

SCI6052 Information documentaire numérique

Cours 8

Web; formats de documents numériques; normalisation, structuration logique; documents structurés

Au programme aujourd'hui

- ➔ Évolution du Web
- ➔ Préoccupations liées aux formats de documents numériques (*rappel cours 1*)
- ➔ Normalisation
- ➔ Formats de documents structurés
 - > XML
 - > SGML
 - > HTML

Cours 8 – Objectifs visés, matériel associé et évaluation

Examen final



OG1 Comprendre les principaux volets de la gestion d'information documentaire numérique

OS d) Expliciter les avantages et désavantages de la structuration de l'information dans différents contextes

Compétences à développer :

- Définir ce qu'est un format de document structuré (acétate 29)
- Distinguer les principaux formats structurés et leurs avantages/désavantages (XML, SGML, HTML et XHTML) (acétates 30-37)
- Pouvoir expliquer les enjeux et les opportunités que représentent les formats structurés pour les milieux documentaires (acétates 38-39)

OS i) Expliquer les avantages et les désavantages des sites Web statiques pour la gestion d'information documentaire numérique

Compétences à développer :

- Pouvoir expliquer au moins un enjeu lié à l'évolution du Web (acétates 6-8)
- Distinguer les différentes évolutions du Web (Web « classique », Web 2.0, Web 3.0) (acétates 11, 18-20)
- Donner des exemples de plateformes 2.0 (acétates 15-17)

Image familière?

Omniprésence des technologies ...

Although everybody is physically present but not mentally

Vie professionnelle

Vie amoureuse

Vie familiale

<https://www.creators.com/read/speed-bump/05/12/8718>

2 mai 2012

Études

<https://www.creators.com/read/speed-bump/03/12/8569>

5 mars 2012

<https://www.creators.com/read/speed-bump/04/12/8536>

28 avril 2012

... dans toutes les
sphères de nos vies

<https://www.creators.com/read/speed-bump/03/12/8587>

26 mars 2012

Principaux enjeux

Nouveaux modèles de circulation de l'information

- ➔ **Contenus générés par les citoyens « branchés »**
 - Processus d'auto-publication en rupture avec le modèle traditionnel de l'édition
 - Journalisme citoyen en parallèle des médias de masse traditionnels
- ➔ **Réutilisation et partage des contenus**
 - Culture et technologies mettant en valeur le partage (par ex., le mouvement du libre accès)
 - Évolution des droits pour faciliter le partage (par ex., *Copyleft* et *Creative Commons*)
- ➔ **Phénomène de la longue traîne**
 - Modèle économique où la multiplicité des niches remplace les marchés de masse
- ➔ **Fractures numérique et sociale**

Principaux enjeux

Identité et réputation numériques

➔ Des problématiques multiples

- Sphères publique et privée en évolution
- Traces difficiles à contrôler, dispersées sur le Web
- Permanence des traces *versus* Droit à l'oubli
- Vitesse de propagation et effet réseau
- « Personnalités » multiples construites pour des audiences diversifiées

<https://www.creators.com/read/spe-ed-bump/11/11/8444>
[28 novembre 2011]

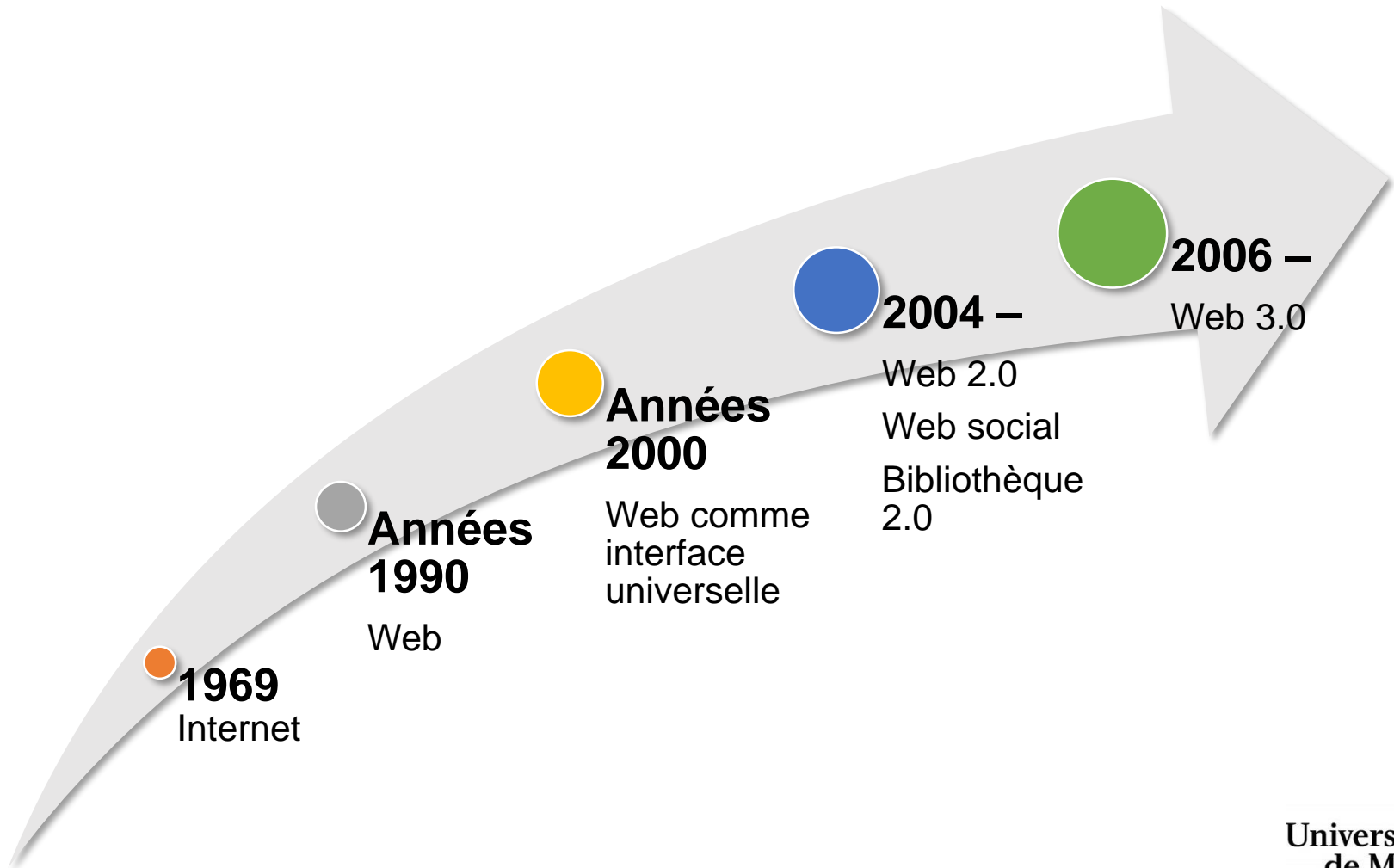
Principaux enjeux

Effet générationnel

- ➔ Différences générationnelles dans l'utilisation des médias sociaux
 - Par ex., la génération Y (18-33 ans) est généralement plus active
- ➔ Présence dans la société de gens ayant une culture et des pratiques informationnelles différentes ainsi que des attentes diversifiées

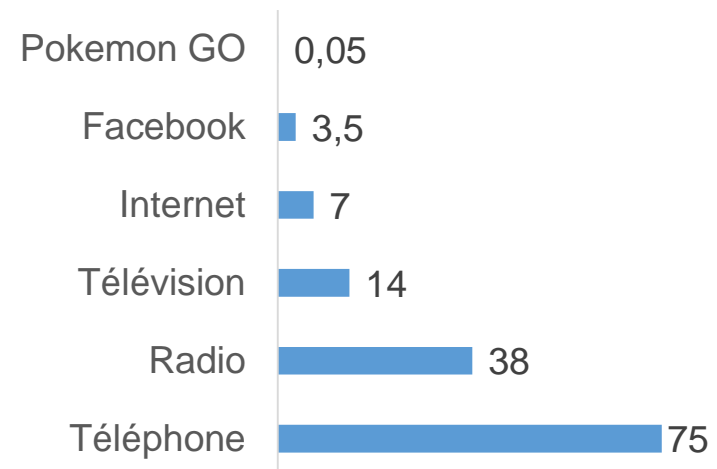
<https://www.creators.com/read/speed-bump/09/12/8772> [18 sept. 2012]

Un peu d'histoire [1/2]



Un peu d'histoire [2/2]

Nombre d'années pour atteindre 50 millions d'utilisateurs



Source : @Guglielminetti, 14 oct. 2017 (via @pixiolabs)
<https://twitter.com/Guglielminetti/status/919245063967911937>

Web 2.0 / Web social

Phénomène **technologique** et **social**

« [...] a supposed **second generation of Internet-based services** such as **social networking sites, wikis, communication tools**, and **folksonomies** that emphasize **online collaboration** and **sharing among users**. » [O'Reilly & Dougherty]

Web 2.0 / Web social Évolution, révolution ~~ou mode passagère?~~

Partage de ressources

Instagram (22-10-2019) [[source](#)]

- **1+ milliard** d'utilisateurs actifs mensuellement
- **500+ millions** d'activités par jour

YouTube (22-10-2019) [[source](#)]

- **Un milliard d'heures** vues par jour
- Plus **d'un milliard** d'utilisateurs

LinkedIn (22-10-2019) [[source](#)]

- Plus de **645 millions** d'utilisateurs

Facebook (22-10-2019) [[source](#)]

- **2,41 milliards** d'utilisateurs actifs mensuellement

Réseaux sociaux

Rédaction collaborative

Wikipedia (22-10-2019) [[source](#)]

- **51 317 159 millions** d'articles
- **296 langues actives**
- **2 148 424 millions** d'articles en français

Blogosphère et micro-blogosphère

Blogpulse (12-01-2012) [[source](#)]

- **182 090 885** blogues identifiés
- **97 084 nouveaux blogues** dans les dernières 24 hrs

Twitter (22-10-2019) [[source](#)]

- **330 millions** d'utilisateurs actifs

Web 2.0 / Web social

Contexte québécois [1/2]

➔ Une population globalement **branchée** et **active** sur les réseaux sociaux

➤ « En 2018, la presque totalité des adultes québécois (95 %) détient au moins un des cinq appareils électroniques suivants : ordinateur (portable ou de table), téléphone intelligent, tablette électronique, bracelet d'exercice connecté et montre intelligente. » (Cefrio. 2018. NETendances 2018 : Portrait numérique des foyers québécois. Volume 9, no 4, p. 7.

<https://cefrio.qc.ca/media/2015/netendances2018-portraitnumeriquefoyersquebecois.pdf>)

➤ « En 2018, 83 % des adultes québécois utilisaient au moins un réseau social dans le cadre de leur utilisation personnelle d'Internet, ce qui représente une hausse notable de 16 points de pourcentage par rapport à 2016. » (Cefrio. 2018. NETendances 2018 : Usage des médias sociaux au Québec. Volume 9, no 5, p. 6. https://cefrio.qc.ca/media/2023/netendances-2018_medias_sociaux.pdf)

Web 2.0 / Web social

Contexte québécois [2/2]

Exemples d'initiatives

Initiatives

But(s)

Projets Wiki de BAnQ

<https://fr.wikipedia.org/wiki/Wikipédia: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/Wikipédia:Cinémathèque québécoise](https://fr.wikipedia.org/wiki/Wikipédia:Cinémathèque_québécoise)

« 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://cafesavoirslibres.com>

« Un collectif de passionnés, souvent bibliothécaires, qui organise des rencontres de partage autour des savoirs libres et des biens communs. »

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

Web 2.0 / Web social

Services & produits-phares [1/3]

Blogues

- ➔ « Site Web ayant la **forme d'un journal personnel**, daté, au contenu antéchronologique et régulièrement mis à jour, où l'internaute auteur peut communiquer ses idées et ses impressions sur une multitude de sujets, en y publiant, **à sa guise**, des textes, **informatifs ou intimistes**, généralement courts, parfois enrichis d'hyperliens, qui appellent **les commentaires du lecteur**. »
(Office québécois de la langue française, 2005)
- ➔ Sans oublier les **micro-blogues**, comme Twitter [<http://twitter.com>]

Sites Wikis

- ➔ « Site Web **collaboratif** où **chaque internaute** visiteur peut **participer à la rédaction** du contenu. »
(Office québécois de la langue française, 2004)
- ➔ Comprend des fonctionnalités pour le suivi des modifications, des espaces de discussion, la révision, etc.
- ➔ Ex.: Wikipedia [<http://www.wikipedia.org/>]

Web 2.0 / Web social

Services & produits-phares [2/3]

Réseaux sociaux

- ➔ Environnements où l'utilisateur **se définit un profil** (personnel ou professionnel) et **réseaute** avec d'autres utilisateurs
 - Facebook [<http://www.facebook.com>] est un site de réseautage personnel très populaire
 - LinkedIn [<http://www.linkedin.com>] est un exemple de site de réseautage professionnel
- ➔ Possibilité d'inclure dans son profil, en sus des informations descriptives, **différents types de fichiers** (images, vidéos, ...)

Partage de ressources

- ➔ Plateformes facilitant la **composition de collections de ressources** diverses et le **partage** de ces dernières; par exemple :
 - Flickr [<http://www.flickr.com>] pour les photos
 - YouTube [<http://www.youtube.com>] pour les vidéos
 - Slideshare [<https://fr.slideshare.net/>] pour les présentations
- ➔ **Mise en commun** des collections **bénéfique** aux communautés visées en autant que les **règles soient respectées** (par ex., droits d'auteurs)

Web 2.0 / Web social

Services & produits-phares [3/3]

Fils de syndication

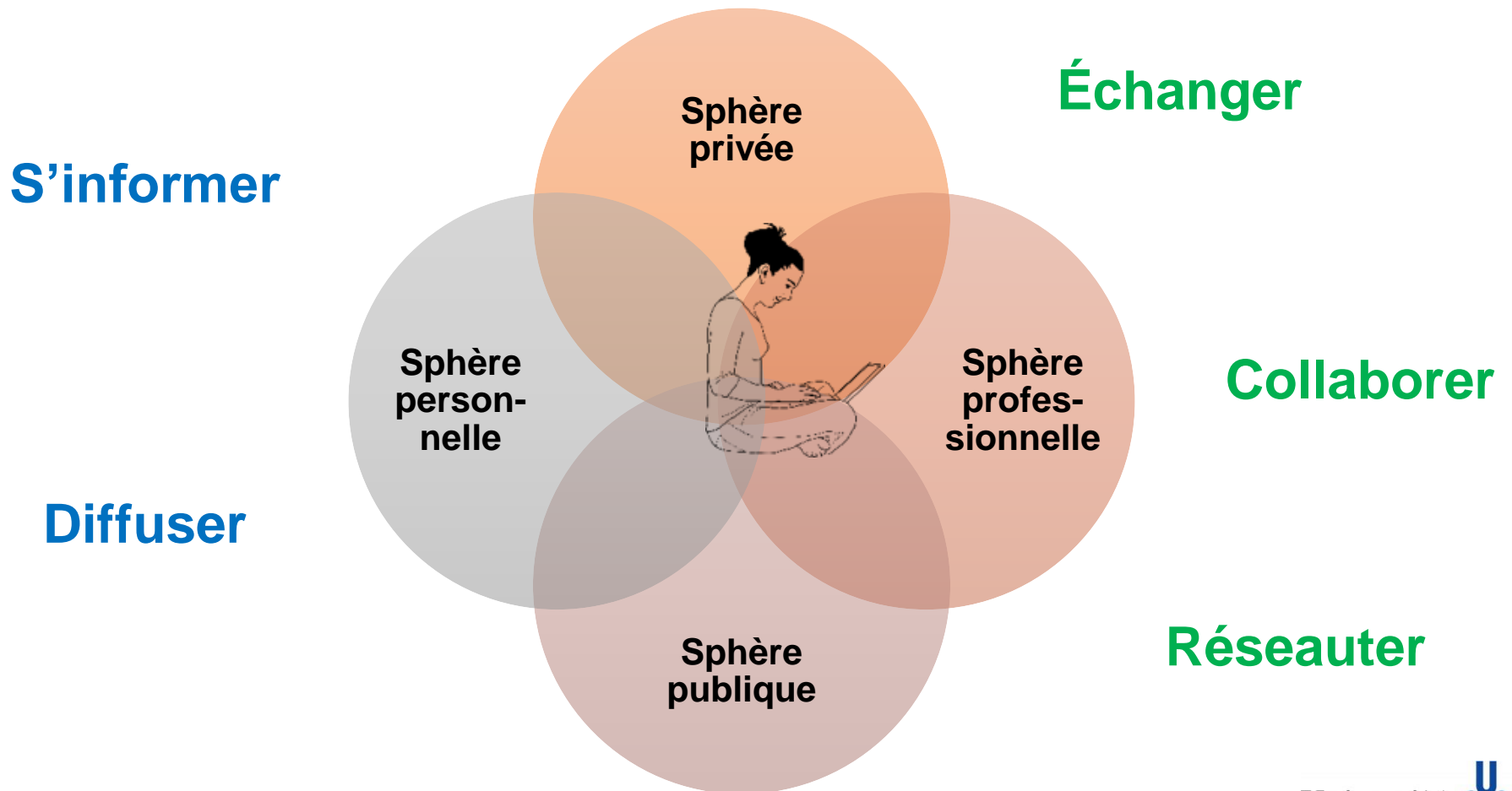
- ➔ Formats facilitant le partage de contenu
 - Permettent aux auteurs et éditeurs d'un site Web de rendre disponibles à la communauté certains **contenus** qui peuvent être **réutilisés** pour intégration dans un autre site
 - Ex.: **RSS** (*Really Simple Syndication*) et **ATOM** - tous deux basés sur XML

Folksonomies

- ➔ Folksonomie = Folks + Taxonomy
 - Mots-clés (étiquettes, descripteurs, *tags*) libres attribués par les utilisateurs à différentes ressources Web
 - Produit de *l'étiquetage social* et *partagé* de ressources sur différentes plateformes du Web 2.0 (plateformes de partage de ressources, blogues, sites Wikis)
- ➔ Facilite le **partage** et le **repérage** des ressources

Web 2.0 / Web social

Utilisations personnelles et organisationnelles



Web 2.0 / Web social

Évolution [1/2]

➔ Web 2.0 : un pas vers le Web sémantique (Web 3.0)

- Folksonomies représentent une *certaine* couche sémantique
- Applications composites (*mashup*) = mise en lien de données diverses possible grâce à l'utilisation d'applications ouvertes (API)
 - Par exemple, *Flickr – Explorer les photos sur une carte* (<https://www.flickr.com/map>) qui met sur une carte les photos géotaguées

Web 2.0 / Web social

Évolution [2/2]

- ➔ **Web 3.0 = Web + couche sémantique (Web sémantique)**
 - Documents « enrichis » compréhensibles non seulement aux humains mais aussi aux machines : vers un « Web de données »
 - Facilitation de la mise en lien des données pour permettre la construction de la connaissance
 - Données liées (*linked data*) : « Ensemble de données munies de leurs métadonnées qui, reliées les unes aux autres, constituent une base de données à l'échelle du Web. » (OQLF, 2013, http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=26520043)
 - Description à l'aide d'un format structuré de données (par ex. RDF – Resource Description Framework)
 - Par exemple
 - Google Knowledge Graph
<https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>
 - Web sémantique à la BNF
http://www.bnf.fr/fr/professionnels/innov_num_web_donnees.html

Normalisation

Préoccupations liées au format *(rappel cours 1)*

➔ **Accessibilité des documents**

- › Le document est dans un format qui est compris par une communauté de lecteurs la plus grande possible

➔ **Pérennité des documents**

- › Le document demeurera accessible dans le futur

➔ **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

Normalisation

Pour la gestion de l'information [1/2]

- ➔ Une « **spécification** » est la description d'un ensemble de règles et/ou façons de faire conventionnelles
- ➔ **Trois types de spécifications pertinentes** en gestion de l'information
 - › Spécifications d'un format (par exemple HTML)
 - › Spécifications d'un protocole (par exemple HTTP)
 - › Procédures pour les humains (par exemple ISO-15489 Records management)
- ➔ Documents **normatifs** (les « spécifications » elles-mêmes) décrivant un format ou un protocole normalisé sont destinés aux humains
 - › Indiquent comment développer des produits ou systèmes qui respectent la norme, et sont donc « interopérables »

Normalisation

Pour la gestion de l'information [2/2]

- ➔ En gestion d'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**, il importe que les systèmes mis en place respectent des normes pour les formats et les protocoles

Normalisation

Niveaux de normalisation

Norme officielle (ou *de jure*) peut être définie à différents niveaux

➔ Normes internationales

- Définies et sanctionnées par une instance officielle internationale de normalisation comme par ex.
 - ISO (*International Organization for Standardisation*)
 - W3C (*World Wide Web Consortium*)
 - Unicode Consortium
 - IEEE (*Institute of Electrical and Electronics Engineers*)
 - IEC (*International Electrotechnical Commission*)
 - ECMA International (*European Computer Manufacturers Association*)
 - ITU ou UIT (Union internationale des télécommunications)

- Ex. W3C : SGML, HTML, XHTML, XML

➔ Normes nationales

- Définies et sanctionnées par un organisme de normalisation national comme, par ex.
 - ANSI (*American National Standard Institute*) (États-Unis)
 - NISO (*National Information Standards Organization*) (États-Unis)
 - CSA (*Canadian Standards Association*)
 - SCC (*Standards Council of Canada*)
- Alignement sur les normes internationales
- Ex. : ASCII (ANSI), Z39.50 (NISO)

Normalisation

Importance de la normalisation

- ➔ 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 : intéressant pour le long terme

Normalisation

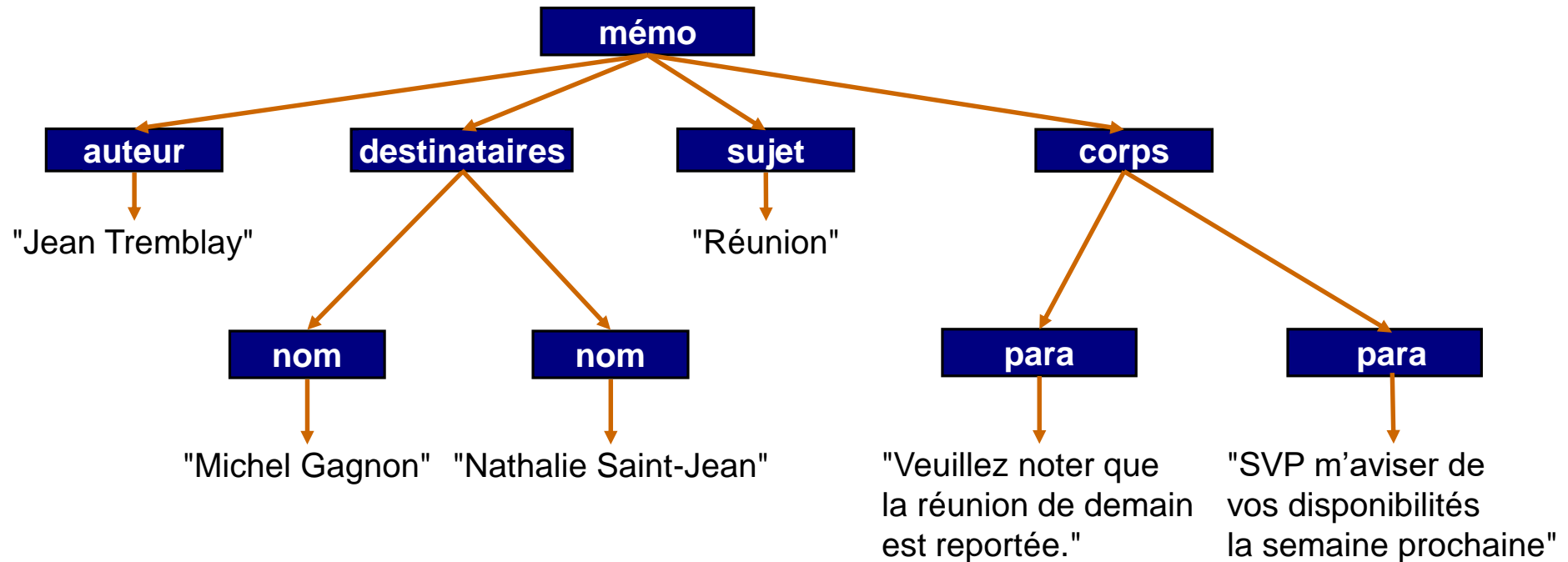
Impact social des normes libres

- ➔ « 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. »

Karl Dubost (cité dans Dumais, Michel. « Technologie : Les normes libres, l'enjeu du moment », Le Devoir, 21 juillet 2003. En ligne <<http://www.ledevoir.com/societe/science-et-technologie/32207/technologie-les-normes-libres-l-enjeu-du-moment>>. (consultée le 2019-10-22))

Documents structurés

Exemple de structure logique : Mémo

**MÉMORANDUM****De:** Jean Tremblay**À:** Michel Gagnon
Nathalie Saint-Jean**Sujet:** Réunion

Veillez noter que la réunion de demain est reportée.

SVP m'aviser de vos disponibilités la semaine prochaine

Documents structurés

Mémo en XML

```

<mémo>
  <auteur>Jean Tremblay</auteur>
  <destinataires>
    <nom>Michel Gagnon</nom>
    <nom>Nathalie Saint-Jean</nom>
  </destinataires>
  < sujet>Réunion</sujet>
  <corps>
    <para>Veillez noter que la réunion de
    demain est reportée.</para>
    <para>SVP m'aviser de vos disponibilités
    la semaine prochaine.</para>
  </corps>
</mémo>

```

MÉMORANDUM

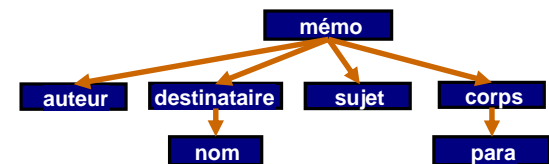
De: Jean Tremblay

À: Michel Gagnon
Nathalie Saint-Jean

Sujet: Réunion

Veillez noter que la réunion de
demain est reportée.

SVP m'aviser de vos disponibilités
la semaine prochaine



Chaînes entre < et > = **Balise**

<nom> = **Balise de début**

</nom> = **Balise de fin**

nom = **Identificateur générique**

- Texte encadré par balise de début et balise de fin = **Élément**
- Identificateur générique indique la **nature** de l'élément
- Imbrication rend compte de la **structure** logique

Documents structurés

Formats de documents structurés

- ➔ Format est dit **structuré**, ou format de documents structurés, s'il permet le **balisage descriptif logique** des documents
- ➔ Ajout de **balises descriptives** au texte pour en indiquer la structure logique
 - › Identifient le début et la fin d'un élément constitutif
 - › Peuvent être imbriquées
 - › Exemples de balisage descriptif logique :
 - `<nom>John Smith</nom>`
 - `<titre><titre principal>Il était une fois un document structuré</titre principal> : <sous-titre>Historique des formats structurés</sous-titre></titre>`
- ➔ Balises ajoutées à un **fichier texte** (exemple d'un fichier texte avec règles supplémentaires)
- ➔ Extensions des fichiers par ex. : .sgml, .html, .xml

Documents structurés : XML

Exemple de format structuré : XML

XML (eXtensible Markup Language) est un format...

... normalisé

- › **Recommandations** du W3C de 1998
 - XML 1.0 : 1^{ère} édition en février 1998; 5^{ième} édition en novembre 2008
 - XML 1.1 : 1^{ère} édition en février 2004; 2^{ième} édition en septembre 2006 (Ajouts mineurs par rapport à XML 1.0)

... de documents numériques

- › Tout comme Word, HTML, PDF, ...
- › Documents = fichiers (.xml) qui peuvent être
 - imprimés, échangés, diffusés sur le Web
 - convertis vers autres formats (ex. PDF, HTML)
 - liés à d'autres documents par des liens hypertextuels

... de documents structurés

- › **Balisage descriptif** (=nature de l'information) **logique** (=structure de l'information) par dessus un format texte (Unicode-UTF-8) [*format texte « enrichi »*]
- › **Balises** identifient le début et la fin des éléments constitutifs du document (par ex. **<nom>**Michel Gagnon**</nom>**)
- › Ne traduit pas la mise en forme ou d'autres traitements (aucune information de mise en page dans un document XML contrairement à d'autres types de balisage tel le balisage procédural)
 - Séparation bénéfique du contenu et des traitements
 - Traitement via d'autres applications (par ex. feuilles de styles pour la restitution des documents)

Documents structurés : XML

XML = Balisage généralisable

➔ XML est en fait un **métalangage**

- › Permet de définir des ensembles de balises pouvant représenter **tout type de documents structurés** (par exemple, les mémos dans une compagnie, les procès-verbaux, ...)
- › Ensemble des balises (*structure logique*) défini dans une *Document Type Definition* (DTD) et/ou un « schéma XML »

➔ **Restitution** des documents (c'est-à-dire mise en page) définie dans une(des) feuille(s) de styles

- › Permet de réutiliser l'information

➔ **Exemples d'application**

- › Documents textuels (rapports, mémoires)
- › Fiches, dossiers (médical, d'employé)
- › Métadonnées (MARC, RDF, GILS)
- › Formulaires, documents administratifs (procès-verbaux)
- › Communication savante (revues)
- › Référence (dictionnaires, encyclopédies)
- › Multimédia, hypermédia (présentations)

Documents structurés : XML

Avantages et désavantages de XML

Avantages

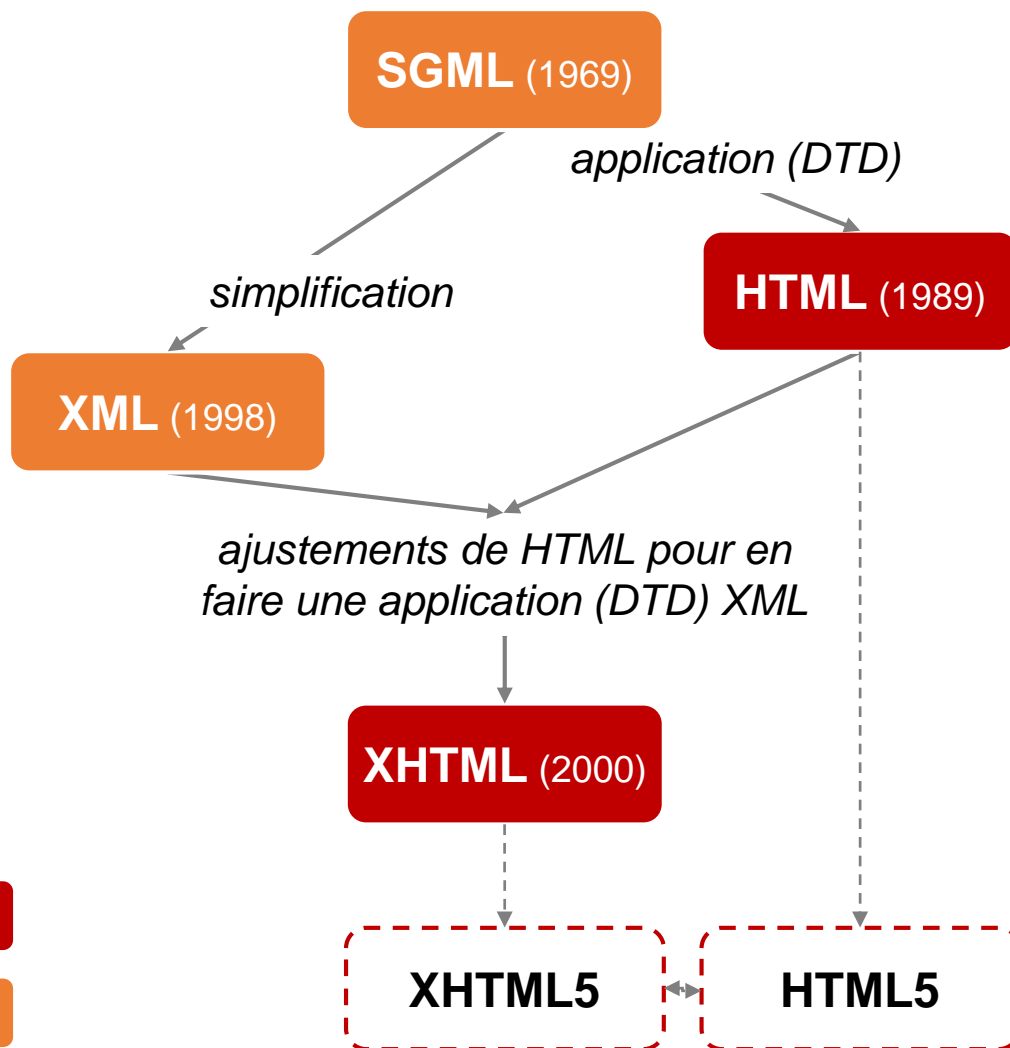
- ➔ Balisage **logique**
- ➔ **Réutilisation** de l'information, versatilité d'accès
- ➔ Spécification des applications **indépendante** des contenus (temps et « espace »)
- ➔ **Factorisation** du travail entre auteurs, typographes, informaticiens, etc.
- ➔ **Indexation** automatique, **recherches** d'information améliorées
- ➔ Format **normalisé**
 - › Diffusion « universelle »
 - › Échange, interopérabilité (important pour le commerce électronique)
 - › Potentiel de pérennité (convient à l'archivage à long terme)
 - › Logique + normalisé : indépendance par rapport aux logiciels et au matériel de restitution
- ➔ Internationalité et internationalisation (i18n) : **Unicode**

Désavantages

- ❖ **Surtout changement « culturel » pour les auteurs qui passent d'un logiciel de traitement de texte**
 - Comparable à obliger l'usage des styles Word

Documents structurés

Évolution des langages de balisage normalisés



Légende

Type de document

Métalangage

Documents structurés : SGML

SGML = ISO/IEC 8879:1986

- ⇒ **SGML** = *Standard Generalized Markup Language*
 - › Langage **normalisé** de balisage **généralisé**
 - › Développé à partir de GML (1969), un produit d'IBM
 - GML = auteurs Goldfarb, Mosher, Lorie!
 - › **Métalangage** permettant de représenter tout type de documents structurés

- ⇒ Par rapport à XML
 - › SGML plus complexe
 - › SGML légèrement plus puissant

Documents structurés : HTML

HTML

- ➔ **HTML** = *HyperText Markup Language*
- ➔ **Application SGML** (c'est-à-dire une modélisation SGML de la structure d'une page Web; HTML n'est ainsi pas un métalangage)
- ➔ Plusieurs versions de HTML
 - › HTML 1 : 1989 (texte, hyperliens)
 - Inventé par Tim Berners-Lee, reconnu comme le créateur du Web
 - › HTML 2 : 1994 (images, formulaires)
 - › HTML 3 : 1996 (son, applets)
 - › HTML 4 : 1998 (vidéo, CSS)
 - › Depuis 2000 : version normalisée de HTML par ISO/IEC (norme internationale ISO/IEC 15445:2000)
 - › HTML 5 : recommandation finale en 2014

Documents structurés : XHTML

XHTML

- ➔ **XHTML 1.0** (2^{ième} édition) : Extensible Hypertext Markup Language
 - Recommandation W3C janvier 2000 (révisée en août 2002)
 - Essentiellement, c'est une réécriture de HTML 4.01 comme une application XML, plutôt que SGML (c'est donc une modélisation en XML de la structure d'une page Web; XHTML n'est ainsi pas un métalangage)

- ➔ **XHTML 1.1** : *Module-based XHTML*
 - Recommandation W3C en mai 2001
 - Deuxième édition en novembre 2010
 - Introduit la possibilité d'extensions modulaires (par exemple, XForms pour les formulaires)

- ➔ **XHTML 5** : *Sérialisation XML de HTML5*

Documents structurés : (X)HTML

HTML & XHTML

Points forts

- ➔ **Simplicité, portabilité**
 - Basés sur des formats texte normalisés
- ➔ **Outils** gratuits et faciles à utiliser
- ➔ **Navigateurs** nombreux et omniprésents sur toutes les plateformes courantes
- ➔ Parfait pour des **documents simples** à diffuser sur le **Web**

Points faibles

- ➔ Inapproprié pour contrôler parfaitement le **rendu visuel des documents** (PDF est plus approprié dans ce cas)
- ➔ Difficile de produire des **documents imprimés** sur papier de qualité professionnelle
 - Ex.: pas de notion de référence à un numéro de page

Documents structurés

Enjeux pour les milieux documentaires

- ➔ L'information numérique est **de plus en plus** produite sous forme structurée (par ex. XML)
- ➔ Pour la gérer de manière « intelligente », il faut la traiter dans son **format natif** (et non en format de présentation)
- ➔ XML est aussi utile pour la **production d'information numérique** (catalogage, thésaurus, ontologies, etc.)

Documents structurés

Conséquences pour le professionnel de l'information

➔ **Compétences à acquérir et nouveaux rôles**

- › **Modélisation** de l'information
- › **Maîtrise** des outils de base de traitement
- › Compréhension des possibilités de traitement automatique (**réutilisation**)
- › Compréhension des **chaînes de traitement**