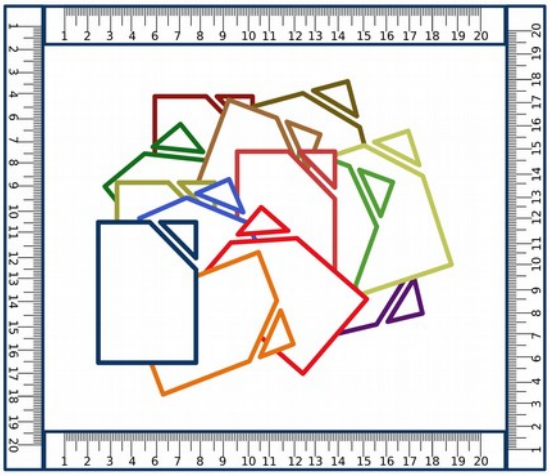


Formats ouverts et métrologie

Parce que les formats ouverts sont essentiels

1er trimestre 2019

Isabelle Dutailly

FORMATION, RÉDACTION, CONSEIL
isabelle@dutailly.net — +33 (0)6 10 28 14 62
www.dutailly.net — www.numericoach.net

Parce que l'écriture numérique requiert des connaissances

Coaching numérique, rédaction, conception et réalisation de tous types de documents, formations bureautique et internet.

Il est fondamental que les données numériques soient aussi aisément accessibles que nos livres papier. Qu'une entreprise ou une autorité publique puisse contrôler la lecture et la conservation des données ou les moyens de communication semble inconcevable dans une société démocratique. L'avenir nous en préservera : l'adoption massive des standards ouverts croît inexorablement.

Ladislas Goffinet, *Les standards ouverts et la normalisation comme conditions de l'interopérabilité logicielle,* Cahiers Droit, Sciences & Technologies, 3-2010

https://journals.openedition.org/cdst/232

Présentation

Pourquoi mêler l'histoire des systèmes de mesures à celle des formats informatiques ?

Tout simplement parce que l'une comme l'autre ont des points et des enjeux communs, que le système métrique est probablement le premier standard ouvert connu et qu'il peut être plus facile de comprendre l'impérieuse nécessité d'utiliser des formats informatiques ouverts par ce biais. Et aussi parce que les uns et les autres font partie de notre vie quotidienne, parfois même sans que l'on s'en rende compte.

L'un amenant l'autre, après une petite révision sur le génie du système métrique et une très brève histoire de l'informatique, on arrivera tout naturellement à comprendre l'importance et la nécessité des formats ouverts et ce que cela signifie dans une société qui s'oriente de plus en plus vers le tout numérique. Et, bien sûr, on saura les « repérer », pour les plus courants et comment les générer, avec quoi les travailler. On verra aussi que, comme monsieur Jourdain, nous utilisons quotidiennement sans le savoir, souvent, des formats ouverts.

La plupart, si ce n'est toutes, des références bibliographiques sont sur internet et ce document contient des liens internes, il sera donc plus facilement lu et consulté sur un support électronique que sur un support papier. Par ailleurs, j'ai essayé de le concevoir de façon à ce qu'il soit le plus accessible possible, que vous lisiez avec vos yeux, vos mains ou vos oreilles.

Quoi qu'il en soit, je vous en souhaite bonne lecture.

Table des matières

Chapitre 1. Un peu de métrologie pour commencer....5
1.1 Des mesures conservées avec un soin religieux....5
1.2 Une diversité enjeu de pouvoir et source de problèmes....5
1.3 Le système métrique : premier format ouvert connu....7
1.4 Le système impérial....8
1.5 L'exemple du tricot et des aiguilles à tricoter....9
Chapitre 2. Un peu d'histoire de l'informatique....11
2.1 Un OS quasiment par constructeur....11
2.2 Chacun son format de fichier....11
2.3 Une émergence de formats standard de fait....12
Chapitre 3. Les formats ouverts en informatique....15
3.1 Définition....15
3.2 Qualités des formats ouverts....15
3.3 Les fichiers aux formats ouverts sont partout....16
Chapitre 4. Quel format pour quel fichier et par quelle application ?....17
4.1 Bureautique et assimilé....17
4.2 Image....17
4.3 Diffusion, échanges et conservation....18
4.4 Multimédia : son et vidéo....19
Chapitre 5. HTML, CSS, et XML, exemples de langages....21
5.1 Le HTML, structure et exemples de balise....21
5.2 Le XML....22
5.3 Les CSS....23
Chapitre 6. Est-ce que ?....25
6.1 Est-ce qu'un fichier protégé est sûr ?....25
6.2 Est-ce que les formats ouverts sont ouverts par toutes les applications équivalentes ?....25
6.3 Est-ce que les formats ouverts sont réservés aux logiciels libres ?....25
6.4 Est-ce qu'un fichier au format ouvert est moins bien qu'un fichier au format privateur ? 25
6.5 Est-ce qu'on est obligé d'utiliser des formats ouverts ?....25
6.6 Est-ce qu'on peut m'obliger à donner des fichiers dans un format privateur ?....26
6.7 Est-ce que l'administration peut m'imposer un format privateur ?....26
Chapitre 7. Crédits, droits, versions....27
7.1 Crédits....27
7.2 Droits....27
7.3 Versions....27
Bibliographie....28

Chapitre 1. Un peu de métrologie pour commencer

La métrologie, c'est la science des poids et mesures et, oui, il y a des points communs entre la métrologie et l'informatique et pas uniquement en termes de poids de fichiers.

1.1 Des mesures conservées avec un soin religieux

Le besoin d'avoir des systèmes de mesures des poids, volumes et longueurs a dû apparaître assez tôt dans l'histoire de l'humanité. Probablement dès lors que, la population augmentant, les échanges aussi, il devenait nécessaire d'avoir des outils permettant de, sinon, garantir, tout au moins faciliter l'équilibre des échanges.

Illustration 1: muid de Ponte Puñide, Musée archéologique d'Espagne, photo Luis García, source Wikipédia.

Partout, pour autant qu'on sache, les poids et mesures étalons étaient conservés, quasiment littéralement religieusement : par exemple dans l'ancienne Rome ils étaient conservés dans le temple de Jupiter et l'empereur Justinien, au 16e siècle, après avoir fait vérifier tous les poids et les mesures en confia la garde des originaux à la principale église de Constantinople. Dans tous les cas, semble-t-il, à l'abri de tous, mais aussi, comme on le verra de personnes chargées de leur vérification.

Ainsi, dans la France de Charlemagne, le système de mesure était uniforme et les étalons conservés au palais royal.

Ensuite est venu le chaos.

1.2 Une diversité enjeu de pouvoir et source de problèmes.

En France, et ailleurs, le système féodal, en accordant aux seigneurs des prérogatives en manière de poids et mesures, et même de monnaie, a fait des systèmes de mesure un enjeu de pouvoir. Et, comme le pouvoir n'aime généralement pas le grand jour, la dissimulation au public des mesures étalon était courante, voire une sorte de norme.

À cette diversité s'ajoutait le fait que les mesures avaient souvent des noms très similaires à identiques. Il fallait donc être vigilant sur le nom exact de la mesure. Il y avait aussi le fait qu'au lieu d'un système de mesures comme celui que nous connaissons qui comporte trois catégories outre les températures et les unités informatiques : les longueurs, les volumes, les poids[1], il y avait aussi une multitude de mesures différentes en fonction du type de denrées. On en retrouve des traces par exemple dans le bois de chauffage qui est vendu au stère, lequel vaut maintenant un mètre cube de buches.

1 À noter d'ailleurs que le système de mesure en informatique est basé sur le système décimal comme le système métrique.

Dans son *Mémoire sur la nécessité et les moyens de rendre uniformes les mesures d'étendue et de pesanteur* présenté à l'Assemblée nationale en 1790[2], Claude - Antoine Prieur signale qu'il a : « présenté le tableau relatif à la ville de Paris, parce qu'il [lui] a paru intéressant de donner, dans un seul coup d'œil, une idée de la diversité des mesures qui y sont en usage. On y verra que, dans cette seule ville, il y a plus de soixante mesures différentes, dont plusieurs cependant portent les mêmes noms, car le nombre de ces derniers n'est que de quarante-cinq. Les mesures nationales n'ont au contraire que vingt noms, et chacun d'eux n'est jamais attribué qu'a des quantités semblables. ». Oui, oui soixante mesures différentes pour la seule ville de Paris et seulement quarante-cinq noms de mesures différents !

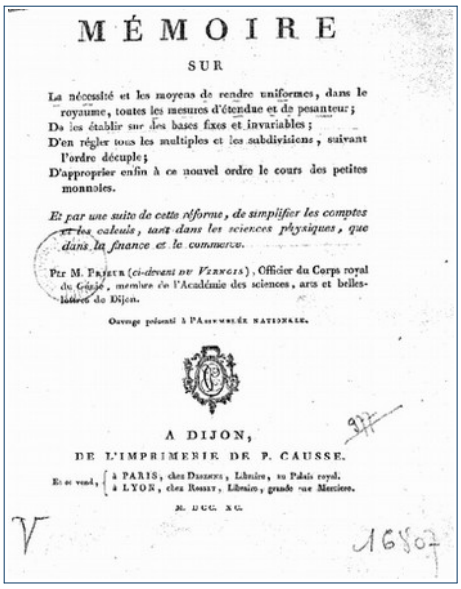

MÉMOIRE
SUR
La nécessité et les moyens de rendre uniformes, dans le royaume, toutes les mesures d'étendue et de pesanteur;
De les établir sur des bases fixes et invariables;
D'en régler tous les multiples et les subdivisions, suivant l'ordre décuple;
D'approprier enfin à ce nouvel ordre le cours des petites monnoies.
Et par une suite de cette réforme, de simplifier les comptes et les calculs, tant dans les sciences physiques, que dans la finance et le commerce.
PAR M. PRIEUR (ci-devant DE VERNOIS), Officier du Corps royal du Génie, membre de l'Académie des sciences, arts et belles-lettres de Dijon.
Ouvrage présenté à l'ASSEMBLÉE NATIONALE.
A DIJON,
DE L'IMPRIMERIE DE P. CAUSSE.
Et se vend, à PARIS, chez DIDIERS, Libraire, au Palais royal. à LYON, chez ROSSET, Libraire, grande rue Mercière.
M. DCC. XC.

Illustration 2 : page de titre du mémoire de Prieur, source BnF Gallica.

On avait par exemple divers muids, de blé, d'avoine, de sel, de pierre, de chaux, de charbon de bois et de charbon de terre, ou encore la perche pour les terres et celle pour les eaux et forêts, l'aune pour les soieries, pour les lainages ou pour les toiles, toutes mesures ayant des contenances et des tailles différentes[3]. On voit sur l'extrait du tableau de comparaison ci-dessous que, par exemple l'aune pour les soieries équivalait à $^{527,5}/_{144}$ de pied de roi tandis que l'aune pour les lainages en valait $^{526,5}/_{144}$ et celle pour les toiles $^{524}/_{144}$. On notera également qu'en introduisant le système décimal, le système métrique a également beaucoup facilité les mesures.

MESURES ANCIENNES			MESURES NATIONALES			RAPPORTS RÉCIPROQUES CORRESPONDANS MULTIPLICATEURS	
						Des nombres de mesures anciennes, et des prix de chaque mesure nationale. A	Des nombres de mesures anciennes, et des prix de chaque mesure nationale. B
Rapports avec le pied de Roi	12 000	la lieue de Paris	Rapports avec le pied	10 000	Le millaire	1,17666	0,84986
	18	La perche pour les terres		10	La perche		
	22	La perche pour les eaux et forêts			Idem		
	6	La toise			Idem	0,58830	1,69981
	$^{527,5}/_{144}$	L'aune pour les soieries		3,5	L'aune		
	$^{526,4}/_{144}$	L'aune pour les lainages			Idem		
	$^{524}/_{144}$	L'aune pour les toiles			Idem		
	1	Le pied de roi		1	Le pied	0,98055	1,01984
	$^{1}/_{12}$	Le pouce		0,1	Le pouce	0,81713	1,22380
	$^{1}/_{144}$	La ligne		0,01	La ligne	0,68093	1,46858

Tableau 1: extrait du tableau de correspondance des mesures du Mémoire sur la nécessité et les moyens de rendre uniformes les mesures d'étendue et de pesanteur de Claude-Antoine Prieur.

2 Claude-Antoine PRIEUR, Mémoire sur la nécessité et les moyens de rendre uniformes les mesures d'étendue et de pesanteur présenté à l'Assemblée nationale en 1790, page 32 du pdf (voir bibliographie).

3 Cf le Tableau de comparaison des mesures anciennes et des nouvelles de Claude-Antoine PRIEUR, Mémoire sur la nécessité et les moyens de rendre uniformes les mesures d'étendue et de pesanteur.

Il fallait donc être sûr d'avoir les bons étalons des poids et mesures qui n'étaient pas toujours fabriqués avec des matériaux fiables eux-mêmes, qui pouvaient avoir des formes qui rendaient les mesures moins fiables et quand ils étaient détruits, perdus, on avait perdu le référentiel. Sans parler de l'énorme frein que posait cette diversité à la communication et aux échanges commerciaux.

Dès 1558, Henri III avait essayé de réduire le nombre des systèmes de mesures employés dans le royaume de France, en vain. Dans son mémoire, Prieur indique que « les esprits n'étoient point encore assez préparés pour recevoir ce changement, même revêtu du sceau de la loi ». On peut aussi subodorer que les seigneurs, qui tiraient profit d'une façon ou d'une autre de cet état de fait, n'étaient pas très empressés à ce que cela change.

1.3 Le système métrique : premier format ouvert connu

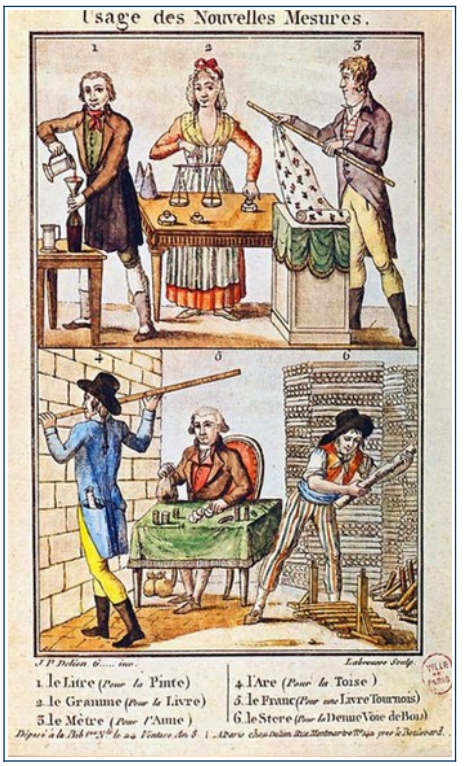


Illustration 3: usage des nouvelles mesures, France circa 1800, source BnF Gallica.

Le système métrique, devenu le système international est une révolution de la révolution française, a uniformisé le système des poids et mesures en imposant un système ouvert, c'est-à-dire dont les bases de calcul sont connues et accessibles de tous. Il est le résultat d'un long et patient travail de recherche. Des savants ont compilé des collections de mesures diverses, les ont analysées pour établir des comparaisons, ont étudié leurs formes et les matériaux et se sont penchés sur la question de leur fiabilité, par exemple, en France, par exemple Alexis-Jean-Pierre PAUCTON dans sa *Métrologie ou Traité des mesures, poids et monnoies des anciens peuples, des modernes* de 1780 et Jean-Baptiste Louis DE ROMÉ DE L'ISLE avec sa Métrologie, ou *Tables pour servir à l'intelligence des poids et mesures des anciens, et principalement à déterminer la valeur des monnoies grecques et romaines d'après leur rapport avec les poids, les mesures et le numéraire actuel de la France* en 1786 que l'on peut télécharger sur le site de la BnF Gallica (cf bibliographie).

En 1790, Prieur, établit les règles que devraient suivre le nouveau système de mesure, sous quelles conditions :

« Je voudrois donc que la longueur du pendule fût tracée sur une regle de platine pur, bien dressée et incrustée dans une autre de bois, pour être déposée à l'hôtel-de-ville de Paris où on l'y conserveroit précieusement dans un étui qu'on y joignît une description de l'opération par laquelle cette longueur auroit été trouvée que dans cette description enfin, on fît mention de la position de l'observatoire en latitude et en longitude, du rayon de la terre en cet endroit, ou au moins de son élévation au-dessus du niveau de la mer, du temps de chaque vibration du pendule, et de la température a laquelle sa longueur a été et doit être toujours mesurée.[4] »

4 Claude-Antoine PRIEUR dans son **Mémoire sur la nécessité et les moyens de rendre uniformes les mesures d'étendue et de pesanteur** présenté à l'Assemblée nationale en 1790, page 9 e t 10 du pdf (voir bibliographie).

Il répond également aux principales objections contre l'uniformité des mesures :

- la diversité est favorable aux négociants ;
- comment faire changer les habitudes de tout un peuple ;
- combien cela coûterait-il ?

Il termine son mémoire par une proposition de loi qui reprend les définitions du nouveau système de mesure, qui devient ainsi public et accessible de tous[5].

Le système décimal et les définitions du mètre seront adoptés en 1793 en France, le système métrique sera rendu obligatoire en 1795, des mètres-étalons de marbre seront placés dans des lieux très passants de la capitale en 1797, subira quelques aléas dus à l'histoire politique de la France dans les années qui ont suivi, puis sera définitivement adopté en France par une loi de 1837 rendant obligatoire l'usage des unités du système métrique à partir du 1er janvier 1840, dans le commerce et dans la vie civile et juridique. Délai nécessaire pour que l'information passe et que les outils de mesures soient aux normes.

À l'origine le mètre était basé sur la circonférence de la terre et valait $^{1}/_{20\,000\,000}$ d'un méridien géographique. Les mesures de masse (poids) et de volume étant basées sur cette première unité. Actuellement, le mètre est la longueur du trajet parcouru dans le vide par la lumière pendant une durée de $^{1}/_{299\,792\,458}$ de seconde. Ce qui revient au même pour les mesures courantes mais ajoute plus de précision. Le kilogramme n'est plus défini comme la masse d'un décimètre cube d'eau depuis 1889. En mai 2019, il devient une valeur numérique fixée à la constante de Planck. Ce qui ne change rien non plus par rapport aux premières définitions pour les mesures courantes mais permet une très grande précision.

1.4 Le système impérial

Il existe aujourd'hui un autre système de mesure répandu, et hélas en informatique[6] car c'est celui utilisé aux États-unis (mais pas dans les milieux scientifiques américains) au Liberia et en Birmanie : le système impérial qui différait entre les deux états jusqu'à ce qu'intervienne un traité pour le préciser en 1959. En effet, jusqu'à cette date le Royaume-Uni, le Commonwealth et les États-unis avaient, pour des mesures aux noms identiques, des définitions différentes. Le résultat est un compromis entre ces mesures. Lesquelles sont définies par rapport au système métrique. Ainsi, un pouce vaut 2,54 cm et un pied 30,48 cm.

Illustration 4: poids en laiton, Angleterre 1826, source Wikipedia,photographe inconnu.

Il importe donc, quand on a des outils de mesures en unités anglo-saxonnes antérieurs à 1959 (voire des années 1960) de vérifier leur origine, car les mesures donneront des résultats différents et pas forcément attendus. Il en va de même pour des modèles et patrons, quel qu'en soit le domaine.

5 Ça rappelle furieusement les préconisations des standards informatiques ouverts, n'est-ce pas ?

6 Notamment les tailles des écrans que les marchands ne prennent pas la peine de transcrire en système métrique.

Le système métrique lui reste valable partout depuis sa création, ce qui fait, aujourd'hui, confortablement plus de deux siècles. Et même si ses définitions ont évolué et qu'il est devenu le Système international, les outils de mesures fabriqués en 1795 restent toujours valables.

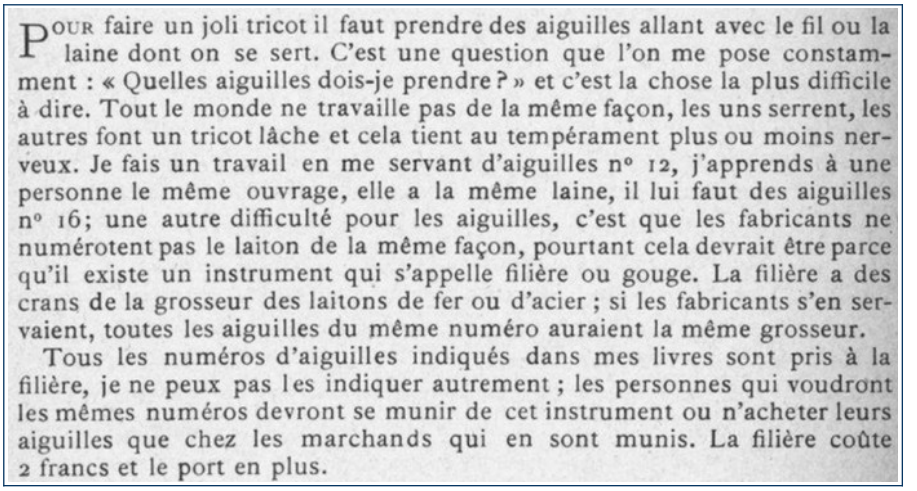
1.5 L'exemple du tricot et des aiguilles à tricoter

C'est un exemple assez typique et qui est susceptible d'empoisonner encore les personnes qui tricotent dans le monde entier.

Globalement, au 19e et au tout début du 20e siècle, quasiment chaque fabricant avait ses aiguilles à son référentiel. Lesdits référentiels étant à base de numéro, le plus souvent (pour les crochets, dans certains cas et encore maintenant aux États-unis notamment, cela pouvait être des lettres).

Cela posait, et continue, à poser divers problèmes. L'un des problèmes principaux concerne l'adéquation entre les modèles de tricot, les aiguilles et les fils à tricoter. Si on ne savait pas de quelle marque étaient les aiguilles, reproduire le modèle était hasardeux.

En tricot, on a une notion d'échantillon : on tricote quelques mailles et rangs, généralement un peu plus de 10 × 10 cm pour faire les calculs sur 10 x 10 cm, afin de déterminer, d'une part si la façon de tricoter est conforme à celle requise par le modèle, d'autre part, le nombre de mailles et de rangs au centimètre. C'est impeccable quand le modèle indique un échantillon parce qu'on peut toujours se débrouiller soit pour changer de fil et d'aiguilles pour s'adapter aux prescriptions de l'échantillon, soit pour calculer les nombre de mailles à monter et de rangs à tricoter en fonction du résultat final. Mais les modèles anciens indiquent rarement un échantillon et encore plus rarement les dimensions ou la taille précise (pointure par exemple) du tricot fini. Prenons, par exemple le *Manuel de tricot : cours élémentaire, contenant 64 descriptions de modèles et illustré de 56 reproductions d'ouvrages*[7] de Mlle M. Baillaud.



POUR faire un joli tricot il faut prendre des aiguilles allant avec le fil ou la laine dont on se sert. C'est une question que l'on me pose constamment : « Quelles aiguilles dois-je prendre ? » et c'est la chose la plus difficile à dire. Tout le monde ne travaille pas de la même façon, les uns serrent, les autres font un tricot lâche et cela tient au tempérament plus ou moins nerveux. Je fais un travail en me servant d'aiguilles n° 12, j'apprends à une personne le même ouvrage, elle a la même laine, il lui faut des aiguilles n° 16; une autre difficulté pour les aiguilles, c'est que les fabricants ne numérotent pas le laiton de la même façon, pourtant cela devrait être parce qu'il existe un instrument qui s'appelle filière ou gouge. La filière a des crans de la grosseur des laitons de fer ou d'acier ; si les fabricants s'en servaient, toutes les aiguilles du même numéro auraient la même grosseur.

Tous les numéros d'aiguilles indiqués dans mes livres sont pris à la filière, je ne peux pas les indiquer autrement ; les personnes qui voudront les mêmes numéros devront se munir de cet instrument ou n'acheter leurs aiguilles que chez les marchands qui en sont munis. La filière coûte 2 francs et le port en plus.

Illustration 5: extrait de l'introduction du manuel de tricot de Mlle Baillaud, source Bnf Gallica.

Dans ce manuel de 1907, mademoiselle Baillaud indique dans l'introduction que, pour faire un joli tricot (et qui a envie d'en faire un moche ?) il faut prendre un fil et des aiguilles assorties. Elle ajoute, qu'il est très difficile de répondre à la question qu'on lui pose souvent « quelles aiguilles dois-je prendre pour ce tricot ? » parce que tout le monde ne travaille pas de la même façon, ce qui est tout à fait vrai. Et elle continue en précisant qu'elle utilise tel numéro d'aiguille et que

7 Téléchargeable sur le site de la BnF Gallica : http://gallica.bnf.fr/ark:/12 148/bpt6k9 334 646.r=Manuel%20de%20tricot?rk=64 378;0

pour déterminer le numéro desdites aiguilles, il faut une filière, qu'elle vend d'ailleurs gentiment au prix de 2 francs plus frais de port. Comment fait-on aujourd'hui ? Le numéro 12 n'est très certainement pas les 12 mm de diamètre des aiguilles N°12 actuelles et, on ne sait pas quelle filière elle a utilisée exactement. Premier problème.

Deuxième problème, c'est typique et assez courant pour l'époque, pas de mesures identifiées.

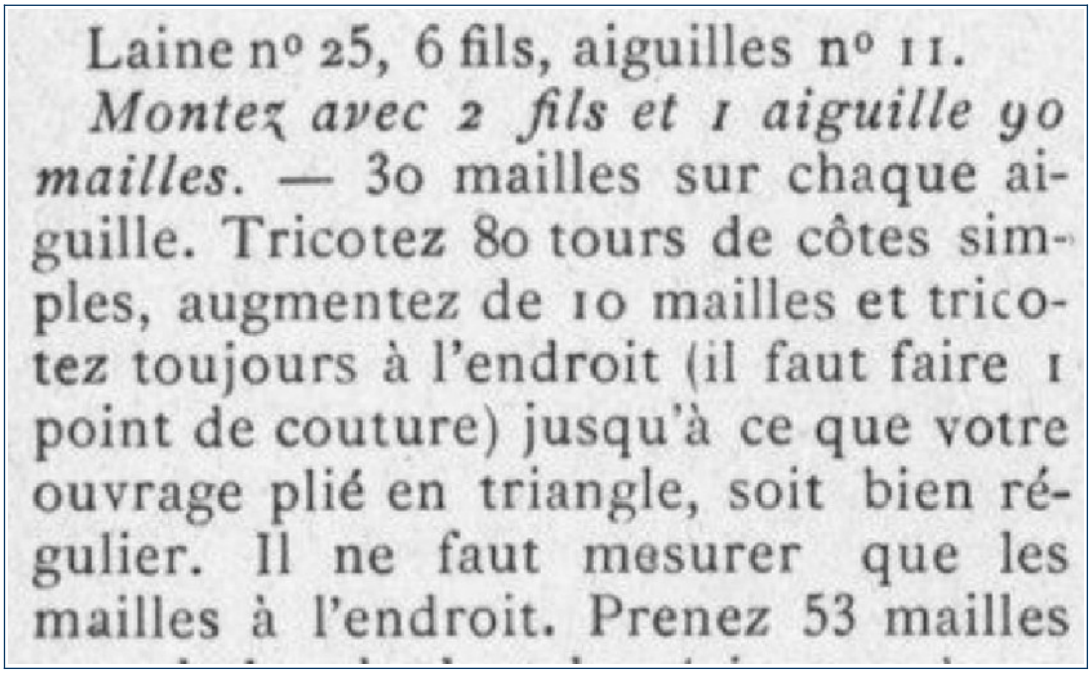

Laine nº 25, 6 fils, aiguilles nº 11.
Montez avec 2 fils et 1 aiguille 90 mailles. — 30 mailles sur chaque aiguille. Tricotez 80 tours de côtes simples, augmentez de 10 mailles et tricotez toujours à l'endroit (il faut faire 1 point de couture) jusqu'à ce que votre ouvrage plié en triangle, soit bien régulier. Il ne faut mesurer que les mailles à l'endroit. Prenez 53 mailles

Illustration 6 : modèle de chaussette pour homme tiré du Manuel de Mlle Baillaud, source : BnF Gallica.

Dans ce modèle de chaussette, Mlle Baillaud, qui l'intitule « Chaussette pour homme », démarre immédiatement, après avoir indiqué le numéro de laine et celui des aiguilles par le nombre mailles à monter et de tours à faire. Aucun échantillon précisé, aucune dimension, sachant qu'entre une pointure 41 et une pointure 47 il y a 3 cm d'écart, ce qui n'est pas rien, qu'a priori un plus grand homme a des pieds plus grands qu'un homme plus petit, mais pas toujours, et, qu'enfin, avec la laine à chaussettes actuelle et des aiguilles 2,5 mm on monte entre 64 et 72 mailles, il est impossible de refaire ce modèle ni de déterminer comment le refaire et avec quelles aiguilles.

C'est un peu comme si on avait un fichier créé avec un logiciel propriétaire qui n'existe plus en fait[8].

8 Accessoirement, il semble que mademoiselle Baillaud n'ait pas été très progressiste : on trouve des modèles de tricot datant des années 1910 qui donnent des numéros d'aiguilles en millimètres.

Chapitre 2. Un peu d'histoire de l'informatique

Des fichiers et des disquettes.

2.1 Un OS quasiment par constructeur

Au début, c'est-à-dire vers la fin des années 1980 quand les ordinateurs ont commencé à envahir les bureaux et entamé leur entrée dans les foyers domestiques, il y avait une foultitude de systèmes informatiques (OS), qui ne se parlaient pas entre eux. Un document généré par un ordinateur d'un système donné ne pouvait être lu par un autre ordinateur d'un autre système. Pire, les disquettes, qui étaient le seul support mobile après que les cassettes magnétiques aient été renvoyées aux oubliettes de l'informatique, n'étaient pas non plus compatibles entre elles, soit parce que les formats matériels étaient différents, soit parce qu'une disquette formatée pour un OS ne pouvait être lue par un ordinateur différent. La palme revenant à la firme Apple qui avait opté pour des disquettes 5 pouces 1/4 pour la série Apple II (et un OS particulier) et des disquettes 3 pouces et un autre OS pour la série Macintosh.

Illustration 7 : disquettes 8, 5 ¼ et 3,5 pouces, Wikipédia, photographe George Chernilevsky.

Au tout début : quasiment chaque fabricant d'ordinateur avait sa version d'OS, voire chacun son format matériel de support, disque, disquette, ou même de connecteur.

2.2 Chacun son format de fichier

Évidemment, dans ces conditions, les logiciels avaient, eux aussi, des formats propriétaires et incompatibles entre eux. Résultat, aujourd'hui une bonne partie des données est inutilisable par manque de matériel ou par défaut de système ou de logiciel.

Tant que les ordinateurs finalement, n'échangeaient pas entre eux, ou alors seulement en interne, cela ne portait pas à conséquence, sauf que… après une assez brève période d'effervescence, le secteur s'est resserré sur quatre OS principaux au début des années 1990 et les systèmes ont commencé à s'ouvrir les uns aux autres.

Mais entre-temps, combien de documents restés dans les machines qu'il devenait impossible d'ouvrir ? Quelle quantité de travail perdu ? Quelles archives précieuses inaccessibles ? S'il est délicat, voire impossible, de donner des réponses précises à ces questions, on peut, cependant, signaler le projet Software Heritage[9].

Lancé en 2015, soutenu par l'Institut national de recherche en informatique et en automatique (Inria) et, depuis 2017 en partenariat avec l'Unesco, le projet Software Heritage s'est donné pour mission d'archiver tous les logiciels de la planète. Ce qui pose d'ailleurs des problèmes divers : juridiques et techniques notamment. En juin 2018, il recensait… plus de 4,5 milliards de fichiers (Source files sur la capture d'écran) pour près de 84 millions de logiciels (Projects) générés par

9 Site partiellement en français.

19,7 millions de développeurs (Authors). Et ce n'est qu'un début et pas uniquement parce que Software Heritage s'est essentiellement penché sur les logiciels libres ou open source.

Size

As of today the archive already contains and keeps safe for you the following amount of objects:

Source files	Directories	Commits	Authors	Projects	Releases
4 579 199 413	4 014 567 447	1 037 021 406	19 696 423	83 835 274	7 706 070

Illustration 8 : capture d'écran du site de Software Heritage, juin 2018.

2.3 Une émergence de formats standard de fait

Avec le début de l'informatique en réseau et des échanges de fichiers entre personnes diverses, il s'est avéré qu'il fallait trouver un format lisible par, disons, la majorité des logiciels de traitement de texte ou des tableurs. Il fallait également, les appareils photo numériques se multipliant, avoir des formats à portée plus universelle[10].

On n'évoquera que quelques formats qui se sont imposés comme standard de fait, les plus connus.

Pour la bureautique, au fil des ans, le bulldozer Microsoft s'étant mis en marche avec efficacité et en laissant de nombreux cadavres dans son sillage, on s'est retrouvés avec des standards de fait. Dans d'autres domaines, image, son et base de données, les standards de fait ont des origines plus variées.

- CSV : c'est un format un peu particulier qui s'est imposé, malgré ses défauts, dus principalement à son manque de spécification formelle, d'échanges de données tabulaires. Lesquelles peuvent être lues par des tableurs, des bases de données mais aussi des traitements de texte par exemple.

- DBF : c'était un format d'échange de bases de données né aux débuts des années 1980 et qui a été très utilisé notamment dans les années 1990. Il n'est plus qu'anecdotiquement utilisé.

- DOC, puis DOCX depuis 2007, Microsoft ayant réussi à imposer son OS et sa suite bureautique à très grande échelle, le format ouvert odt étant, quant à lui encore balbutiant, c'est ce format propriétaire qui était conseillé pour des textes que l'on devait retravailler, ou simplement voir, les pdf étant encore relativement rares et il fallait le logiciel spécifique de la firme Adobe pour en générer. Cette hégémonie qui a commencé son essor à la fin du vingtième siècle devait céder le pas au début de cette décade 2010. Il avait l'avantage d'être le format « universel de fait » dont on savait que le résultat serait identique d'un OS (Microsoft ou Apple) à l'autre pour peu que les machines fussent équipées de la même police de caractère et l'inconvénient[11], outre d'être un format propriétaire, d'être généré par un logiciel qui n'existe pas pour Linux. Le format doc est remplacé en 2007 par le format docx et n'est plus maintenu depuis 2014, c'est un format mort.

10 Certaines marques d'appareils photo numériques de première génération avaient leur propre format d'image qui ne pouvait être lu et travaillé qu'avec le logiciel spécifique livré avec l'appareil et pas toujours enregistrable dans un autre format.

11 Sans parler des caractéristiques techniques et des défauts propres à tous les formats des logiciels des suites bureautiques Microsoft et des problèmes de rétrocompatibilité.

➲ JPG ou JPEG : actuellement le format d'image le plus répandu sur internet et probablement dans les appareils photo du fait de la légèreté des fichiers à ce format. Les spécifications du format sont publiques, mais pas son codage arithmétique, ça n'est pas un standard ouvert.

➲ MP3 ou MPEG : ce format de compression audio, le plus répandu, le seul souvent que l'on connaît est un format dont les sources sont très anciennes[12] et qui a connu un plein essor depuis le début de ce siècle en lien avec celui d'internet et, évidemment, des baladeurs numériques et de smartphones. Il est entré dans le domaine public en avril 2017.

➲ PDF : format d'échange de fichiers, les pdf est une sorte de « photocopie numérique » qui a débuté en 1998 et s'est imposé complètement au début de ce siècle. Le format, au départ privateur de la société Adobe est devenu ouvert en 2008.

➲ PPT puis PPTX depuis 2007 : formats du logiciel de présentation de la suite Ms Office, le .ppt n'est plus maintenu depuis 2014. comme le format doc il s'est imposé « par défaut », comme pour le format doc, c'est le format natif de logiciels qui n'existent pas sous Linux.

➲ RTF (Rich Text Format), développé par Microsoft, c'était le format de Wordpad, espèce d'intermédiaire entre un vrai traitement de texte et un vrai éditeur de texte. Sans reprendre tous les éléments d'un document en traitement de texte complexe, notamment les styles, il garde les enrichissements typographiques (gras, italiques) et le rendu du texte est lisible (ce qui n'est pas vraiment le cas du txt). En revanche, les fichiers générés sont assez lourds, et c'est, de toute façon, un format mort. Aux tous débuts d'internet c'était le meilleur format d'échanges de fichiers envisageable.

➲ XLS puis XLSX depuis 2007 : formats du tableur Excel de la suite Ms Office, le .xls n'est plus maintenu depuis 2014. comme le format doc il s'est imposé « par défaut », comme pour le format doc, c'est le format natif de logiciels qui n'existent pas sous Linux.

12 Selon l'article de Wikipédia, https://fr.wikipedia.org/wiki/MPEG-1/2_Audio_Layer_III : « *les techniques de compression de données audio employées par le format MP3 trouvent leurs origines dans l'industrie de la téléphonie. Conceptuellement, le modèle établi par l'industrie téléphonique dans les années 1920 est l'ancêtre du format MP3.* »

Chapitre 3. Les formats ouverts en informatique

3.1 Définition

Les formats ouverts sont l'équivalent informatique du système métrique. Avant le système métrique, comme on a pu le voir au premier chapitre, tous les systèmes de mesures étaient « propriétaires » et les spécifications plutôt cachées ce qui était aussi la porte ouverte à diverses tromperies et fraudes. Le système métrique, étant, quant à lui, un système ouvert dont les spécifications sont diffusées largement, notamment à l'école pour la France.

À l'instar du système métrique, les spécifications des formats ouverts sont accessibles de tous, gratuitement, sans restriction et peuvent être utilisées par tous sans autre formalité. Ils garantissent ainsi non seulement la pérennité des documents mais aussi le fait qu'ils puissent être ouverts et travaillés quels que soient les systèmes (interopérabilité).

La notion est très clairement définie par le droit français dans l'article 4 de la loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (la fameuse LCEN) :

> *« On entend par standard ouvert tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre. »*

Les autres définitions légales, y compris celle de l'Union européenne, sont très voisines, bien que rédigées différemment.

Les formats ouverts sont donc des standards accessibles à tous, uniformisés et dont les spécifications techniques sont publiques. Ils ne sont pas forcément liés à un logiciel spécifique, ce serait même plutôt une faiblesse. Cela signifie aussi qu'il n'y a donc pas de format LibreOffice, de format Inkscape ou encore de format Gimp. Mais des formats que ces logiciels utilisent nativement parmi d'autres.

En France, le référentiel général d'interopérabilité (RGI) détermine les formats acceptés, arrêté du 22 avril 2016, téléchargeable sur le site legifrance.fr (voir la bibliographie).

3.2 Qualités des formats ouverts

Les formats ouverts ont, au moins, quatre qualités.

➲ **Interopérabilité** : c'est le fait pour un fichier de pouvoir être ouvert, et travaillé quel que soit l'OS, actuel, passé ou futur. Un fichier interopérable pourra ainsi être lu et travaillé sur les Linux, Mac OS ou Windows. C'est une notion essentielle, notamment dans le cadre d'échanges d'informations médicales.

➲ **Fiabilité** : parce qu'on sait comment est construit le fichier, on sait aussi comment le travailler (d'un point de vue informatique s'entend). On peut aussi mieux vérifier son intégrité et corriger un fichier corrompu.

➲ **Pérennité** : dans vingt, trente, cent ans, on pourra toujours accéder aux fichiers et les retravailler car, si les premiers logiciels qui ont servi à les créer n'existent plus ou ne sont plus téléchargeables nulle part, il sera toujours loisible de partir des sources pour refaire un logiciel ou en adapter un susceptible de les utiliser. Ce n'est pas, ou très difficilement, le cas des fichiers aux formats privateurs, notamment lorsqu'ils sont anciens[13].

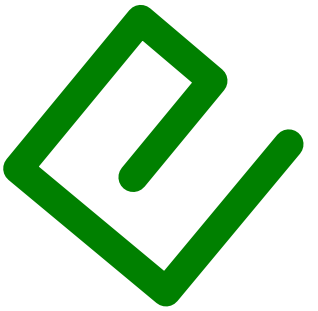
➲ **Sécurité** : d'un point de vue technique, la possibilité d'accéder aux sources du logiciel en permet l'examen par des experts indépendants et facilite notamment la découverte et la correction de failles de sécurité. D'un point de vue juridique, un format ouvert est une garantie de ne pas subir de problème en matière de propriété intellectuelle (plagiat, brevet, etc.).

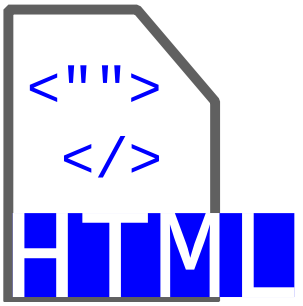
Les professionnels de la gestion électronique de documents et les archivistes préfèrent les formats ouverts pour ces raisons.

3.3 Les fichiers aux formats ouverts sont partout

On utilise tous les jours des formats ouverts : dans nos téléphones dont les images sont au format svg, en utilisant internet évidemment pour consulter des sites dont les pages sont au format html, en écoutant de la musique au format mp3 ou encore en ouvrant ou en créant des fichiers au format pdf.

Les liseuses actuelles[14], quelles que soient leurs marques lisent des fichiers epub, soit directement, soit de façon détournée pour une liseuse de marque américaine qui a son propre format privateur.





13 Voir l'article de Wikipédia : L'Âge sombre numérique
https://fr.wikipedia.org/wiki/%C3%82ge_sombre_num%C3%A9rique.

14 Le format EPUB est devenu un standard en 2007. Le format EPUB 3 date de 2011.

Chapitre 4. Quel format pour quel fichier et par quelle application ?

Sans détailler pour tous les types de fichiers, ce qui est assez difficile, quelques exemples de formats ouverts courants[15] et dont on a déjà probablement vu les extensions ou qu'on a déjà utilisés même sans savoir qu'il s'agissait de formats ouverts.

> ⚠ **À noter** : ces formats ne sont pas spécifiques à un logiciel particulier, libre ou non. Ils sont spécifiques à des types de logiciel : un tableur génère des classeurs tandis qu'un logiciel de traitement de texte est fait pour créer et travailler des textes par exemple et qu'un logiciel de traitement d'image peut être amené à travailler plusieurs formats.

4.1 Bureautique et assimilé

En bureautique, la norme générique est l'ODF, Open Document File, qui est gérée par le consortium international OASIS. Ces formats sont basés sur le langage XML.

À l'intérieur de cette norme, les extensions sont subdivisées par catégorie de document :

- **base de données** : ODG, Open Document Base ;
- **dessin**[16] : ODG, Open Document Graphic et OTG, Open Template Graphic pour les modèles[17] de dessin ;
- **présentation** : ODP, Open Document Presentation et OTP, Open Template Presentation pour les modèles de présentation ;
- **tableur** : ODS, Open Document Spreadsheet et OTS, Open Template, Spreadsheet pour les modèles ;
- **traitement de texte** : ODT, Open Document Text et OTT, Open Template Text pour les modèles.

4.2 Image

Outre le format ODG dont on a parlé ci-dessus, il y a plusieurs types de formats d'image.

Concernant **les images matricielles**, à savoir celle dont l'unité de base est le pixel et qui sont, par exemple, le type d'image généré par les appareils photo numériques et les smartphones, on indiquera notamment les formats :

- BMP : Windows Bitmap, format d'image très lourd et, de fait assez peu utilisé ;

15 Classement qui m'est personnel, d'autres sources peuvent les répartir autrement.

16 À probablement parler, le format ODG n'est pas franchement de la bureautique, mais il fait partie du « lot » des extensions gérées par OASIS. Par ailleurs, l'application Draw est incluse dans les deux suites bureautiques libres LibreOffice et OpenOffice.

17 **À noter** : le format ODG est un format de travail qui doit ensuite être exporté pour être utilisé dans un autre type de fichier.

➲ GIF : dont la palette de couleurs est très réduite (256) mais qui permet de faire des animations, c'est d'ailleurs uniquement pour cela que ce format est encore utilisé ;

➲ OpenJPEG : variante ouverte du JPEG qui, lui, n'est pas vraiment un format ouvert, ce format est assez peu utilisé bien qu'il gère mieux les images que le JPG, Gimp depuis sa version 2.10 supporte le format OpenJPEG ;

➲ PNG, qui a le mérite de générer des images de bonne qualité et de gérer la transparence[18] mais au prix d'un poids informatique accru.

Les images vectorielles sont basées sur le langage XML, comme les fichiers des suites bureautiques. Une image vectorielle est une succession de points dans les coordonnées sur la feuille sont en enregistrées. Les cartes géographiques par exemple sont des images vectorielles mais pas que. Le format ouvert des images vectorielles est le SVG.

L'intérêt : les fichiers de dessin sont légers et les images SVG restent « vivantes » et modifiables.
Inconvénient : dès lors qu'on veut un rendu « réaliste » les fichiers SVG deviennent très lourds en raison du très grand nombre de coordonnées que cela génère. Dans ce cas, le format PNG est plus adapté.

Fichiers graphiques : ces formats sont des documents de travail qui ne peuvent pas être utilisés tel quel notamment parce qu'ils peuvent comporter des calques (plusieurs couches de document) qui permettent de mieux travailler les images. Outre le format ODG déjà signalé, il y a le format XCF qui est, entre autres, le format de travail du logiciel de traitement d'image GIMP.

4.3 Diffusion, échanges et conservation

Pour échanger des fichiers et les diffuser, s'il s'agit de documents qui n'ont pas besoin d'être retravaillés, deux formats sont à retenir, le pdf et l'epub.

➲ PDF : standard de fait puis devenu ouvert et faisant l'objet d'une norme ISO, le format PDF peut incorporer les polices du document pour une meilleure conservation, par contre, selon la façon avec laquelle le document a été conçu il n'est pas forcément très lisible des dispositifs de lecture Braille ou audio.

⚠ **À noter** : les deux suites libres OpenOffice et LibreOffice permettent de générer des pdf hybrides qui contiennent la version originale du document et pèsent deux fois plus, logiquement, qu'un pdf « normal ». En ouvrant ce type de format à partir de l'application, on peut modifier le fichier d'origine et générer un nouveau pdf actualisé.

➲ EPUB : c'est le format privilégié des liseuses numériques. Le format epub est basé sur le XML. Il est très accessible. C'est un format à privilégier pour les dispositifs de lecture Braille ou audio.

⚠ À savoir : le format epub 3 exporte les formules de mathématiques d'une façon lisible par les dispositifs d'assistance. LibreOffice exporte correctement et nativement en epub 2 et 3 depuis la version 6.1.

18 Globalement, une image est carrée ou rectangulaire, la gestion de la transparence permet d'avoir un fond invisible autour des éléments de l'image et donner l'impression qu'elle a une autre forme : ronde, ovale, triangulaire, forme des sujets, etc.

Concernant l'échange de données « pures » afin de pouvoir les utiliser, on signalera trois types de formats, qui ne sont pas les seuls mais sont très utilisés :

- CSV : qui a été signalé précédemment car c'est un standard de fait ;
- JSON : il est utilisé notamment par les navigateurs internet et les agrégateurs de flux rss pour récupérer des informations, listes et structures des marques-pages par exemple ;
- EDF : pas très connu du grand public, c'est un format d'échanges et de stockages des données médicales ou biologiques utilisable quel que soit le constructeur du matériel et sa version. Il permet d'échanger de manière fiable des données entre équipes qui travaillent sur des machines différentes.

Pour l'archivage ou la compression de données qui permettent aussi d'échanger plusieurs fichiers réunis en un seul fichier, on peut signaler :

- 7Z : lu de la plupart des logiciels d'archivage depuis 2011 ;
- TAR : format d'archivage de données non compressées ; on le retrouve souvent plutôt sous la forme compressée avec l'extension du type de compression, par exemple : .TAR.7Z ou .TXZ, .TAR.Z ou .TAZ, .TAR.GZ ou .TGZ ;
- ZIP : probablement le plus connu de ces types de format et qui fait qu'on parle souvent plus facilement d'un fichier « zippé » que d'un fichier compressé.

4.4 Multimédia : son et vidéo

En matière de son, le format le plus courant est le format MP3, entré dans le domaine public en avril 2017 qui a l'avantage de générer des fichiers légers, et l'inconvénient de trop compresser les données et donc de ne pas être adapté à tous les types de sons ou de musique dont il ne reflètera pas les nuances avec la précision voulue.

Néanmoins, sous l'extension générique OGG[19] (comme pour les ODF des fichiers bureautiques), il existe un format ouvert et libre pour le son et un pour la vidéo :

- OGA ou OGG Vorbis, plus performant que le MP3 et moins populaire est celui utilisé par Wikipédia, il est, notamment, lisible des navigateurs internet récents ;
- OGV ou OGV Theora pour la vidéo.

À noter : si le format Mpeg est ouvert car il fait l'objet de spécifications ISO, il n'est pas libre.

19 Ces extensions nécessitent des codecs spécifiques d'où les noms plus courants d'OGG Vorbis pour le son et d'OGV Theora.

Chapitre 5. HTML, CSS, et XML, exemples de code

5.1 Le HTML, structure et exemples de balise

Illustration 9: logo officiel de l'HTML 5, W3C

Le HTML (Hypertext Markup Language) est le langage de balisage avec lequel sont faites toutes les pages internet. Il permet de séparer la mise en forme du contenu, d'incorporer des liens hypertextes, qui sont la base de la navigation sur internet et d'inclure d'autres médias : images, sons, vidéos, etc. Puisque le « projet, baptisé "World Wide Web", a été initialement conçu et développé pour que les physiciens et les physiciennes travaillant dans les universités et les instituts du monde entier puissent s'échanger les informations dont ils avaient besoin. »[20], il fallait que cela soit parfaitement interopérable afin que des chercheurs aux machines très diversifiées puissent y accéder sans problème (et facilement, car ce n'étaient pas des informaticiens).

Le html est un langage de balise : absolument tous les éléments sont entre des éléments, les balises, qui comportent les informations nécessaires pour afficher le contenu comme il faut. Ces balises doivent être fermées. Ainsi, quand il y a une balise, un élément entre des chevrons < et >, on a forcément un balise de fin contenue entre les mêmes signes mais dont le nom est précédé d'une barre oblique /, par exemple, pour indiquer le corps de la page à afficher :

```
<body> contenu </body>
```

Une page html contient obligatoirement certains éléments pour être bien interprétée, à commencer par le *doctype* qui comporte les éléments suivants, ici en html 5 :

```
<!doctype html>
<html lang="fr">
<head>
  <meta charset="utf-8">
  <title>Titre de la page</title>
  <link rel="stylesheet" href="style.css">
  <script src="script.js"></script>
</head>
```

Ces éléments indiquent la langue du document, ici fr pour la langue française puis dans le « head » (qui n'est pas l'entête des pages que l'on voit dans son navigateur), la table de caractères, ici Unicode (utf8), le titre de la page, sa mise en forme qui est le lien (link) vers la ou les feuilles de style .css (stylesheet) et les scripts éventuels (systématiques en fait). La balise <html> ouvre la page et </html> la clôt.

Le contenu de la page, ce qui est affiché à l'écran figure entre les balises <body> et </body> et en dessous du « head » :

- on a un contenu entre balises du type de celui ci-dessous qui est l'entête des pages html, header, donc visible ;

20 Giampetro (Marina), Noyes (Dan), O'Luanaigh (Cian), Les 20 ans d'un web libre et gratuit, CERN 2015.

- la balise <div class> indique de quel bloc il s'agit et sa mise en forme ;
- la balise <a> est celle qui indique qu'il y a un lien, dans le cas présent, l'accueil du site ;
- la balise <h1> qu'il s'agit d'un titre de niveau 1 (heading en anglais).

On notera que toutes ces balises sont fermées par : </> dans l'ordre inverse de leur « apparition » dans le code : </h1>, et

```
<header id="header">
<div class="accueil">
<a rel="start home"
href="https://numericoach.net/" title="Accueil"
><h1>Votre coach numérique</h1> </a>
</div>
</header>
```

Cette partie de la page, à savoir son entête ou header est ensuite fermée </header>. On peut (on a, en fait) avoir ainsi plusieurs éléments entre cet entête et la fin de la page.

5.2 Le XML

Le XML est un langage de balisage développé par le W3C qui en gère les spécifications. Le langage est très proche de l'html, à la différence qu'on peut y ajouter ses propres balises. C'est ce qui « tourne derrière » les fichiers bureautiques actuels, les fichiers au format epub ou encore les fichiers json. Il est très bien interprété par les machines transcriptrices en Braille ou en audio.

Ci-dessous, un exemple du début du fichier de contenu, content.xml d'un fichier odt avec la première ligne qui est son « doctype » et le début de l'équivalent du « head » en html. Les balises s'écrivent de façon très similaire à celles du html. Historiquement, le html est antérieur au xml, mais il est considéré comme un dérivé du xml.

```
<office:document-content
office:version="1.2"><office:scripts/>
<office:font-face-decls><style:font-face
style:name="OpenSymbol2" svg:font-
family="OpenSymbol" style:font-charset="x-
symbol"/>
</office:font-face-decls>
<office:automatic-styles>
<style:style style:name="P1"
style:family="paragraph" style:parent-style-
name="Header">
<style:text-properties officeooo:paragraph-
rsid="001a98fb"/>
</style:style>
```

Ci-dessous un paragraphe tiré du fichier de contenu, content.xml d'un fichier odt, « p » indique un paragraphe comme avec le html.

```
<text:p
text:style-name="P59">Présentation</text:p>
<text:p text:style-name="P60">Pourquoi mêler
l'histoire des systèmes de mesures à celle des
formats informatiques ?</text:p>
```

5.3 Les CSS

Les CSS pour Cascading Sheet Style ou Feuilles de styles en cascade en version française, sont les fichiers qui s'occupent de préciser la mise en forme d'un site. De cette façon, le contenu et la forme étant séparés, on peut changer très facilement[21] et très rapidement l'aspect d'un site en changeant ses feuilles de styles.

Ci-dessous, un exemple de code CSS, il est balisé, à l'instar du HTML et du XML.

```
<?xml-stylesheet href="#s1"
   type="text/css"?>
<doc>
  <s id="s1">
    s { display: none }
    p { display: block }
    p.note { color: red }
  </s>
  <p>Some text... </p>
  <p class="note">A note... </p>
</doc>
```

On retrouve cette notion de feuille de styles dans les documents ODF, à cette différence que, pour l'utilisateur, elles ne sont pas conçues à l'extérieur du document, mais dans le document. Mais, le fichier lui-même est en fait la réunion, ou compression, de six fichiers dont le fichier styles.xml qui est l'équivalent de la feuille de styles CSS d'un site internet.

Illustration 10: comment se décompose un document odf.

21 Une fois que la ou les feuilles de styles ont été faites, s'entend. La conception des feuilles elles-mêmes nécessite des connaissances et du métier.

Chapitre 6. Est-ce que ?

6.1 Est-ce qu'un fichier protégé est sûr ?

Format ouvert n'est pas synonyme de fichier, et donc de contenu ouvert à tous les vents. En fait, le fait même qu'il soit ouvert permet justement de vérifier la qualité de la protection du fichier (qui peut être très forte[22]) et garantir son intégrité parce qu'on sait comment il est fabriqué.

6.2 Est-ce que les formats ouverts sont ouverts par toutes les applications équivalentes ?

Cela dépend par exemple MS Office pour Mac OS ne reconnaît les formats ouverts en bureautique que depuis la version 2016, MS Office Windows depuis 2007. Mais comme OpenOffice[23] et LibreOffice existent pour tous les OS, et que les vieilles versions sont toujours téléchargeables, les propriétaires desdits ordinateurs peuvent ouvrir des formats bureautiques ouverts. Les formats multimédias, sons et vidéos, ouverts sont assez peu utilisés, hormis le mp3, et on trouve assez peu de logiciels qui les ouvrent.

6.3 Est-ce que les formats ouverts sont réservés aux logiciels libres ?

Non d'ailleurs on les retrouve partout, qu'on le sache ou non. En fait les logiciels privateurs auraient même intérêt à les utiliser, ce qui leur éviterait une part de développement et de bugs associés.

6.4 Est-ce qu'un fichier au format ouvert est moins bien qu'un fichier au format privateur ?

Pourquoi serait-il moins bien ? Parce qu'on en connaît les sources ? Parce qu'on peut, souvent mais pas systématiquement, les générer et les travailler avec des logiciels gratuits alors que les formats privateurs sont souvent, mais pas toujours, le fait de logiciels parfois fort chers ?

Il n'y a aucune raison objective de préférer un format privateur quand il existe un format ouvert, qui, lui, sera pérenne.

6.5 Est-ce qu'un logiciel gratuit utilise forcément des formats ouverts ?

Non.

D'une part, un logiciel peut être gratuit sans être libre, par exemple le lecteur de pdf le plus répandu sous Windows est gratuit mais n'est certainement pas un logiciel libre.

22 À titre d'exemple, il est extrêmement difficile de casser la protection par mot d'un passe d'un fichier protégé généré par LibreOffice.

23 OpenOffice semble en stase depuis 2014, il est donc préférable d'opter pour LibreOffice qui est toujours développé. Néanmoins, les forums d'aide d'OpenOffice sont toujours vivants et la plupart des échanges, conseils et solutions qui y sont délivrés sont valables également pour la suite LibreOffice.

D'autre part, si beaucoup de logiciels libres sont distribués gratuitement, la gratuité n'est pas une condition sine qua none des logiciels libres. Il est tout à fait possible de développer un logiciel et de le vendre tout en le mettant sous une licence libre.

6.6 Est-ce qu'on est obligé d'utiliser des formats ouverts ?

Légalement non, mais le référentiel général d'interopérabilité (RGI) de l'administration française les recommande pour garantir l'interopérabilité. Accessoirement (?!?), les professionnels de l'archivage ont une nette tendance à les préférer. Donc, on n'est pas obligé, mais on a intérêt à les utiliser.

6.7 Est-ce qu'on peut m'obliger à donner des fichiers dans un format privateur ?

En France, en aucun cas un service administratif ne saurait exiger un format privateur, en aucun cas et ce d'autant plus que la plupart ne sont pas acceptés dans le RGI.

Ça c'est la théorie. L'administration française, pas plus que l'européenne d'ailleurs, qui, les préconise aussi, ne met vraiment en pratique ces recommandations.

6.8 Est-ce que l'administration peut m'imposer un format privateur ?

En théorie non, en pratique, notamment dans le cadre de certains pdf, il arrive qu'une administration conçoive des formulaires pas interopérables et que l'on ne peut ouvrir qu'avec un lecteur de pdf qui n'a pas été développé pour Linux.

Dans ce cas, il faut le signaler à l'administration concernée et aux associations d'usagers des services publics, ou encore aux associations qui promeuvent le logiciel libre par exemple. Ça n'est pas forcément immédiatement utile, mais, si on ne fait pas remonter le problème, il n'y a aucune chance qu'il soit réglé un jour. À défaut d'autre solution, si le dialogue avec l'administration n'aboutit pas, squatter un ordinateur qui peut ouvrir le document.

6.9 Est-ce que l'Open XML est libre et ouvert ?

Le standard Office Open XML (OOXML) qui est une norme de formats bureautiques développée par Microsoft a fait l'objet, d'une norme ISO, il est donc théoriquement ouvert.

Cette norme est contestée à plus d'un titre :

- elle avait fait l'objet d'un premier refus de la part de l'ISO (organisme de normalisation international) ;
- il existe déjà une norme bureautique performante : l'ODF ;
- elle n'est pas libre ni parfaitement interopérable ;
- elle est extrêmement complexe et, d'une manière générale, assez peu mise en application complètement, même par Microsoft.

Pour le RGI, version 2015, l'OXML est en observation et seulement sa variante stricte.

Chapitre 7. Crédits, droits, versions

7.1 Crédits

Gnu du « logo » interoperabilité : GNUx Gnunguin d'ecogex[24], openclipart.org

Logo SVG : https://www.w3.org/Graphics/SVG

Logo HTML5 : https://www.w3.org/html/logo

Concernant les autres illustrations, soit j'en suis l'auteure (la plupart), soit elles sont dans le domaine public sans qu'il m'ait été possible d'en trouver l'auteur (essentiellement le logo mp3).

7.2 Droits

Licence CC BY-SA 4.0. Attribution - Partage dans les Mêmes Conditions 4.0 International.

7.3 Versions

Date	Objet
juin 2018	Première rédaction.
janvier 2019	Mise à jour notamment : système international, OpenJPEG et EPUB. Corrections mineures.

24 http://ecogex.com/the-other-linux-logo

Bibliographie

Métrologie

- Bureau International des Poids et Mesures (BIPM), *Réalisation pratique des définitions des principales unités.*

  https://www.bipm.org/fr/publications/mises-en-pratique

- La métrologie.com, *Ressources.*

  https://www.lametrologie.com/ressources

- Metrodiff.org, *Extraits des cahiers de doléance de 1789 sur les poids et mesures.*

  http://www.metrodiff.org/cmsms/index.php/hi%EF%AC%86oire/cahiers-de-doleances.html

- La Métrologie française, *Unités de mesure – le Système international d'unités.*

  http://www.metrologie-francaise.fr/fr/si/unites-mesure.asp

- Paucton (Alexis-Jean-Pierre), *Métrologie ou Traité des mesures, poids et monnoies des anciens peuples, des modernes*, Veuve DESAINT, libraire, 1780.

  Fac-simile pdf : http://gallica.bnf.fr/ark:/12 148/bpt6k55 491 755?rk=21 459;2

- Prieur (Claude-Antoine), *Mémoire sur la nécessité et les moyens de rendre uniformes les mesures d'étendue et de pesanteur[...]* présenté à l'Assemblée nationale en 1790.

  Version epub, odt et pdf : https://numericoach.net/Memoire-sur-la-necessite-et-les-moyens
  Fac-simile pdf : http://gallica.bnf.fr/ark:/12 148/bpt6k62 066m

- Romé de L'Isle (Jean-Baptiste Louis de), *Métrologie, ou Tables pour servir à l'intelligence des poids et mesures des anciens, et principalement à déterminer la valeur des monnoies grecques et romaines d'après leur rapport avec les poids, les mesures et le numéraire actuel de la France*, Imprimerie de Monsieur, 1789.

  Fac-simile pdf : http://gallica.bnf.fr/ark:/12 148/bpt6k9 643 199s?rk=300 430;4

- *Unités de mesure anglo-saxonnes*, article de Wikipédia.

  https://fr.wikipedia.org/wiki/Unit%C3%A9s_de_mesure_anglo-saxonnes

Un peu d'histoire de l'informatique

- *Chronologie de l'informatique*, article de Wikipédia.

  https://fr.wikipedia.org/wiki/Chronologie_de_l%27informatique

- Frechin (Jean-Louis), *Petite histoire de l'informatique*, ENSCI.

  https://www.ensci.com/creation-industrielle/ateliers-de-projets/jean-louis-frechin/petite-histoire-de-linformatique

- Pépin (Guénaël), *Software Heritage, le défi d'une archive mondiale du code source*, Nextinpact, 14 juin 2018.

  https://www.nextinpact.com/news/106 719-software-heritage-defi-dune-archive-mondiale-code-source.htm

Formats

- Âge sombre numérique, article de Wikipédia.

  https://fr.wikipedia.org/wiki/%C3%82ge_sombre_num%C3%A9rique

- April, dépliant *Formats ouverts.*

  https://media.april.org/docs/formats_ouverts/mini-site/formats-ouverts.html

- *Arrêté du 20 avril 2016 portant approbation du référentiel général d'interopérabilité.*

  https://www.legifrance.gouv.fr/eli/arrete/2016/4/20/PRMJ1 526 716A/jo

- Becquet (Jean-Christophe), *Quatre saisons de formats ouverts et d'interopérabilité*, apitux.org, 25 juin 2018.

  http://www.apitux.org/index.php?2008/01/22/137-quatre-saisons-de-formats-ouverts-et-d-interoperabilite

- Dutailly (Isabelle), *Liste de formats ouverts*, avril 2018.

  Pdf hybride : https://numericoach.net/Liste-de-formats-ouverts

- Feilner (Markus), *Incompatibilités et incohérences du format Office Open XML*, trad. Jean-François Nifenecker ; 2014

  https://numericoach.net/Incompatibilites-et-incoherences-du

- Format ouvert, article de Wikipédia.

  https://fr.wikipedia.org/wiki/Format_ouvert

- Giampetro (Marina), Noyes (Dan), O'Luanaigh (Cian), *Les 20 ans d'un web libre et gratuit*, CERN 2015.

  https://home.cern/fr/about/updates/2013/04/twenty-years-free-open-web

- *Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique.*

  Version initiale :
  https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000 000 801 164&categorieLien=id
  Version en vigueur au 19 juin 2018 :
  https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000 000 801 164&dateTexte=20 180 619

- Pour les formats ouverts ! Site qui semble en stase depuis 2015, mais comporte nombre d'informations utiles.

  https://formats-ouverts.org

- *Référentiel général d'interopérabilité 2015*, version pdf.

  http://references.modernisation.gouv.fr/sites/default/files/Referentiel_General_Interoperabilite_V2.pdf

HTML, CSS, et XML, exemples de langages

- Simon-K, *HTML5 : structure globale du document*, alsacreations.com, mai 2014.

  https://www.alsacreations.com/article/lire/1374-html5-structure-globale-document.html

- W3C, *Spécifications du HML5.2*, en anglais.

  https://www.w3.org/TR/html52

- W3C, *XML Technology*, en anglais.

  https://www.w3.org/standards/xml

- W3C, *Feuilles de style en cascade.*

  https://www.w3.org/Style/CSS/Overview.fr.html

Index des illustrations

Illustration 1: muid de Ponte Puñide, Musée archéologique d'Espagne, photo Luis García, source Wikipédia.....5
Illustration 2 : page de titre du mémoire de Prieur, souc re BnF Gallica.....6
Illustration 3: usage des nouvelles mesures, France circa 1800, source BnF Gallica.....7
Illustration 4: poids en laiton, Angleterre 1826, source Wikipedia,photographe inconnu.....8
Illustration 5: extrait de l'introduction du manuel de tricot de Mlle Baillaud, source Bnf Gallica.....9
Illustration 6 : modèle de chaussette pour homme tiré du Manuel de Mlle Baillaud, source : BnF Gallica.....10
Illustration 7 : disquettes 8, 5 ¼ et 3,5 pouces, Wikipédia, photographe George Chernilevsky.....11
Illustration 8 : capture d'écran du site de Software Heritage, juin 2018.....12
Illustration 9: logo officiel de l'HTML 5, W3C.....21
Illustration 10: comment se décompose un document odf.....23