

Les métadonnées, c'est simple comme un pot de confiture !

Aimez-vous les confitures? Et comment les dégustez-vous? À la cuillère? Sur du pain? Le pot est-il vite terminé? Ou préférez-vous le garder pendant longtemps?

Si vous avez des réponses à ces questions alors les métadonnées seront un jeu d'enfant pour vous! Car les métadonnées sont, comme l'indique son étymologie, des données sur des données ou bien plus précisément, des informations de contexte, de contenu et de structure qui qualifient un contenu d'information, un peu comme l'étiquette du pot de confiture vous permet d'identifier ce qu'il y a à l'intérieur du pot.



Maday Confitures

L'étiquette du pot de confiture

L'étiquette du pot de confiture, selon les normes alimentaires en vigueur en Europe, porte des informations vitales pour votre santé (la date de péremption, composition, valeurs caloriques) ou pour votre agrément: vous éviterez de servir du coulis de tomate au petit déjeuner car l'étiquette indique bien «confiture de framboises», et précise que les framboises proviennent d'Espagne et que «grand-maman», votre fabricante préférée, en est l'auteur. Si votre confiture contient un peu d'alcool, on vous avertira qu'elle ne convient pas aux enfants et femmes enceintes.

Il en est de même pour les métadonnées : un contenu d'information n'est compréhensible et pérennisable que lorsqu'on est capable d'identifier rapidement, précisément ce contenu, et qu'il est décrit avec un niveau de finesse suffisant pour les besoins métiers présents et futurs.

«L'étiquette du pot de confiture porte des informations vitales.»

Pareillement, pour accéder au contenu du pot de confiture, il vous faut du muscle, pour ouvrir le pot, et à votre convenance, une cuillère, un couteau ou un doigt gourmand pour recueillir la bonne quantité une fois le pot ouvert. Et si vous souhaitez conserver longtemps cette merveilleuse confiture de framboises, alors il vous faudra, selon les écoles, la réserver au frais ou dans un placard, à l'abri de la lumière et d'autres gourmands.

C'est bien la même chose pour les contenus d'information: il vous faudra, pour y accéder, des informations concernant la syntaxe et la structure des informations, pour pouvoir bien la comprendre et l'interpréter, mais également des informations indispensables pour assurer l'intégrité et la pérennité de ce contenu: le format de fichier, le type MIME, le logiciel qui l'a produit, ...

Les normes pour les métadonnées

Alors qu'il y a des normes alimentaires pour vous apporter de l'information en tant que consommateur, il y a également des normes définissant un cadre pour les métadonnées.

Ainsi, pour l'archivage électronique, le cadre est fixé par le modèle conceptuel OAIS, qui identifie plusieurs types de métadonnées nécessaires à la bonne compréhension du contenu:

- les informations de représentation qui permettent de définir l'interprétation des séquences d'octets et les traduire dans une forme lisible,
- les informations de description, qui permettent de comprendre le contexte de création et de vie d'un contenu (nom, provenance, droits d'accès),

- les informations d'emballage qui explicitent les relations entre les différents composants d'un paquet d'information à archiver.

«Les métadonnées sont indispensables pour assurer l'intégrité, l'authenticité, la fiabilité et la lisibilité des documents et leur contenu.»

D'autres normes, plus directement liées au métier des archivistes, existent pour décrire les ressources et contenus: pour les métadonnées descriptives, on pourra s'appuyer sur les normes ISAD(G) et ISAAR(CPF) ou plus récemment sur le modèle RiC (*Records in Context*), ou encore le Standard d'échanges de données pour l'archivage, qui en France définit des dictionnaires de données et des schémas xml permettant de les encoder, tout comme PREMIS et la DTD EAD.

Trop de métadonnées tuent la métadonnée

Mais attention: on sait que la confiture ça dégouline et ça colle, quand on en met trop! Aussi, il est important de bien évaluer les besoins métiers dans la définition d'un jeu de métadonnées: les utilisateurs ne sont pas friands de remplir des champs à la main et s'en lassent assez vite, alors qu'en phase projet toute l'attention a été portée sur une définition fine des besoins en termes d'identification des ressources. De nombreux systèmes de gestion électronique des documents (GED) avec des champs de métadonnées précis et nombreux, mais vides, sont là pour nous rappeler tous les jours qu'il s'agit d'un écueil trop commun pour ne pas être évitable.

Pour pallier ces désagréments, automatiser au maximum la complétion des champs est une solution, tout comme limiter et factoriser le nombre de champs pour chaque type de contenu.

Ainsi, les Archives nationales d'Australie <http://www.naa.gov.au/information-management/digital-transition-and-digital-continuity/information-is-interoperable/metadata/index.aspx> ont défini un set minimal de 9 métadonnées dont 3 vitales: l'identifiant, le créateur et la date de création. Quant aux Archives nationales des États-Unis d'Amérique, elles se réfèrent aux métadonnées d'un jeu normalisé, le Dublin Core, pour qualifier les documents fédéraux <https://www.archives.gov/records-mgmt/bulletins/2015/2015-04.html>.

Notons enfin l'initiative de l'Association des archivistes français, qui, en collaboration avec l'AFNOR (l'organisme de normalisation français), a proposé un set minimal de métadonnées à associer aux contenus d'information, suite à une étude et un croisement de plusieurs sets de métadonnées en usage dans des entités privées, ou publiques, à l'échelon local, national et international. Ce travail, publié sous forme de livre blanc, sera bientôt disponible à tous sur le site de l'ISO. En attendant, nous pouvons toujours nous référer au 4e livre blanc édité en 2013 par l'AFNOR sur les métadonnées pour les documents d'activité http://www.ademec.com/sites/default/files/livre_blanc_4.pdf.

Les métadonnées n'ont donc plus de secret pour vous: bonne dégustation!



Charlotte Maday

Charlotte Maday est archiviste. Elle a exercé successivement comme expert archives pour la plateforme d'archivage électronique du CINES puis en étant responsable des archives de l'université Paris Diderot-Paris 7 jusqu'en 2016, avant de rejoindre le secteur privé (Saint-Gobain) et l'édition de logiciels de gestion des archives (Spark archives, KleeGroup).

Abstract

Français

Les métadonnées qualifiant les documents d'activité et les archives sont désormais chose commune: mais comment bien les identifier et les appréhender? Quelles informations vitales doivent accompagner un document tout le long de son cycle de vie? Quels sont les textes qui peuvent nous aider à mieux les identifier? Quels écueils éviter? L'étiquette d'un pot de confiture nous aide à répondre à ces questions.

Deutsch

Metadaten, die aktive und archivierte Dokumente beschreiben, haben gemeinsame Eigenschaften. Aber wie ist es möglich, diese Metadaten adäquat zu identifizieren und zu erfassen? Welche Informationen sind unentbehrlich und müssen ein Dokument während seines gesamten Lebenszyklus begleiten? Wie lauten die Beschreibungen und Standards, die uns am besten helfen, diese Metadaten zu identifizieren? Welche Klippen gilt es dabei zu umschiffen?