

SCI6005 Information numérique et informatique documentaire (A2025)

Christine Dufour, EBSI, UdeM

A2025 2 octobre 2025

Cours 5- Environnements en réseau : Internet et Web
SCI6005

Table des matières

I - Cours 5 - Environnements en réseau : Internet et Web	3
1. + Au programme aujourd'hui	3
2. + Alignement pédagogique	3
3. Introduction	5
4. Évolution d'Internet et du Web	5
4.1. Historique du Web et d'Internet	5
4.2. Culture informationnelle et numérique sur le Web	16
5. Données sur le Web	19
5.1. Données ouvertes	19
5.2. Données liées	20
5.3. Données de recherche	21
6. Importance de la normalisation.....	22
7. Ressources en lien avec le cours	24
Glossaire	25
Bibliographie	28
Webographie	30
Index	31
Crédits des ressources	32

Cours 5 - Environnements en réseau : Internet et Web



1. + Au programme aujourd'hui

- Évolution du Web et d'Internet
- Types de données sur le Web
- Importance de la normalisation

2. + Alignement pédagogique



Objectifs visés, matériel du cours et évaluation : Examen mi-session

Objectif général : Comprendre la place des technologies et de l'information numérique en contexte documentaire		
Objectifs spécifiques	Compétences à développer	Matériel associé
Expliquer les concepts fondamentaux des environnements en réseau	Définir ce qu'est Internet, le Web 1.0, le Web 2.0, le Web 3.0 et l'intelligence artificielle générative	Section <i>Évolution d'Internet et du Web</i>
	Donner des exemples de plateformes 2.0	Section <i>Évolution d'Internet et du Web</i> > <i>Historique du Web et d'Internet</i> > <i>Web</i> > <i>Web 2.0 : Le Web participatif</i>
	Expliquer les enjeux liés à l'évolution du Web	Section <i>Évolution d'Internet et du Web</i> > <i>Culture informationnelle et numérique sur le Web</i>
	Expliquer ce que sont les données ouvertes, les données liées et les données de recherche	Section <i>Données sur le Web</i> <i>TP Données</i>
	Donner des raisons justifiant l'importance de la normalisation	Section <i>Importance de la normalisation</i>

Lien entre les objectifs, les compétences à développer et le matériel du cours 5



Objectifs visés, matière du cours et activités associées

Objectif général : Comprendre la place des technologies et de l'information numérique en contexte documentaire		
Objectifs spécifiques	Matière	Activités
Expliquer les concepts fondamentaux des environnements en réseau	Données ouvertes	<i>TP Données</i>
	Données liées	<i>TP Données</i>
	Données de recherche	<i>TP Données</i>

Lien entre les objectifs, la matière du cours 5 et les activités associées

3. Introduction

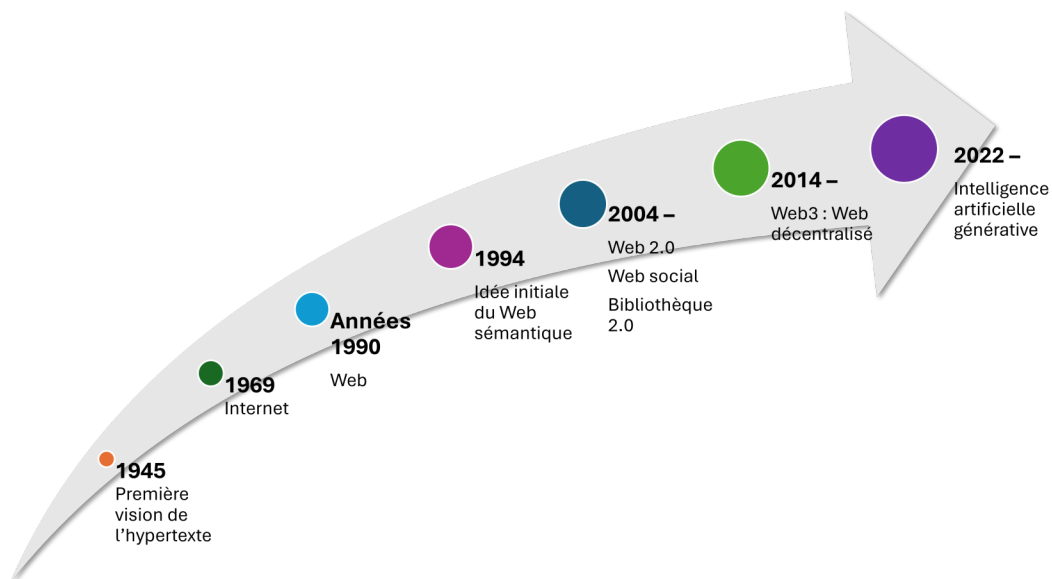
Afin de bien vivre dans l'environnement réseauté omniprésent dans le quotidien des professionnelles et professionnels de l'information, en sus des concepts de bases des environnements réseautés, il est important de comprendre un des réseaux que l'on croise au quotidien, soit le Web. Les milieux documentaires ont intégré le Web comme plateforme pour leurs ressources et leurs services avec les défis qui en découlent. *Lankes et al. (2007)*^{LANKES et al. 2007 p.28} soulignent en effet que ces derniers sont multiples entre autres pour l'intégration du Web 2.0 au sein de leurs systèmes d'information, tant sur le plan politique et éthique qu'opérationnel et technique. Une attention particulière doit être portée à chacune des composantes de l' "organisation 2.0" (acteurs, infrastructures, etc.) et à son insertion dans son environnement informationnel (*Dufour & Cadieux, 2013*^{DUFOUR 2013 p.28}).

Après avoir précisé les bases de l'évolution du Web, nous nous intéresserons à la culture numérique qui en découle. Nous enchaînerons par la suite sur les principaux types de données qu'on y retrouve ainsi que sur l'importance de la normalisation dans ces environnements réseautés.

4. Évolution d'Internet et du Web

4.1. Historique du Web et d'Internet

Dans sa **conceptualisation**, l'histoire de l'Internet et du Web remonte aux **années 40** lorsque Vannevar Bush a rêvé d'un univers de documents interconnectés en réaction à l'explosion documentaire qu'il observait. Bien qu'il n'ait pas lui-même inventé le terme "hypertexte", c'est une des premières instances de ce concept. Dans sa **concrétisation technique**, c'est dans les **années 60** que le premier nœud de ce qui deviendra Internet a été mis en place par l'équipe de Tim Berners Lee.



Quelques dates importantes dans l'histoire d'Internet et du Web

On remarque que l'**écart** entre les moments clés **se raccourcit** plus le temps avance. En fait, cette observation peut se faire plus globalement lorsque l'on observe l'**adoption des technologies**. Ceci est évident si on compare le temps pris pour atteindre *50 millions d'auditeurs* pour certaines *technologies marquantes* dans notre histoire : la radio a pris 38 ans, la télévision 13 ans, Internet 4 ans, le iPod 3 ans, et Facebook 2 ans (*source* : Did you know 3.0? 2012, https://www.youtube.com/watch?v=YmwwrGV_aiE; il est à noter que d'autres sources donnent des chiffres un peu différents, par exemple <https://x.com/Guglielminetti/status/919245063967911937>, mais une tendance similaire est observable). À titre d'exemple supplémentaire, certaines sources disent que ChatGPT a pris 2 mois pour atteindre 100 millions d'utilisateurs (*source* : After the ChatGPT moment: Measuring AI's Adoption, 2025, <https://epoch.hai.substack.com/p/after-the-chatgpt-moment-measuring>)!

Si le rythme est essouffant pour l'utilisateur et l'utilisatrice qui veut rester à jour, il l'est encore plus pour les organisations dont les activités sont étroitement liées à ces environnements numériques. Cela leur demande d'être de plus en plus *agiles* dans leur intégration des technologies et de s'habituer aux *changements perpétuels*, ce qui n'est pas toujours facile d'un point de vue humain et organisationnel.

L'histoire d'Internet et du Web est jalonnée de moments clés qui seront décrits dans les sections qui suivent.

a) Internet

D'entrée de jeu, il est important de bien comprendre que les termes *Internet* et *Web* **ne sont pas synonymes**! On retrouve parfois une certaine confusion entre les deux termes. Plus précisément, on peut définir **Internet** ainsi :

« A huge computer network available to everyone with a personal computer and a means to connect to it. It is the actual physical network made up of wires, cables, and satellites as opposed to the web, which is the multimedia interface to resources available on the Internet. »

(Source : O'Leary et al. O'LEARY, O'LEARY & O'LEARY 2019 p.28, 2019, p. 367)

Le **Web** est ainsi **une des sphères d'Internet**, Internet étant un réseau plus large regroupant différentes sphères (Web, courriel, FTP, etc.).

Internet est né d'une initiative militaire américaine. Le premier nœud du réseau ARPANET (*Advanced Research Project Agency Network*) à l'origine d'Internet a été mis en place en 1969. Avant l'apparition du Web, les échanges sur Internet étaient uniquement textuels. Leiner et al. (1997)^{LEINER et al. 1997 p.30} précisent qu'une des clés du développement rapide d'Internet est le fait que la documentation, et plus particulièrement les spécifications et les protocoles, était accessible gratuitement et librement.

b) Web

Comme précisé dans la section précédente, le **Web** est une *interface graphique* à des contenus sur *Internet* :

« Prior to the introduction of the web in 1992, the Internet was all text. The web made it possible to provide a multimedia interface that includes graphics, animations, sound, and video. »

(Source : O'Leary et al. O'LEARY, O'LEARY & O'LEARY 2019 p.28, 2019, p. 377)

Au cœur du **Web** se retrouve l'idée des *documents interconnectés* proposés en 1945 par Vannevar Bush. Dans son texte "*As We May Think*",^{BUSH 1945 p.28} il propose l'idée d'une machine, nommée MEMEX, pour résoudre les problèmes liés à l'explosion documentaire. Cette dernière serait en effet en mesure de ranger et rendre accessible tous les documents en permettant :

- La visualisation des documents sur microfilm sur un écran, documents pouvant être complétés par des images;
- La localisation des documents;
- L'inclusion de pistes associatives entre les documents.

Le **néologisme "hypertexte"** a été proposé par le philosophe et sociologue américain Theodore Nelson en 1965. Il est à l'origine du projet Xanadu dont l'objectif était de regrouper tous les documents de l'humanité. Ce système, accessible à tous, permettrait l'ajout de documents et la définition de liens entre ces derniers. Au départ, il s'agissait uniquement d'une conceptualisation sans implémentation. Depuis, des démonstrations ont vu le jour après bien des années, dont OpenXanadu¹ en 2014^{HERN 2014 p.28}.

1. <https://xanadu.com/xanademos/MoeJusteOrigins.html>

Le *premier logiciel hypertexte grand public*, créé par Bill Atkinson, est **HyperCard** (Apple) en 1987, logiciel dont le développement s'est terminé en 1998. Il permettait de présenter des contenus multimédias sur des fiches virtuelles cartonnées, cartes qui pouvaient être liées entre elles par des boutons. Pour les curieux et les curieuses (ou les nostalgiques!), Internet Archive a mis en ligne en 2017 pour fêter les 30 ans d'HyperCard une collection de projets HyperCard².

Finalement l'*application la plus connue du modèle hypertextuel*, le **Web**, a été créée au CERN par l'équipe de Tim Berners-Lee en 1989. Il s'agit d'une version "allégée" des visions initialement proposées pour l'hypertexte entre autres par T. Nelson et V. Bush :

- Absence de gestion des droits d'auteur;
- Liens unidirectionnels se cassant lorsque la ressource disparaît;
- Absence d'aperçu des ressources liées;
- Absence de système d'annotation et de gestion de versions.

Le Web est passé par *différentes phases* (Web 1.0, Web 2.0 et Web 3.0) qui seront décrites dans les sections subséquentes.

i) Web 1.0 : Les premières années "statiques" du Web

La "première version du Web", le *Web 1.0* en quelque sorte que certains désignent comme un *Web de documents*, correspond à un Web où un *petit nombre d'internautes* étaient des *créateurs ou créatrices de contenus*, cette création demandant en effet des compétences informatiques plus avancées. La *majorité* des internautes était ainsi des *observateurs ou des observatrices* pouvant chercher parmi les informations existantes et les consulter. Maignien (2014) le compare à une *bibliothèque distribuée sans murs* :

« Le modèle documentaire est alors clairement celui d'une bibliothèque distribuée, sans murs, autrement appelée bibliothèque virtuelle ou numérique, mais où l'information n'est pas classée de façon hiérarchisée, dans un arbre, mais distribuée en graphe (il existe autant de relations non hiérarchisées que de liaisons possibles entre les documents ou des parties de ces documents). »

(Source : Maignien, 2014^{MAIGNIEN 2014 p.28}, p. 78)

ii) Web 2.0 : Le Web participatif

La génération suivante, le *Web 2.0*, est un phénomène à la fois *technologique* et *social*, ce que fait bien ressortir cette définition :

« Web doté d'outils et de contenus interactifs qui permettent aux internautes de participer à la création de contenus Web, de partager de l'information en ligne et de communiquer entre eux. »

(Source : Grand dictionnaire terminologique, 2018, <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26519897/web-20>)

Les mots-clés ici sont "*participer*", "*création de contenus*" et "*partager*". Certains *services* et certaines *technologies phares* sont associés au Web 2.0.

2. <https://archive.org/details/hypercardstacks?tab=about>

Principaux services et technologies phares du Web 2.0

Services/Technologies	Définition	Exemples
Blogues	« Site Web ou section de site Web généralement tenus par une seule personne, consacrés à une chronique personnelle ou à une thématique particulière présentées sous forme de billets ou d'articles. » (OQLF, 2018, https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8370242/blogue)	Bibliomancienne ³ Convergence AAQ : Le blogue de l'Association des archivistes du Québec ⁴
Microblogues	« Blogue permettant de communiquer en temps réel au moyen de messages courts de 140 à 240 caractères, le plus souvent par l'intermédiaire d'une plateforme de microblogage, afin de partager de l'information avec une communauté d'internautes. » (OQLF, 2018, https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26502311/microblogue)	Bruno Guglielminetti ⁵
Wikis	« Site Web collaboratif où chaque internaute visiteur peut participer à la rédaction du contenu. » (OQLF, 2011, https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8362053/site-wiki)	Wikipédia ⁶
Réseaux sociaux	Environnements où l'utilisateur ou l'utilisatrice se définit un profil (personnel ou professionnel) et réseaute avec d'autres utilisateurs et utilisatrices	Facebook ⁷ (personnel) LinkedIn ⁸ (professionnel)
Plateformes de partage de ressources	Plateformes facilitant la composition de collections de ressources diverses et le partage de ces dernières	Flickr ⁹ (photographies) YouTube ¹⁰ (vidéos) Slideshare ¹¹ (présentations powerpoint)
Fils de syndication	« Fil d'information consistant en un fichier XML, généré automatiquement, dont le contenu formaté, exploitable dynamiquement par d'autres sites Web, est récupérable par l'entremise d'un agrégateur qui permet de lire le nouveau contenu de ce fil répertorié, dès qu'il est disponible. »	Formats les plus connus : RSS et ATOM

3. <https://bibliomancienne.com/>

4. <https://archivistesqc.wordpress.com/>

5. <https://bsky.app/profile/guglielminetti.bsky.social>

6. https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal

7. <https://www.facebook.com/>

8. <https://www.linkedin.com/>

9. <https://www.flickr.com/>

10. <https://www.youtube.com/>

11. <https://fr.slideshare.net/>

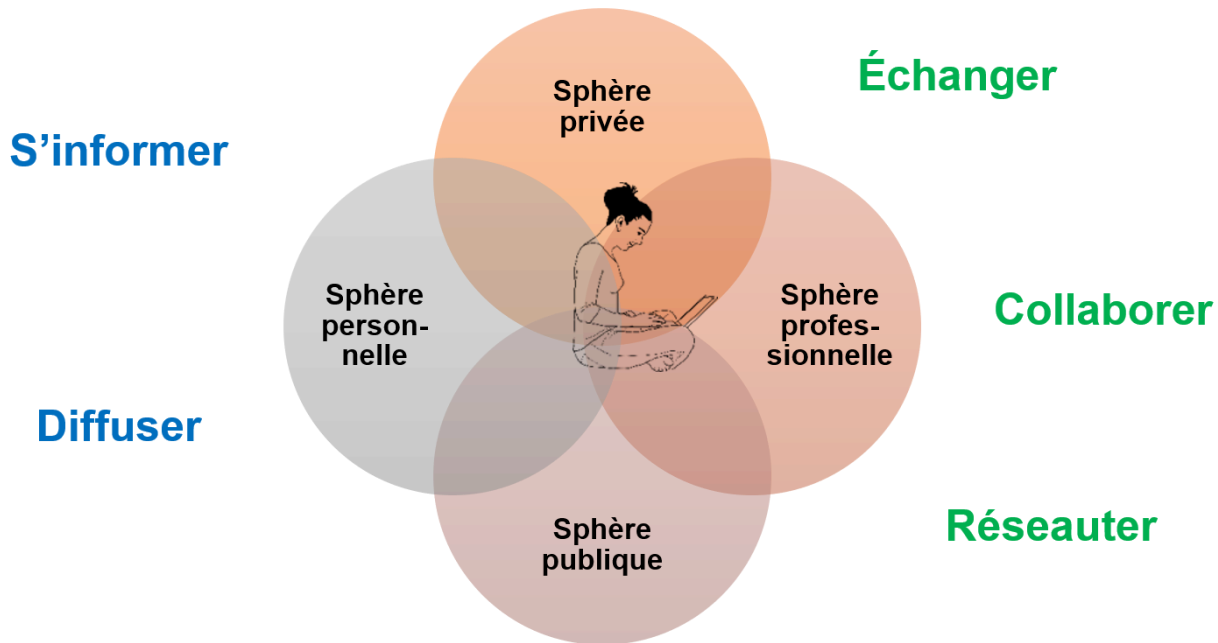
Services/Technologies	Définition	Exemples
	(OQLF, 2006, https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8869144/fil-de-syndication)	
Folksonomies	« Système de classification collaborative et spontanée de contenus Internet, basé sur l'attribution de mots-clés librement choisis par des utilisateurs non spécialistes, qui favorise le partage de ressources et permet d'améliorer la recherche d'information. » (OQLF, 2006, https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8351986/folksonomie)	Se retrouvent par exemple sur les plateformes de partage de ressources pour décrire les ressources ou sur les blogues pour décrire les billets

L'expression Web 2.0 a été utilisée pour la première fois en 2004 par l'équipe d'O'Reilly Media qui, reprenant la métaphore des numéros de version pour les logiciels, voulait indiquer une version améliorée du Web. Certains lui préfèrent maintenant l'expression *Web social* qui est en effet une de ses principales facettes. À ses débuts, certains voyaient le Web 2.0 comme un effet de mode et, à l'autre extrême, d'autres le percevaient comme un changement de paradigme. Les chiffres, qui croissent d'année en année, montrent sans l'ombre d'un doute qu'il ne s'agissait pas d'un effet de mode.

Type	Plateforme	Chiffres	Source
Partage de ressources	Instagram	2 milliards d'utilisateurs ou utilisatrice	https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (Février 2025)
	YouTube	20 millions de vidéo téléchargés quotidiennement en moyenne Plus de 100 pays en 80 langues	https://blog.youtube/press/ (2025-09-18)
Réseaux sociaux	LinkedIn	Plus de 850 millions d'utilisateurs et utilisatrices Plus de 200 pays et territoires	https://about.linkedin.com/fr-fr (2025-09-18)
	Facebook	Plus de 3 milliards d'utilisateurs et utilisatrices Plus de 510 000 commentaires rédigés en une seule minute	https://digitiz.fr/statistiques-facebook/ (2025-09-18)
Rédaction collaborative	Wikipédia	65 516 514 articles 2 709 830 articles en français 343 langues actives	https://meta.wikimedia.org/wiki/List_of_Wikipedias et https://wikistats.wmcloud.org/display.php?t=wp (2025-09-18)
Micro-blogsphère	X	586 millions d'utilisateurs et utilisatrices	https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (Février 2025)
	Bluesky	38 millions d'utilisateurs et utilisatrices	https://www.statista.com/statistics/1536616/global-bluesky-users/ (Août 2025)

Quelques chiffres sur certaines plateformes types du Web 2.0

Si on peut hésiter à parler d'une *révolution*, il s'agit très clairement d'une *évolution* tant des *technologies* que de leur *utilisation*. L'internaute qui le désire a maintenant à sa portée des technologies qui lui permettent de passer, s'il ou elle le désire, d'un rôle uniquement d'observation à un rôle de création, que ce soit dans ses activités personnelles ou professionnelles. Les internautes utilisent le Web 2.0 sur une base personnelle, mais aussi professionnelle, dans un cadre public ou privé, que ce soit pour informer et s'informer (à la « Web 1.0 »), mais aussi pour collaborer, réseauter ou échanger dans le cadre de leurs activités.



Sphères d'utilisation du Web 2.0

Dans l'univers du Web 2.0, on observe que les *sphères publique et privée, personnelle et professionnelle* deviennent *perméables*, leurs frontières étant plus floues. Par exemple, certains blogueurs mélangent parfois anecdotes personnelles et réflexions liées à leur travail sur leurs blogues. Le caractère privé de certaines formes d'expression, comme les journaux intimes, s'ouvre à la sphère publique, par exemple sur les blogues et les réseaux sociaux.

Adoption du Web 2.0 au Québec

L'utilisation faite des technologies du Web 2.0 est aussi tributaire du contexte des internautes. Une *fracture numérique* existe toujours pour certaines régions ou certains pays ayant un accès limité à ces plateformes pour des raisons économiques ou sociales. Au Québec, la population est globalement branchée et active sur les réseaux sociaux. En 2024, 92 % des adultes québécois ayant participé à l'enquête NETendances avaient utilisé Internet au cours des 3 derniers mois (Académie de la transformation numérique. 2024. NETendances 2024 : Portrait numérique des foyers québécois. Volume 15, no 5. <https://transformation-numerique.ulaval.ca/wp-content/uploads/2025/02/netendances24-intelligence-artificielle-generative-et-donnees-personnelles.pdf>).

Les jeunes ne sont pas en reste! Une grande majorité des 13 à 17 ans, selon l'enquête NETendances, possèdent un profil sur un réseau social (81 %), TikTok étant le plus populaire (67 %), suivi par Instagram (59 %) et Snapchat (52 %) (Académie de la transformation numérique. 2024. NETendances 2024 : Famille numérique. Volume 15, no 6. <https://transformation-numerique.ulaval.ca/wp-content/uploads/2025/03/netendances24-famille-numerique.pdf>). Selon cette même étude, une des activités les plus pratiquées par les enfants de 6 à 17 ans est l'écoute de vidéos par exemple sur YouTube ou Tiktok (73 %).

On retrouve plusieurs initiatives d'exploitation des plateformes 2.0 entre autres au sein des institutions culturelles québécoises.

Initiatives	But(s)
Projets Wiki de BAnQ https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:BAnQ	« Ces projets visent à mettre en valeur les fonds et les collections de BAnQ, à rendre disponibles aux wikimédiens les ressources numérisées de l'institution et à contribuer à l'enrichissement des articles de Wikipédia (et ses projets frères) concernant le Québec, le Canada français et, plus largement, l'Amérique française. »
Projets Wiki de la Cinémathèque québécoise https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Cin%C3%A9math%C3%A8que_qu%C3%A9b%C3%A9coise	« La Cinémathèque québécoise s'investit dans la production et la diffusion de connaissances libres sur le cinéma québécois. Depuis septembre 2017, nous organisons notamment des activités citoyennes dans le but d'enrichir les projets Wikimedia dans ce domaine. »
Café des savoirs libres https://www.savoirslibres.ca/a-propos/ ¹²	« Le Café des savoirs libres est un collectif montréalais de bibliothécaires, enseignants, chercheurs et autres passionné(e)s qui organise des rencontres de partage, des projets et des événements autour des savoirs libres et des communs numériques. »

Quelques initiatives québécoises sur le Web 2.0

(Source d'inspiration : compte-rendu de la conférence-midi du 21 septembre 2017 sur la culture ouverte et le savoir libre par @bibliomancienne (Marie D. Martel) <https://bibliomancienne.com/2017/09/22/culture-ouverte-et-savoir-libre-a-lebsi-sqil-copibecpasenmonnom/>)

iii) Web 3.0 : Le Web sémantique

Le *World Wide Web Consortium* (W3C) définit le *Web sémantique* ainsi :

« The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. »

(Source : W3C, 2013^{W3C 2013 p.30})

L'idée centrale de ce qu'ils identifient comme un *Web de données* est celle d'un Web où les documents sont « **enrichis** » de manière à les rendre compréhensibles non seulement aux humains, mais aussi aux machines pour en faciliter la réutilisation. Cet enrichissement se fait à l'aide de *métadonnées* ^{p.26} permettant de décrire différents aspects des documents (par exemple, l'auteur, la date de création).

En fait, les folksonomies du Web 2.0 représentent une *certaine* couche « sémantique » ajoutée aux ressources et illustre le potentiel du « Web sémantique » en ce qu'elles permettent la création de nouvelles connaissances; « *certaine* » car c'est souvent une sémantique plus personnelle que collective qu'on y retrouve comme les étiquettes que l'on met à des photos sur Flickr par exemple. À un certain moment, parmi les étiquettes les plus populaires sur Flickr se retrouvait le mot "*me*" comme plusieurs utilisateurs l'employaient pour les photos où ils se retrouvaient!

Les applications composites – ou les *mashups* – que l'on retrouve sur le Web 2.0 démontrent l'intérêt du croisement de diverses données. Pensons par exemple à la possibilité d'explorer les photos déposées sur Flickr à l'aide d'une carte (<https://www.flickr.com/map>). La carte créée est le résultat du croisement des étiquettes géographiques apposées par les utilisateurs de Flickr à leurs photos avec une carte.

¹². <https://www.savoirslibres.ca/>

Cette idée d'ajouter aux documents du Web une couche sémantique pour en faciliter le partage et la réutilisation est attribuée à Tim Berners-Lee. Une des premières mentions de cette idée lui est attribuée lors de l'*International World Wide Web Conference*¹³ en 1994. Le *Web sémantique* est ainsi une *extension du Web* que l'on connaît par l'ajout de cette *couche sémantique* et le *développement de technologies permettant de l'exploiter*. Différents standards y sont rattachés que Bermès et coll.^{BERMÈS et al. 2013 p.28} (2013, p. 28) décrivent ainsi :

« Ainsi, le Web sémantique a pour objet de fournir des standards qui vont rendre possible ce Web plus intelligent et plus efficace. Ces standards incluent le modèle **RDF (*)**, des outils comme **RDFS (*)** (RDF Schema) et **OWL (*)** (Web Ontology Language) pour décrire la sémantique et la logique des données, et aussi des standards pour manipuler et traiter les données, comme **SPARQL (*)**, le langage et protocole de requête de RDF. »

Dans leur ouvrage, ces auteurs s'intéressent au Web sémantique en bibliothèque et espèrent réussir par cet ouvrage à "mobiliser vos [bibliothécaires] compétences d'analyse de l'information et de traitement des données pour vous [bibliothécaires] montrer comment elles peuvent être mises en valeur dans le nouvel environnement apporté par le Web sémantique." (Bermès et al., 2013, p. 13). Comme *Da Sylva*^{DA SYLVA 2017 p.28} le précise (2017, p. 28), Bermès et al. dans leur ouvrage reconnaissent le rôle des bibliothèques dans le Web sémantique, rôle en continuité avec leurs missions et leurs activités traditionnelles. Il importe ainsi comme professionnelles et professionnels de l'information d'en connaître les bases et tout particulièrement les types de données qu'on y retrouve, soit les *données ouvertes liées* décrites plus loin dans les notes de cours.

Google Knowledge Graph



Un exemple du Web sémantique que l'on croise fréquemment dans nos recherches sur le Web est le *Google Knowledge Graph*. En effet, on voit apparaître lors de certaines recherches dans le moteur de recherche Google un encart à la droite de la liste des résultats retournés qui présente des informations souvent factuelles en lien avec la recherche. C'est le cas par exemple si vous faites une recherche sur une personnalité connue ou un pays. Cet encart est produit par la mise en relation de plusieurs jeux de données distincts en lien avec l'objet de votre recherche. C'est un exemple du potentiel du Web sémantique qui permet, par la couche sémantique ajoutée à certaines données, de générer de nouvelles données (les encarts de *Google Knowledge Graph*). Plus d'informations sur cette fonctionnalité sont disponibles sur le blogue de Google à l'URL <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>.

iv) Intelligence artificielle générative

Si l'intelligence artificielle générative (IAg) est un phénomène relativement récent pour le grand public, l'origine de l'intelligence artificielle (IA) remonte quant à elle à plusieurs décennies. Voici quelques jalons des développements en lien avec l'IA à titre d'exemple :

- Le premier article fondateur sur les neurones formelles, qui a servi de base théorique à l'intelligence artificielle, a été publié par Warren McCulloch et Walter Pitts en 1943 (*WIKIPEDIA - Réseau de neurones artificiels 2025*^{WIKIPEDIA - Réseau de neurones artificiels 2025 p.29}).
- L'intelligence artificielle est en fait officiellement devenue une discipline scientifique en 1956 lors d'une conférence à Dartmouth (*WIKIPEDIA - Conférence de Dartmouth 2025*^{WIKIPEDIA - Conférence de Dartmouth 2025 p.29}).
- ELIZA, Le premier programme capable de simuler une conversation humaine, précurseur de l'intelligence artificielle générative telle qu'on la connaît actuellement, a été présenté par Joseph Weinbaum en 1965 (*WIKIPEDIA - ELIZA 2025*^{WIKIPEDIA - ELIZA 2025 p.29}).
- C'est le lancement de ChatGPT, en novembre 2022, qui a rendu l'IAg accessible au grand public.

¹³. <https://www.w3.org/Talks/WWW94Tim/>

L'IAg se définit comme :

« AI that can create new content such as text, images, or audio by learning patterns from existing data and generating novel outputs that reflect those patterns. » (BENGIO et al. 2025^{BENGIO et al. 2025 p.28}, p. 222)

L'IAg repose sur des modèles avancés comme les grands modèles de langage (*Large Language Models - LLMs*) qui vont générer des contenus à partir d'ensembles de données massives en y identifiant des structures probables. Ces nouveaux outils ouvrent une multitude de possibilités pour soutenir les internautes dans différentes tâches impliquant entre autres la création, la modification, l'analyse de contenus diversifiés, pour un gain en efficacité étonnant. Mentionnons, par exemple :

- Création de texte : ChatGPT, Gemini et NotebookLM
- Création d'images : DALL-E et Midjourney
- Création de vidéos : Synthesia

L'IAg fait partie maintenant de la vie de nombreuses personnes, parfois de manière consciente, en utilisant certaines IAgS telles que ChatGPT, DALL-E ou Scite. D'autres fois l'usage n'est pas conscient, comme l'IAg s'intègre de plus en plus de manière transparente dans différents logiciels d'application. C'est le cas, par exemple, de la dernière version du correcteur Antidote qui offre des fonctionnalités de reformulation de phrase basées sur l'IAg. Ces outils ouvrent la porte à des usages variés pour, par exemple, la création de contenus (textes, images, vidéos, etc.), la synthèse de textes ou la recherche documentaire.

La littérature fait état de l'intégration de l'IAg dans différents secteurs, comme en santé, dans le domaine du marketing ou de l'éducation. Des recherches portent entre autres sur l'utilisation de l'IAg dans le domaine de l'enseignement supérieur et présentent des usages tant des personnes étudiantes, par exemple pour soutenir la recherche documentaire ou le processus d'idéation, que des personnes enseignantes qui peuvent s'en servir, notamment en soutien à la création d'activité pédagogique.

Malgré tous les avantages de l'IAg, il y a certains risques à garder à l'esprit que l'on peut regrouper en trois catégories (BENGIO et al. 2025^{BENGIO et al. 2025 p.28}) qui peuvent aussi s'appliquer à l'IAg :

- Risques liés à une utilisation malicieuse
- Risques liés à un mauvais fonctionnement
- Risques systémiques

1 Risques liés à une utilisation malicieuse

L'IAg peut être exploitée à de fins malicieuses comme, par exemple, pour générer de fausses informations pour frauder des gens ou les manipuler. Les hypertrucages qu'elle permet de créer peuvent en effet induire en erreur tant ils peuvent être réalistes.

2 Risques liés à un mauvais fonctionnement

L'IAg peut, de manière non intentionnelle, générer du biais provenant des corpus de connaissances utilisés. Elle peut aussi produire des informations qui sont fausses. Ces outils ne comprennent en effet pas réellement le contenu généré et les requêtes. Leur fonctionnement se base sur le calcul de ce qui est le plus probable en fonction des requêtes par rapport aux bases de connaissance exploitées. Ces outils sont ainsi faillibles et peuvent se tromper (VANGRUNDERBEECK et al. 2024^{VANGRUNDERBEECK et al. 2024 p.29}, p. 5)! Il faut donc demeurer très critique dans leur utilisation par rapport aux contenus générés. L'IAg est une source d'information parmi d'autres dont il faut valider l'information, comme on le ferait avec d'autres sources. Nous reviendrons dans un prochain module sur la notion de la fiabilité de l'information. L'usage efficace de l'IAg repose aussi sur le développement d'une littératie de l'intelligence artificielle qui, en sus des questions de son usage éthique et responsable, comprend une bonne connaissance de son mode de fonctionnement ainsi que de l'« art » de la *rédictique*^{p.26} (*prompt engineering*).

3 Risques systémiques

Le déploiement rapide et à grande échelle de l'IAg provoque des changements à plusieurs niveaux. Par exemple, certains secteurs du marché de l'emploi peuvent être bousculés du fait de l'automatisation qu'elle rend possible de certaines tâches.

La question des droits d'auteur et de la propriété intellectuelle est aussi un enjeu important par rapport à l'utilisation de l'IAg. D'une part, l'éthique de certains outils d'IAg est remise en question comme ils utilisent pour s'entraîner des données sans s'assurer d'avoir les droits pour le faire. En fait, les données que l'on injecte soi-même lorsque l'on utilise une IAg peuvent être exploitées à des fins d'entraînement par cette dernière; il ne faut donc pas y déposer des informations personnelles ou confidentielles. D'autre part, il est aussi légitime de se demander à qui appartiennent les droits des contenus générés par l'IAg et si l'usage de l'IAg est accepté dans certains contextes. Dans le cadre des études universitaires, par exemple, il est important que les personnes étudiantes s'assurent qu'elles peuvent utiliser l'IAg dans leurs travaux afin de ne pas que leur travail soit reconnu comme du plagiat.

La problématique environnementale fait aussi partie des risques systémiques de l'IAg. Bien que l'IAg peut être utilisée pour participer à trouver des solutions à des enjeux environnementaux, elle est aussi une grande consommatrice d'énergie, d'une part, et d'eau, d'autre part. Sa consommation vient notamment des centres de données utilisés pour entreposer et traiter les données nécessaires à l'IAg. Un centre de données peut consommer autant d'énergie qu'une petite ville, certains des plus grands centres situés en Virginie, par exemple, pouvant consommer l'équivalent de 25 000 résidences (RADIO-CANADA INFO (réalisateur) 2025^{RADIO-CANADA INFO (réalisateur) 2025 p.29}). Le nombre total de centres de données exploités pour l'intelligence artificielle générative n'est pas connu, il s'agit d'un secret bien gardé! On estimait par exemple à la fin 2024 qu'il y avait 239 centres de données au Canada pour tous types d'usage (et non uniquement pour l'intelligence artificielle générative) (GOUVERNEMENT DU CANADA 2024^{GOUVERNEMENT DU CANADA 2024 p.28}). Ces centres de données consomment de l'eau directement pour le refroidissement de leurs ordinateurs ainsi qu'indirectement lors de la production de l'énergie consommée. Cette consommation n'est pas négligeable : certaines études montrent qu'une interaction avec ChatGPT peut consommer 2 litres d'eau (RADIO-CANADA INFO (réalisateur) 2025^{RADIO-CANADA INFO (réalisateur) 2025 p.29}) et certains estiment qu'il y a plus d'un milliard de requêtes envoyées à ChatGPT sur une base quotidienne. Une utilisation responsable de ces outils tient pour acquise une conscientisation sur ces impacts environnementaux; il faut se demander : « Est-ce bien le bon outil? ».

4.2. Culture informationnelle et numérique sur le Web

Le Web, par les plateformes et services qu'il propose, vient modifier la *culture numérique* ^{p.25} et *informationnelle* ^{p.25} des internautes. Le Web 2.0, par exemple, en facilitant la création de contenu et la participation, amène les internautes qui y contribuent à développer une culture informationnelle basée sur le partage de l'information. Certains disent que le Web 2.0 a ainsi démocratisé la création du contenu, une affirmation qu'il nous semble important de nuancer comme un *fossé numérique* ^{p.25} existe encore pour certaines populations, auquel s'ajoute une certaine *fracture sociale*. En effet, même en ayant accès au Web, ce ne sont pas tous les internautes qui sont prêts et sont en mesure de

participer activement à la création de contenu. Li & Bernoff (2011)^{LI 2011 p.28} proposent une échelle décrivant **sept niveaux d'interaction avec le Web**, l'échelon 1 étant celui avec le moins d'interaction et l'échelon 7, celui avec l'interaction la plus intense :

Échelon	Profil	Caractéristiques
7	Créateur	<ul style="list-style-type: none"> • Publie un blogue • Publie son propre site Web • Télécharge des vidéos qu'il a créés • Télécharge de la musique qu'il a créée • Publie des articles qu'il a écrit
6	Causeur	<ul style="list-style-type: none"> • Met à jour son statut sur un réseau social • Publie des mises à jour sur Twitter
5	Critique	<ul style="list-style-type: none"> • Publie des évaluations/critiques de produits ou de services • Commente sur le blogue d'une autre personne • Contribue à des forums en ligne • Contribue à des articles sur un wiki
4	Collectionneur	<ul style="list-style-type: none"> • Utilise les fils RSS • "Vote" pour des sites Web • Ajoute des étiquettes à des pages Web ou des photos
3	Social	<ul style="list-style-type: none"> • Possède un profil sur un réseau social • Visite des réseaux sociaux
2	Spectateur	<ul style="list-style-type: none"> • Lit des blogues • Écoute des baladodiffusions • Regarde des vidéos d'autres utilisateurs • Lit des forums en ligne • Lit des évaluations/critiques de produits • Lit des gazouillis
1	Inactif	<ul style="list-style-type: none"> • Aucun des éléments présentés ci-dessus

Social Technographics ladder (Li & Bernoff, 2011, p.43, notre traduction et représentation)

Plusieurs **enjeux** peuvent être soulevés en lien avec la culture informationnelle et numérique sur le Web dont les quatre suivants :

Enjeu 1 : Journalisme citoyen

On observe, par l'accessibilité de plus en plus grande des plateformes Web, l'apparition d'un « journalisme citoyen » où l'individu impliqué dans certains événements se sert des plateformes sociales pour diffuser ses pensées sur l'actualité, en parallèle des médias de masse traditionnels. Le message ainsi partagé contourne les filtres potentiels des médias de masse. Ceci ne veut toutefois pas dire que l'information ainsi publiée est exempte de biais; cette forme d'auto-publication, en l'absence de validation, *transfère le rôle de validation au lecteur* qui doit pouvoir distinguer les fausses nouvelles des vraies.

Enjeu 2 : Réutilisation de l'information

La culture de partage de contenu s'accompagne de pratiques de « remixage » où les internautes utilisent et réutilisent les contenus. Cette réutilisation ne se fait pas toujours dans le respect de la propriété intellectuelle et des droits des auteurs. Pour certains, les lois sur les droits d'auteur sont trop restrictives et viennent en contradiction avec cet esprit de partage. On a ainsi vu au fil des ans se développer des cadres moins restrictifs que les lois sur le droit d'auteur afin de faciliter ce « remixage » et respecter ainsi l'esprit du Web 2.0. Mentionnons à cet égard le Copyleft¹⁴ (gauche d'auteur) et les autorisations Creative Commons¹⁵ qui permettent aux auteurs d'accorder certains droits de réutilisation de leur contenu. N'en demeure pas moins que cette évolution des droits d'auteur doit, à notre avis, s'accompagner d'une sensibilisation des internautes sur le respect des droits d'auteur.

Enjeu 3 : Identité numérique^{p.25} et réputation numérique^{p.27}

Un enjeu important réside dans les traces laissées par l'internaute sur le Web qui concurrent à lui construire une, voire des *identités numériques*, ce qui n'est pas sans conséquence. Que ce soit volontairement ou non, tout internaute laisse sur le Web des traces, en commentant des blogues, en mettant des photos sur le Web, en ayant un profil public sur Facebook, traces qui pourront servir à la communauté pour se faire une opinion sur lui.

Cette *réputation numérique* peut parfois les desservir s'ils ne sont pas attentifs aux traces qu'ils laissent. De nombreuses histoires existent d'Internautes n'ayant pas obtenu un emploi, ou ayant perdu leur emploi, pour des photos disgracieuses sur le Web ou pour avoir émis une opinion négative sur son employeur sur la place publique. Le phénomène n'est pas nouveau; le Web 2.0 vient simplement l'amplifier et le compliquer.

Les difficultés proviennent en partie de l'évolution de la notion de sphère publique et de sphère privée. Certains internautes perçoivent comme privées certaines plateformes sans se rendre compte que, comme elles sont ouvertes, l'information qui s'y trouve devient publique.

Ceci se complique du fait que l'information est facilement repérable et récupérable sur le Web et circule très rapidement. Les traces sont presque permanentes. Le "droit à l'oubli" n'est pas un droit automatique dans le Web actuel et l'internaute qui veut rétablir sa réputation numérique n'a pas la tâche facile; des compagnies offrent même maintenant des services afin d'effacer les traces.

À l'opposé, la conscience des auditoires multiples qu'ont certains internautes peut les amener à développer des « personnalités numériques » multiples modulant, selon les occasions, les traces laissées. Un adolescent ayant dans son profil Facebook ses parents comme amis et en étant conscient – ce qui n'est pas toujours le cas – n'y interviendra pas de la même manière qu'un autre inconscient de leur présence ou ne les ayant pas inclus dans son réseau social. Un même individu pourrait ainsi avoir plusieurs « visages » 2.0 rendant difficile à l'observateur extérieur l'identification de sa vraie personnalité.

Enjeu 4 : Effet générationnel

Comme dernier enjeu, nous pouvons penser aux *différences générationnelles* qui existent dans l'utilisation des médias sociaux. Force est de constater, comme le montrent certaines études sur l'utilisation des technologies de l'information, que différentes générations utilisent le Web différemment. En fait, ce n'est pas uniquement le Web, mais les technologies de l'information en général. Le regard que l'on porte sur ces dernières peut entre autres être influencé par notre génération. Par exemple, les personnes nées avant l'apparition de la téléphonie mobile auront eu comme première utilisation des téléphones un contexte très privé (à la maison), les téléphones étant fixes. Cette conception de l'usage d'un téléphone pourrait se révéler très différente pour quelqu'un n'ayant connu que la téléphonie mobile. On retrouve aussi cette potentielle différence de perspective du numérique entre les personnes *natives numériques*^{p.26} (*digital natives*) et les personnes dites *immigrantes numériques*^{p.25} (*digital immigrants*).

14. <https://www.gnu.org/licenses/copyleft.fr.html>

15. <https://creativecommons.org/licenses/?lang=fr>

Se côtoieront dans la société des gens ayant une culture et des pratiques informationnelles différentes ainsi que des attentes diversifiées. Les organisations doivent en prendre conscience et évaluer le meilleur moyen de rejoindre ces différences.

5. Données sur le Web

« Les professionnels de l'information, dans leur pratique aujourd'hui, doivent considérer non seulement les documents et l'information, mais aussi différents types de données. Cela aura un impact également sur le développement de la discipline des sciences de l'information, sur la recherche qui est menée et sur la formation offerte. »

(Source : Da Sylva, 2017, p. 30^{DA SYLVA 2017 p.28})

Comme le fait bien ressortir Da Sylva, il est important d'un point de vue de la gestion de l'information de s'intéresser à la question des données sur le Web. On retrouve en effet différents types de données qui demandent aux professionnelles et professionnels de l'information de développer des compétences variées entre autres (Da Sylva, 2017, p. 30) :

- Pour les **données ouvertes** : compétences en *description* des données, en *évaluation* de l'information, en *formation* des usagers
- Pour les **données liées** : compétences en *description*, en *recherche* d'information
- Pour les **données de recherche** : compétences en *organisation* et en *préservation*, aussi en *archivistique* pour planifier leur gestion.

Mais, avant toute chose, il faut s'assurer de bien comprendre ce que sont des données ouvertes, des données liées et des données de recherche.



La question des différents types de données sur le Web est vaste. L'objectif visé dans le cours est de vous introduire à ces dernières. Ainsi, une brève description de ces trois types de données est présentée dans les sections qui suivent, description qui sera complétée par une exploration concrète des trois types de données couverts dans le cadre du *TP Données*. Si la gestion des données vous intéresse, le cours SCI6201 Introduction à la gestion des données¹⁶ est pour vous!

5.1. Données ouvertes

Un premier type de données important à connaître comme professionnelle et professionnel de l'information sont les **données ouvertes**. L'adjectif "ouvertes" rejoint une idée d'ouverture similaire à celle des logiciels ouverts, soit celle de leur *accessibilité publique* que la définition ci-dessous fait bien ressortir :

« C'est typiquement de l'information, surtout sous forme de statistiques, chiffres ou autres formats tabulaires, qui provient d'un organisme public (mais parfois privé) et qui est rendue disponible publiquement sur le Web. »

(Source : Da Sylva, 2017^{DA SYLVA 2017 p.28}, p. 8)

¹⁶. <https://admission.umontreal.ca/cours-et-horaires/cours/sci-6201/>

Il s'agit ainsi de jeux de données produits par des particuliers, des organisations, des gouvernements par exemple et qui sont mis à disposition des internautes publiquement sur le Web. Il est ainsi possible de les consulter, voire de les réutiliser. Parmi les enjeux propres aux données ouvertes, nous retrouvons :

- **Format de données** : Le choix du format pour les données partagées est crucial dans une perspective de partage et de réutilisation. Certains portails de données ouvertes comme *Données Québec* du Gouvernement du Québec (<https://www.donneesquebec.ca/>) vont prescrire certains formats pour les jeux de données, les principaux étant CSV, XML, JSON et GeoJSON (<https://www.donneesquebec.ca/lignes-directrices-sur-la-diffusion-de-donnees-ouvertes/>).
- **Qualité des données** : Sur le portail *Données Québec* se trouvent précisés des critères pour évaluer la qualité des jeux de données, en sus du format, tel que la licence utilisée pour leur partage (<https://www.donneesquebec.ca/exploiter/reperer-la-qualite-des-donnees/>).
- **Licence pour le partage** : Afin de favoriser le partage des données ouvertes, il est important d'utiliser une licence qui le permet. On utilise par exemple, sur le portail *Données Ouvertes* de la Ville de Montréal, la licence *Creative Commons CC-BY 4.0* afin de favoriser le partage et l'utilisation des données ouvertes (<https://donnees.montreal.ca/pages/licence-d-utilisation>).

Vous explorerez plus avant le portail de données ouvertes *Données Québec* dans le cadre du TP Données.

5.2. Données liées

Abordées lorsque le Web sémantique a été présenté, les données liées sont aussi d'intérêt pour les professionnels et professionnelles de l'information. Il s'agit de données auxquelles ont été associées des *métadonnées* ^{p.26} afin de rendre possible leur manipulation et croisement :

« Ensemble de données munies de leurs métadonnées qui, reliées les unes aux autres, constituent une base de données à l'échelle du Web. »

(Source : Office québécois de la langue française, 2013, <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26520043/donnees-liees>)

Différentes technologies sont exploitées avec les données liées pour favoriser leur interconnexion, leur accessibilité et leur réutilisation. Mentionnons, notamment :

- Les URI (Uniform Resource Identifier) qui sont des identifiants uniques pour nommer et identifier des ressources. Les URL (Uniform Resource Locator) et les URN (Uniform Resource Name) sont des URI. Les URL représentent le chemin vers une ressource (par exemple, <https://ebsi.umontreal.ca/>) et les URN sont des identifiants permanents pour des ressources de la forme : `urn:<identifiant de noms>:<chaîne de caractères spécifiques pour l'espace de noms>` (par exemple, `urn:isbn:0875426`).
- Le RDF (Resource Description Framework), développé par le W3C, qui est le langage de base du Web sémantique pour permettre la représentation et l'échange des données sous forme de graphes. Les relations entre les ressources sont décrites en RDF comme des triplets composés d'un sujet (ressource décrite), d'un prédicat (propriété ou relation entre le sujet et l'objet) et d'un objet (une autre ressource ou une valeur).
- SPARQL, qui est un langage de requête pour pouvoir interagir avec des graphes RDF.
- Les graphes sémantiques pour représenter les données en réseau.

St-Germain^{ST-GERMAIN 2017 p.29} (2017, p. 36) résume ainsi les **cinq niveaux de qualité des données liées** proposés par Berners-Lee en 2010 :

1. Les données sont *disponibles sur le Web*, peu importe leur format, à l'aide d'une *licence ouverte*;
2. Les données sont disponibles dans un *format de données structurées* comme, par exemple Excel;
3. Les données sont disponibles dans un *format non-propriétaire* comme, par exemple en format texte délimité par des virgules;

4. Les données sont identifiées à l'aide d'*URI* afin que d'autres puissent les pointer;
5. Les données sont *reliées à d'autres données* pour fournir du contexte.

Des données qui répondent à ces cinq critères peuvent être considérées comme des données liées et ouvertes.

Plusieurs auteurs s'entendent quant à l'**importance des données liées en contexte bibliothéconomique** (voir, par exemple, *Bermès et al. 2013*^{BERMÈS et al. 2013 p.28}, *Da Sylva 2017*^{DA SYLVA 2017 p.28} et *St-Germain 2017*^{ST-GERMAIN 2017 p.29}). St-Germain (2017, p. 37-38) souligne le **gain de visibilité** que l'adoption de données ouvertes et liées peut apporter aux **données des catalogues de bibliothèques** ainsi que leur **enrichissement** du fait de pouvoir les **lier à d'autres ressources externes**. Elle mentionne aussi l'**augmentation de l'interopérabilité** des données par l'adoption d'un **format universel** qui facilitera le partage et la réutilisation. Plusieurs bibliothèques nationales ont commencé à intégrer des données ouvertes et liées, dont la BnF que vous explorerez dans le cadre du *TP Données*.

Selon St-Germain (2017, p. 40) "la place des professionnels de l'information au sein de l'équipe de développement [de projet de données ouvertes et liées] est absolument nécessaire". Les **compétences** que les bibliothécaires et autres professionnels et professionnelles de l'information doivent mobiliser dans un projet de données ouvertes et liées touchent **trois aspects** (Zengenene, Casaros & Meghini, 2014, cités dans St-Germain, 2017, p. 40) :

1. Les *systèmes d'information* comme tels (téléchargement, installation, configuration) et les *formats XML et RDF* qui y sont utilisés ;
2. Les *métadonnées* produites lors du catalogage
3. La *modélisation* afin d'être en mesure de comprendre la structure des données et de pouvoir convertir de manière efficace les données vers RDF

5.3. Données de recherche

Finalement, on ne peut passer sous silence les **données de recherche** qui sont entre autres le produit du mouvement de la *science ouverte*. La **gestion des données de recherche (GDR)**, de leur création jusqu'à leur partage et leur réutilisation présente plusieurs avantages(*Scholars Portal, 2021*^{Scholars Portal, 2021 p.30}, diapositive 6) :

- "Consignation et partage des données de recherche pour faciliter leur découverte
- Préservation et accès à long terme aux données
- Dépôts certifiés conformes aux principes TRUST et FAIR
- Augmentation du rayonnement de la recherche grâce à l'accès aux données et à leur réutilisation
- Métadonnées correctement structurées et enrichies avec des balises de métadonnées pour une meilleure découverte des données de recherche"

De plus, certains organismes subventionnaires peuvent exiger des chercheurs et chercheuses qui obtiennent une subvention de mettre en place un plan de gestion des données de recherche. C'est ainsi dans une idée de partager les données issues de la recherche pour rendre la science plus accessible que le domaine de la GDR s'est développé. Les bibliothèques universitaires ont entre autres un rôle de premier plan en matière de GDR en étant activement impliquées avec la communauté de recherche universitaire pour les aider à développer des plans de gestion des données de recherche et à gérer leurs données de recherche. On a ainsi vu apparaître des postes de "bibliothécaire à la gestion des données de recherche" dans certaines universités. Au Canada, l'Association des bibliothèques de recherche du Canada¹⁷ a lancé le réseau Portage en 2015 afin de coordonner les efforts en matière de GDR. En avril 2021, le réseau Portage a été intégré à la Nouvelle organisation d'infrastructure de recherche numérique (NOIRN) (<https://alliancecan.ca/fr/nouveautes/nouvelles/le-reseau-portage-se-jo-int-la-noirn-pour-continuer-de-faire-progresser-la-gestion-des-donnees-de>).

¹⁷. <https://www.carl-abrc.ca/fr/>

On peut définir les données de recherche comme suit :

« [E]nregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche. »

(Source : OCDE, 2007^{OCDE 2007 p.30}, p. 18)

On retrouve aussi pour les données de recherche les **enjeux** liés aux **formats** et aux **licences** pour le partage. Les **enjeux éthiques** sont aussi bien présents par rapport aux données qui sont partagées.

Vous explorerez un dépôt de données de recherche dans le cadre du *TP Données*.

À noter!



Les concepts de données ouvertes, données liées et données de recherche ne sont pas mutuellement exclusifs. Des données liées peuvent être ouvertes ainsi que des données de recherche peuvent être ouvertes.

En sus de ces trois types de données, on retrouve aussi le concept de données massives (méga-données) qui désigne un « [e]nsemble d'une très grande quantité de données, structurées ou non, se présentant sous différents formats et en provenance de sources multiples, qui sont collectées, stockées, traitées et analysées dans de courts délais, et qui sont impossibles à gérer avec des outils classiques de gestion de bases de données ou de gestion de l'information. » (Office québécois de la langue française, 2020, <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26507313/megadonnees>)

6. Importance de la normalisation



Matériel adapté de Marcoux, Yves. 2007. Notes de cours du SCI6052 Information documentaire numérique. EBSI, FAS, UdeM.

Les **préoccupations liées au format** présentées lors du deuxième cours sont aussi présentes dans le contexte du Web. Dans ces espaces virtuels, des documents numériques naissent, circulent et s'échangent et on doit se préoccuper :

- De l'**accessibilité** des documents : un document sur le Web doit être dans un format qui est compris par une communauté de lecteurs la plus grande possible;
- De la **pérennité** des documents : le document doit demeurer accessible dans le futur;
- De la **réutilisabilité** des documents : on peut récupérer un document archivé et continuer à travailler dessus pour créer un nouveau document.

L'adoption de **formats normalisés** est un moyen de répondre à ces préoccupations. Pour la gestion de l'information, la normalisation est particulièrement cruciale en raison de l'importance de l'accessibilité, la pérennité et la réutilisabilité des documents. On y retrouve trois types de **spécifications**^{P.27} pertinentes :

- Des spécifications d'un *format* (par exemple HTML)
- Des spécifications d'un *protocole* (par exemple HTTP)
- Des procédures pour les *humains* (par exemple ISO-15489 Records management)

Ces documents normatifs (les « spécifications » elles-mêmes) sont destinés aux humains. Ils décrivent un format ou un protocole normalisé et indiquent comment développer des produits ou systèmes qui respectent la norme, et sont donc « interopérables ».

En gestion de l'information numérique, les deux premiers types (formats et protocoles) sont les plus déterminants sur l'efficacité et l'efficience des systèmes automatisés. Si on vise la réutilisation de l'information et l'*interopérabilité* ^{p.26}, il importe que les systèmes mis en place respectent des normes pour les formats et les protocoles.

Niveaux de normalisation

Il existe différents **niveaux de normalisation**. Une **norme officielle** (ou *de jure*) peut être définie sur le plan national ou sur le plan international. Les normes nationales s'alignent aux normes internationales. Le tableau qui suit présente quelques exemples d'instances de normalisation et de normes qu'elles prennent en charge.

Niveau	Normes internationales	Normes nationales
<i>Définition</i>	Normes définies et sanctionnées par une instance officielle internationale de normalisation	Normes définies et sanctionnées par un organisme de normalisation national
<i>Exemples d'instances</i>	<ul style="list-style-type: none"> • ISO (<i>International Organization for Standardisation</i>) • W3C (<i>World Wide Web Consortium</i>) • Unicode Consortium • IEEE (<i>Institute of Electrical and Electronics Engineers</i>) • IEC (<i>International Electrotechnical Commission</i>) • ECMA International (<i>European Computer Manufacturers Association</i>) • ITU ou UIT (<i>Union internationale des télécommunications</i>) 	<ul style="list-style-type: none"> • ANSI (<i>American National Standard Institute</i>) (États-Unis) • NISO (<i>National Information Standards Organization</i>) (États-Unis) • CSA (<i>Canadian Standards Association</i>) • SCC (<i>Standards Council of Canada</i>)
<i>Exemples de normes</i>	SGML, HTML, XHTML, XML (W3C)	ASCII (ANSI), Z39.50 (NISO)

Normes internationales et nationales

Être une norme officielle (*de jure*) n'est pas tout, la **reconnaissance** et l'**adoption** dans la société sont encore plus importantes. Les spécifications normalisées sont **plus stables** (changent moins vite) que les spécifications guidées par des intérêts commerciaux, ce qui est intéressant pour le long terme tout particulièrement en gestion de l'information.

Karl Dubost (cité dans *Dumais, 2003* ^{DUMAIS 2003 p.28}) amène des arguments très convaincants quant à l'importance des normes libres et ouvertes et tout particulièrement à leur impact social :

« L'impact social, à mon avis, est l'impact le plus important des normes libres. Ne pas être prisonnier d'une entité commerciale décidant de l'avenir de votre information, de vos données. Avoir le choix en toutes circonstances, la possibilité de changer, de faire autrement, voilà l'enjeu des normes libres. »

L'importance des normes libres et ouvertes est aussi soulignée par l'*OFE (2008)*^{OFE 2008 p.28} :

« In conclusion, choosing open standards is highly strategic. Their benefits and positive impact are debated and seen at the highest decision making levels. Interoperability is a major requirement for the ICT sector as societies, governments and industry increasingly move towards global collaboration and integration. »

7. Ressources en lien avec le cours

Matériel de cours

- *Notes de cours [cf. sci6005_a2025_c5.pdf]*

Lectures complémentaires facultatives (pour aller plus loin si désiré)

- *DA SYLVA, 2017*^{DA SYLVA 2017 p.28}
- *DUFOUR & CADIEUX, 2013*^{DUFOUR 2013 p.28}
- *DUMAIS, 2003*^{DUMAIS 2003 p.28}

Glossaire



Culture de l'information

"[U]n système de valeurs, d'attitudes et de comportements, de connaissances et d'aptitudes qui conduisent non seulement à un usage intelligent et approprié de l'information externe, mais surtout à contribuer à la diffusion et à la bonne utilisation de l'information tant externe qu'interne (ou produite/reconfigurée par soi-même). Donc une culture de partage et d'enrichissement collectif. "

(Menou M. (2008). *Culture de l'information*. In Cacaly S., Le Coadic Y.-F., Pomart P.-D., Sutter E. (dir.), Dictionnaire de l'information, 3e éd., Paris, Armand Colin, p. 65-67)

Culture numérique

"Ensemble de connaissances et de pratiques se rapportant aux nouvelles technologies de l'information et des communications."

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2012, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26540982 - définition de Culture Montréal, 2015)

Fossé numérique

"Écart existant entre les pays développés et les pays en voie de développement, en matière d'accès aux technologies de l'information et de la communication."

Notes

Le fossé numérique peut exister non seulement entre les pays industrialisés et les pays en développement, mais aussi à l'intérieur même de chaque pays. Plus ce fossé se creuse, plus il augmente l'écart entre les femmes et les hommes, entre les riches et les pauvres, entre les jeunes et les vieux, entre les gens qui savent lire et écrire et les analphabètes, et entre les populations urbaines et les populations rurales.

Le fossé numérique est associé notamment à l'aggravation des inégalités d'équipement en ordinateur personnel et Internet, et aux disparités croissantes selon le niveau de revenu, la localisation géographique et l'appartenance ethnique."

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2005, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8360705)

Identité numérique

"Ensemble des traces numériques laissées de manière active par une personne dans Internet et qui fournissent des renseignements à son sujet. "

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2017, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26508663)

Immigrant numérique

"Personne née avant 1974, qui n'a pas grandi dans un monde numérique, mais qui a appris à s'adapter à ce nouvel environnement, en intégrant l'ordinateur et les nouvelles technologies à sa façon de travailler."

Note

On considère que les natifs numériques ont une connaissance de l'informatique et des nouvelles technologies (le langage du numérique) comme si c'était leur langue maternelle (donc ils la parlent sans accent), alors que les immigrants numériques ont une connaissance de celles-ci comme s'ils parlaient une langue seconde (tels des immigrants parlant une autre langue que la leur, mais avec un accent). Chez ces derniers, qui ont un pied dans le passé (celui d'avant l'informatique), cet accent, tel celui d'un immigrant, est maintenu à travers certains de leurs comportements, par exemple, imprimer leurs courriels."

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2004, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8358524)

Interopérabilité

"Capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner accès à leurs ressources de façon réciproque."

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2007, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8395687)

Métadonnées

"Une métadonnée est, à proprement parler, une donnée à propos de données (le préfixe grec meta indiquant l'autoréférence). En documentation, une métadonnée est une donnée servant à définir ou décrire d'autres données (par exemple un document). La valeur sémantique d'une métadonnée est généralement décrite et définie dans un schéma de métadonnées. À titre d'exemple, une date associée à un document est une métadonnée servant à déterminer la date de publication, de dépôt légal, de copyright, de création, de dernière modification, etc. Les métadonnées peuvent être de nature descriptive, structurelle ou administrative." (Arsenault, Clément, Terminologie en sciences de l'information - EBSI / Métadonnées (Metadata)¹⁸)

Natif numérique

"Personne née après 1974, qui a grandi dans un monde numérique, qui est familière avec les ordinateurs, les jeux vidéo et Internet, et qui passe une grande partie de sa vie en ligne."

Note

On considère que les natifs numériques ont une connaissance de l'informatique et des nouvelles technologies (le langage du numérique) comme si c'était leur langue maternelle (donc ils la parlent sans accent), alors que les immigrants numériques ont une connaissance de celles-ci comme s'ils parlaient une langue seconde (tels des immigrants parlant une autre langue que la leur, mais avec un accent). Chez ces derniers, qui ont un pied dans le passé (celui d'avant l'informatique), cet accent, tel celui d'un immigrant, est maintenu à travers certains de leurs comportements, par exemple, imprimer leurs courriels."

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2004, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8358516)

Rédactique (prompt engineering)

"Ensemble des techniques de rédaction orientées vers la formulation précise de requêtes envoyées à un système d'intelligence artificielle, en particulier à un grand modèle de langage, en vue d'en améliorer le fonctionnement ou d'en exploiter efficacement les capacités pour obtenir des résultats plus pertinents."

(Office québécois de la langue française, 2023, <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26571214/redactique>)

¹⁸. <https://clip.ebsi.umontreal.ca/terminologie/sci/fr/index.php?tema=101&/metadonnee>

Réputation numérique

"Réputation basée sur la perception que des internautes ont du double numérique d'une personne."

(Office Québécois de la Langue Française, Grand dictionnaire terminologique, 2017, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26508673¹⁹)

Spécification

Une « spécification » est la description d'un ensemble de règles et/ou façons de faire conventionnelles.

¹⁹. http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26508673

²⁰. http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26508663

Bibliographie



[BENGIO et al. 2025] BENGIO, Y.; MINDERMAN, S.; PRIVITERA, D.; BESIROGLU, T.; BOMMASANI, R. et al.. *International AI Safety Report 2025* (No. DSIT 2025/001), 2025. <https://www.gov.uk/government/publications/international-ai-safety-report-2025/international-ai-safety-report-2025>

[BERMÈS et al. 2013] BERMÈS, Emmanuelle; ISAAC, Antoine (coll.); POUPEAU, Gautier (coll.). *Le Web sémantique en bibliothèque*. Paris : Éditions du Cercle de la Librairie, 2013.

[BUSH 1945] BUSH, Vannevar. *As We May Think*. *The Atlantic Monthly*, 176(1), 112-114, 1945 <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

[DA SYLVA 2017] DA SYLVA, Lyne. *Les données et leurs impacts théoriques et pratiques sur les professionnels de l'information*. *Documentation et bibliothèques*, 63(4), 5-34. 2017. <https://www.erudit.org/fr/revues/documentation/2017-v63-n4-documentation03290/1042308ar/>

[DUFOUR 2013] DUFOUR, Christine; CADIEUX, Stéphanie. *Web 2.0 et milieux documentaires : qu'en retenir ?*. *Archives*, 45(1), 27-46. 2013-2014. https://www.archivistes.qc.ca/revuearchives/vol45_1/45_1_dufour_cadieux.pdf

[DUMAIS 2003] DUMAIS, Michel. *Technologie : Les normes libres, l'enjeu du moment*. *Le Devoir*, juillet 2003. <http://www.ledevoir.com/societe/science-et-technologie/32207/technologie-les-normes-libres-l-enjeu-du-moment>

[GOUVERNEMENT DU CANADA 2024] Gouvernement du Canada, Régie de l'énergie du Canada. *ONÉ - Aperçu du marché: Le développement de l'intelligence artificielle contribue considérablement à l'augmentation constante de la demande d'énergie des centres de données*, 3 octobre 2024 <https://www.cer-rec.gc.ca/fr/donnees-analyse/marches-energetiques/aperçu-marches/2024/aperçu-marche-developpement-de-lintelligence-artificielle-contribue-considerablement-laugmentation-constante-demande-denergie-des-centres-de-donnees.html>

[HERN 2014] HERN, Alex. *World's most delayed software released after 54 years of development*. *The Guardian*, 6 juin 2014 <https://www.theguardian.com/technology/2014/jun/06/vapourware-software-54-years-xanadu-ted-nelson-chapman>

[LANKES et al. 2007] LANKES, R. David; Joanne SILVERSTEIN; NICHOLSON, Scott. *Participatory Networks : The Library as Conversation*. *Information Technology and Libraries*, 26(4), 17-33. 2007. <https://ejournals.bc.edu/index.php/ital/article/view/3267>

[LI 2011] LI, Charlene; BERNOFF, Josh. *Groundswell: Winning in a world transformed by social technologies*. Boston : Harvard Business Press, 2011.

[MAIGNIEN 2014] MAIGNIEN, Yannick. *Chapitre 5. Les enjeux du web sémantique*. IN *Pratiques de l'édition numérique [en ligne]*, p. 77-93. Montréal : Presses de l'Université de Montréal, 2014 <http://books.openedition.org/pum/320>

[O'LEARY, O'LEARY & O'LEARY 2019] O'LEARY, Timothy; O'LEARY, Linda I.; O'LEARY, Daniel A. *Computing Essentials 2019*. New-York : McGraw-Hill Higher Education, 2019.

[OFE 2008] OPENFORUM EUROPE (OFE). *The Importance of Open Standards in Interoperability*. OFE Onepage Brief, (1), 2008-10-31. <https://openforumeurope.org/publications/the-importance-of-open-standards-in-interoperability-ofe-onepage-brief-no-1-31-10-08/>

[RADIO-CANADA INFO (réalisateur) 2025] Radio-Canada Info (Réalisateur). *Intelligence artificielle: Quand les centres de données grugent nos ressources | Découverte* [Enregistrement vidéo], 5 avril 2025. <https://www.youtube.com/watch?v=q0q3OY20BHQ>

[ST-GERMAIN 2017] ST-GERMAIN, Marielle. *Étapes pour le développement d'un projet de données ouvertes et liées en bibliothèque*. *Documentation et Bibliothèques*, 63(4), 35-45. 2017. <https://www.erudit.org/fr/reviews/documentation/2017-v63-n4-documentation03290/1042309ar.pdf>

[VANGRUNDERBEECK et al. 2024] VANGRUNDERBEECK, P.; DEVILLE, Y.; GUISET, M.; BAUR, M.; BIOT, M.; RAUCENT, B.; WOUTERS, P. *Intégrer l'IA générative dans les stratégies pédagogiques*. Louvain Learning Lab - UCLouvain, 2024. <https://oer.uclouvain.be/jspui/handle/20.500.12279/1089.3>

[WIKIPEDIA - Conférence de Dartmouth 2025] *Conférence de Dartmouth*. 2025. Dans Wikipédia. https://fr.wikipedia.org/w/index.php?title=Conf%C3%A9rence_de_Dartmouth

[WIKIPEDIA - ELIZA 2025] *ELIZA*. 2025. Dans Wikipédia. <https://fr.wikipedia.org/w/index.php?title=ELIZA>

[WIKIPEDIA - Réseau de neurones artificiels 2025] *Réseau de neurones artificiels*. 2025. Dans Wikipédia. https://fr.wikipedia.org/w/index.php?title=R%C3%A9seau_de_neurones_artificiels

Webographie



[**LEINER et al. 1997**] LEINER, Barry M.; CERF, Vinton G.; CLARK, David D.; Kahn, Robert E.; KLEINROCK, Leonard; LYNCH, Daniel C.; POSTEL, Jon; ROBERTS, Larry G.; WOLFF, Stephen. *Brief History of the Internet*. 1997. <https://www.internetsociety.org/internet/history-internet/brief-history-internet/>

[**OCDE 2007**] ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE). *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics*. 2007. <http://www.oecd.org/fr/sti/inno/38500823.pdf>

[**Scholars Portal, 2021**] Scholars Portal. *Modules 101 sur les dépôts de données de recherche - Module 1 : Contexte et objectifs d'apprentissage (Série de modules de formation de Portage)*. 2021. <https://learn.scholarsportal.info/fr/modules-dapprentissage/portage-fr/modules-101-sur-les-depots-de-donnees-de-recherche-module-1/>

[**W3C 2013**] WORLD WIDE WEB CONSORTIUM (W3C). *W3C Semantic Web activity homepage*. 2013. <https://www.w3.org/2001/sw/>

Index



Accessibilité.....	22
Culture informationnelle	16
Culture numérique.....	16
Données de recherche	19
Données liées	19
Données ouvertes	19
Internet.....	5
Normalisation.....	22
Pérennité	22
Ressources en lien avec le cours	24
Réutilisabilité	22
Web	5
Web 2.0 / Web social	5
Web 3.0 / Web sémantique	5

Crédits des ressources



p. 3

<http://creativecommons.org/licenses/publicdomain/4.0/fr/>, johnny_automatic

p. 4

<http://creativecommons.org/licenses/publicdomain/4.0/fr/>, maoriveros