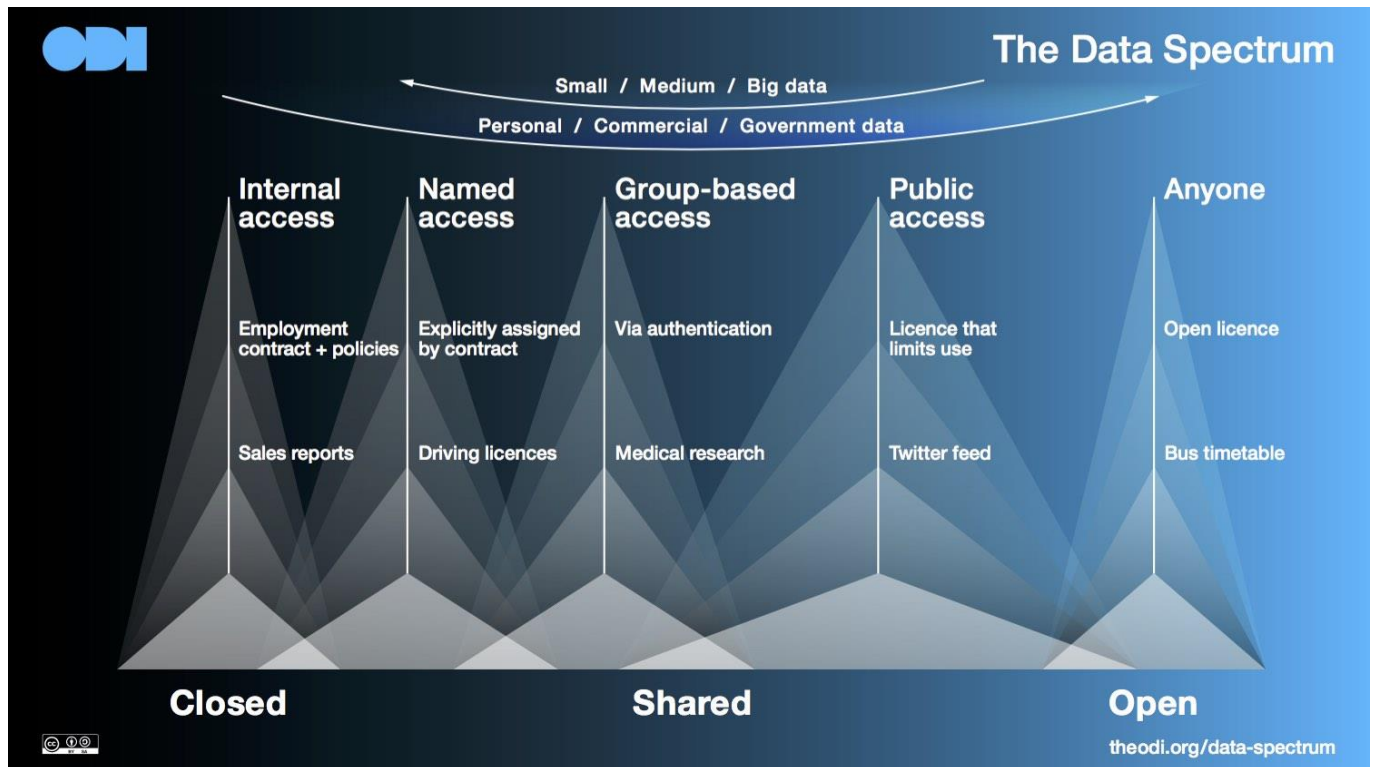
Fiche de cours

S-I4 Données ouvertes et Big Data

Contenus

Données ouvertes

Les données ouvertes sont des « données brutes non nominatives et libres de droits, produites ou recueillies par un organisme public ou privé, qui sont accessibles aux citoyens par Internet » (Office québécois de la langue française 2013).



(Open Data Institute 2015)

Principes

L'Open Data Charter (s.d.) a émis six principes concernant les données ouvertes :

1. **Ouvertes par défaut** : les données sont présumées ouvertes à la publication, sauf exception pour des raisons de sécurité ou de protection des données (personnelles par exemple) ;
2. **Diffusées en temps opportun et exhaustives** : publication rapide et complète des données, sous leur forme originale et non modifiée idéalement ;
3. **Accessibles et utilisables** : lisibles par les machines et faciles à trouver, dans des formats ouverts, en licences ouvertes ;
4. **Comparables et interopérables** : nécessité d'avoir des ensembles de données de qualité qui puissent communiquer entre eux ;

5. **Servant à l'amélioration de la gouvernance et de la participation citoyenne** : transparence qui peut améliorer les services publics ;
6. **Favorisant le développement inclusif et l'innovation** : stimule le développement économique.

Enjeux et considérations archivistiques

Conservation permanente

L'archiviste a un rôle à jouer dans la mise en place de dispositifs afin de garantir l'accès et l'exploitabilité durables des données ouvertes.

Évaluation

À travers l'évaluation archivistique, il convient d'identifier clairement les données qui peuvent être partagées de celles qui ne le peuvent pas, comme les données personnelles et / ou sensibles.

Protection et droits d'accès

Les données personnelles et sensibles sont soumises à des législations encadrant leur protection. Il convient donc de s'assurer que les droits d'accès soient correctement régulés.

Big Data

Le Big Data « désigne à la fois la production de données massives et le développement de technologies capables de les traiter afin d'en extraire des corrélations ou du sens » (Grinbaum 2017, p. 2).

Caractéristiques (Imbert 2016)

- 3 V
 - **V**olume : peta, exa, zettaoctet...
 - **V**élocité : fréquence à laquelle les données sont générées...
 - **V**ariété : données brutes, structurées, texte, image...
- + 2 autres :
 - **V**éracité: précision, pertinence → fiabilité
 - **V**aleur: génération d'une valeur pour l'institution

Enjeux et considérations archivistiques

Contrôle et gouvernance

Le Plan stratégique fédéral de recherche et de développement sur les grandes données de la Maison-Blanche, publié en mai 2016, traite de la création des capacités de la prochaine génération, de la compréhension de la fiabilité des données et des connaissances qui en découlent, de l'augmentation de la valeur des données grâce à des politiques, de la compréhension des grandes collectes de données en matière de protection de la vie privée, de sécurité et d'éthique et de la création de grands centres de comparaison des données (Marciano et al. 2018).

Conservation permanente

Dispositifs et garanties de l'accès et de l'exploitabilité durables des données massives

Évaluation et élimination

Hugues Ferreboeuf (2018, pp. 15-16) et son équipe estiment que « la consommation énergétique du Numérique dans le monde augmente d'environ 9% par an (période 2015 à 2020) » et que « [l]a part du Numérique dans la consommation finale d'énergie (elle-même en croissance de 1,5% par an) aura ainsi augmenté de presque 70% entre 2013 et 2020 ».

Face à ce problème, van Bussel, Smit et van de Pas (2015) proposent par exemple un *Green Archiving Model*, combinant les éléments de la *Green IT* et l'évaluation archivistique.

Protection

De nombreuses données sont récoltées à travers différentes applications, comme des dispositifs pour rendre les maisons ou les voitures intelligentes, ou encore les bracelets connectés. Il s'agit de l'Internet des Objets. Or si ces différentes données sont anonymisées, leur recoupage peut amener à une perte d'anonymisation (PFPDT [s.d.]).

Types d'évaluation

- Examen de connaissances
- Étude de cas
- ...

Ressources bibliographiques

BODER, Willy, 2016. Adrian Lobsiger : « L'internet des objets va nous occuper de plus en plus ». *Le Temps* [en ligne]. 9 septembre 2016. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.letemps.ch/economie/adrian-lobziger-linternet-objets-va-occuper-plus-plus>

Clefs : les Voix de la recherche [en ligne]. Juin 2017. N° 64. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.cea.fr/multimedia/Pages/editions/clefs-cea/big-data.aspx>

COMMISSARIAT À L'ÉNERGIE ATOMIQUE ET AUX ÉNERGIES ALTERNATIVES (CEA), 2017. L'essentiel sur... Le Big Data. *Découvrir & Comprendre: L'espace de culture scientifique du CEA* [en ligne]. 16 juin 2017. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.cea.fr/multimedia/Documents/publications/clefs-cea/CLEFS64-BIGDATA.pdf>

DENOIX, Antoine, 2018. Big data : comment collecter ses données ?. *Archimag.com* [en ligne]. 29 août 2018. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://www.archimag.com/univers-data/2018/08/29/big-data-comment-collecter-donnees>

DODDS, Leigh, 2016. How to write a good open data policy. *Open Data Institute's Website* [en ligne]. 26 décembre 2016. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://theodi.org/article/how-to-write-a-good-open-data-policy/>

FERREBOEUF, Hugues (dir.), 2018. *Lean ICT - Pour une sobriété numérique* [en ligne]. The Shift Project, octobre 2018. [Consulté le 25 septembre 2019]. Disponible à l'adresse : <https://theshiftproject.org/wp-content/uploads/2018/11/Rapport-final-v8-WEB.pdf>

GRINBAUM, Alexei, 2017. Big Data : de quoi parle-t-on ?. *Clefs : les Voix de la recherche* [en ligne]. Juin 2017. N° 64, p. 2. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.cea.fr/multimedia/Documents/publications/clefs-cea/CLEFS64-BIGDATA.pdf>

IMBERT, Clément, 2016. *L'encyclopédie des big data 2016* [en ligne]. EB G-Elenbi – 55 – IBM – PwC – Turn. [Consulté le 14 octobre 2019]. Disponible à l'adresse : https://www.ebg.net/publication/L_EncyclopA_die_des_Big_Data_2016-51232.htm

JOST, Clémence, 2017. "L'open data est un nouveau défi pour les archivistes". *Archimag.com* [en ligne]. 26 mai 2017. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://www.archimag.com/archives-patrimoine/2017/05/26/open-data-nouveau-defi-archivistes>

JOST, Clémence, TEXIER, Bruno, 2015. Infographie : le big data en mots et en chiffres. *Archimag.com* [en ligne]. 26 novembre 2015. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://www.archimag.com/univers-data/2015/11/26/infographie-big-data-mots-chiffres>

MARCIANO, Richard, LEMIEUX, Victoria, HEDGES, Mark, ESTEVA, Maria, UNDERWOOD, William, KURTZ, Michael, CONRAD, Mark, 2018. Archival Records and Training in the Age of Big Data. In: PERCELL, Johnna, SARIN, Lindsay C., JAEGER, Paul T. and BERTOT, John Carlo, (Ed.). *Re-envisioning the MLS: Perspectives on the Future of Library and Information Science Education* [en ligne]. Emerald Publishing Limited, 17 mai 2018, pp. 179-199. *Advances in Librarianship*, Vol. 44B. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://doi.org/10.1108/S0065-28302018000044B010>

Programme de conservation des documents numériques (PCDN)
Écoles d'études pour les enseignants en archivistique
École de Bibliothécaires, Archivistes et Documentalistes (EBAD), Dakar, 21-25 octobre 2019

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL (NSTC), 2016. *The Federal Big Data Research and Development Strategic Plan* [en ligne]. Executive Office of the President, National Science and Technology Council, mai 2016. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.nitrd.gov/pubs/bigdatardstrategicplan.pdf>

OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, 2013. Données ouvertes. *Le grand dictionnaire terminologique (GDT)* [en ligne]. [Consulté le 14 octobre 2019]. Disponible à l'adresse : http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26519745

OPEN DATA CHARTER, [s. d.]. Principes. *Charte internationale sur les données ouvertes* [en ligne]. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://opendatacharter.net/principles-fr/>

OPEN DATA INSTITUTE, 2015. The Data Spectrum. *Open Data Institute's Website* [en ligne]. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://theodi.org/about-the-odi/the-data-spectrum/>

PRÉPOSÉ FÉDÉRAL À LA PROTECTION DES DONNÉES ET À LA TRANSPARENCE (PFPDT), [s. d.]. Explications relatives aux Big Data (données massives). *Site du PFPDT* [en ligne]. [Consulté le 16 octobre 2019]. Disponible à l'adresse : https://www.edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/Internet_und_Computer/services-en-ligne/explications-relatives-aux-big-data--donnees-massives-/explications-relatives-aux-big-data--donnees-massives-.html

ROSSEL, Pierre, 2014. Le grand potentiel des mégadonnées. *Annale des Journées d'informatique juridique* [en ligne]. 5 novembre 2014. [Consulté le 16 octobre 2019]. Disponible à l'adresse : https://rechtsinformatik.ch/wp-content/uploads/2014/11/Rossel_Big-data-PR_05112014-v3.pdf

TALÈS, Romain, SUCEVIC, PÉRICA, 2017. Archives et open data : les conditions de l'ouverture des données. *Archimag.com* [en ligne]. 26 mai 2017. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://www.archimag.com/archives-patrimoine/2017/05/26/archives-open-data-conditions-ouverture-donnees>

TEXIER, Bruno, 2017a. Archivistes, emparez-vous de l'open data !. *Archimag.com* [en ligne]. 26 mai 2017. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://www.archimag.com/archives-patrimoine/2017/05/26/archivistes-emparez-vous-open-data>

TEXIER, Bruno, 2017b. Open data : "les archivistes doivent tirer parti de leurs compétences". *Archimag.com* [en ligne]. 26 mai 2017. Mis à jour le 21 juin 2017. [Consulté le 15 octobre 2019]. Disponible à l'adresse : <https://www.archimag.com/archives-patrimoine/2017/05/26/open-data-archivistes-tirer-part-competences>

VAN BUSSEL, Geert-Jan, SMIT, Nikki, VAN DE PAS, John, 2015. Digital Archiving, Green IT and Environment. Deleting Data to Manage Critical Effects of the Data Deluge. *The Electronic Journal Information Systems Evaluation* [en ligne]. Vol. 18, n° 2, pp. 187-198. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <http://www.ejise.com/issue/download.html?idArticle=986>

WUTHRICH, Bernard, 2017. La protection des données est mise à mal par le Big Data. *Le Temps* [en ligne]. 26 juin 2017. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.letemps.ch/suisse/protection-donnees-mise-mal-big-data>

Vidéo

BUNDESARCHIV, 2016. opendata.swiss - le portail des données ouvertes de l'administration publique suisse (OGD). YouTube [en ligne]. 26 mai 2016. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.youtube.com/watch?v=4ow8rhxNmdI>

COOKIE CONNECTÉ, 2017a. BIG DATA : Pourquoi les entreprises s'en emparent! (Part 1). *YouTube* [en ligne]. 5 septembre 2017. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.youtube.com/watch?v=JJxj8Oz7gFk>

COOKIE CONNECTÉ, 2017. BIG DATA : Et nos données personnelles alors ? (Part 2). *YouTube* [en ligne]. 5 septembre 2017. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.youtube.com/watch?v=yfN5I5-3F5I>

COOKIE CONNECTÉ, 2017c. L'Opendata expliqué en emojis. *YouTube* [en ligne]. 24 octobre 2017. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.youtube.com/watch?v=pnrc6ZaYrwg>

DATA GUEULE, 2014. Big data : données, données, donnez-moi ! - #DATAGUEULE 15. *YouTube* [en ligne]. 15 novembre 2014. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.youtube.com/watch?v=5otaBKsz7k4>

DATA GUEULE, 2017. Gouvernés par les nombres ? Échecs et maths. - #DATAGUEULE 76. *YouTube* [en ligne]. 1er décembre 2017. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.youtube.com/watch?v=3Xct2cLcsWY>