Fichiers, dossiers, formats et interopérabilité : cours

3. les formats

Pourquoi et comment

Chacune des vues va contribuer à la représentation numérique d'un document textuel, mais, selon le domaine d'application, elle sera p importante ou nécessaire. La première question est : **que veut-on représenter en vue de quels usages ?** Des choix techniques seront répondre à la question : **comment représenter ?** Cette distinction entre le quoi et le comment est, en informatique comme dans beau sciences, une approche essentielle des problèmes.

Différents formats pour différents usages

Les choix effectués pour répondre à la question **comment représenter des documents textuels** aboutissent à des **formats** de représe connaissez sans doute certains de ces formats précisés avec les abréviations suivantes :

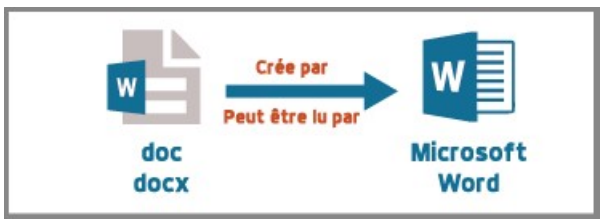
- le format `txt` pour les textes,
- le format `doc` ou le format `docx` du traitement de textes Word,
- le format `odt` des traitements de textes LibreOffice ou OpenOffice,
- le format `pdf` pour l'impression,
- le format `html` pour les hypertextes.

Vous noterez, que pour de mêmes usages, comme la composition de documents textuels, il existe des formats différents comme `doc` (également que les formats évoluent avec les usages et les technologies. Par exemple `HTML` a été défini dans des versions successives (années 90 jusque `HTML5` , paru en 2014.

Formats et logiciels

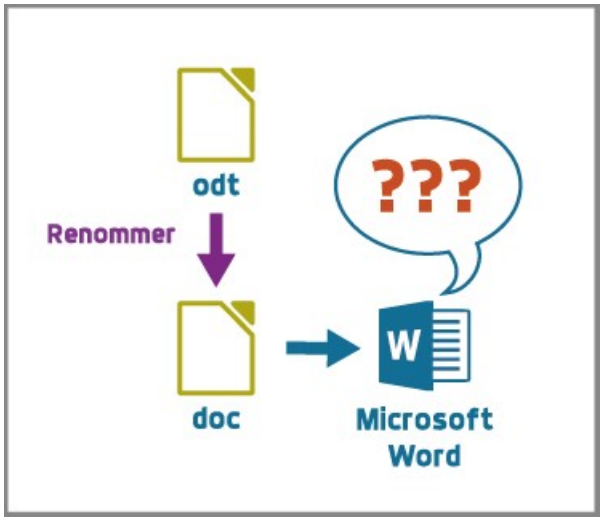
Nous avons expliqué la relation forte entre le choix de la représentation et les traitements qui peuvent être faits sur une donnée numéi vue très concret, cette relation se traduit souvent par la liaison entre un format et une application spécifique d'un éditeur logiciel. Par document textuel au format `doc` est associé au logiciel `Word` de Microsoft. Il aura souvent été créé avec ce logiciel et pourra être lu e logiciel.

Un document dans un format pourra être stocké dans un fichier. Pour des raisons historiques, le format d'un document est souvent pré du nom de fichier constituée de trois ou quatre lettres après le point. On désigne même abusivement un format par cette extension, cc fait précédemment en parlant de format `doc` par exemple. Cette extension peut être vue comme une métadonnée qui dit : "le docume respecte le format de représentation de documents utilisé par le logiciel `Word` ".

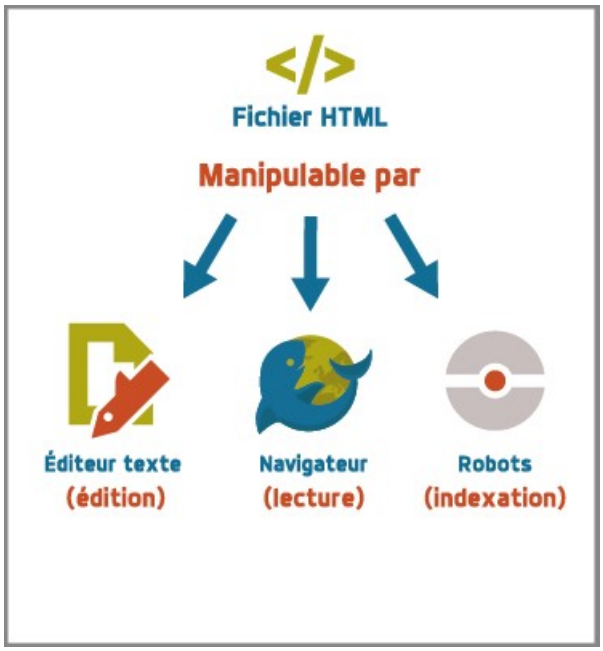


Si nous avons un document textuel au format `odt` , il ne suffit pas de le renommer avec l'extension `doc` pour le rendre lisible par Wo **conversion** d'un format dans l'autre, opération qui peut être difficile voire impossible. Pourquoi ? Pour au moins deux raisons :

1. Tout d'abord, les choix qui ont été opérés pour définir les formats ne sont pas toujours compatibles. On peut donc perdre des in cette conversion.
2. Ensuite, les choix ne sont pas toujours rendus publics. On ne peut donc pas toujours écrire de programme de conversion.



Par ailleurs, un document textuel dans un format peut être parfois manipulé avec des logiciels différents pour des besoins différents. l fichier `html` peut être ouvert par un navigateur pour le visualiser. Le même fichier peut être ouvert avec un éditeur de texte pour le m l'avez vu dans le cours du Web, il sera également manipulé par les robots des moteurs de recherche qui contribuent à indexer le web.

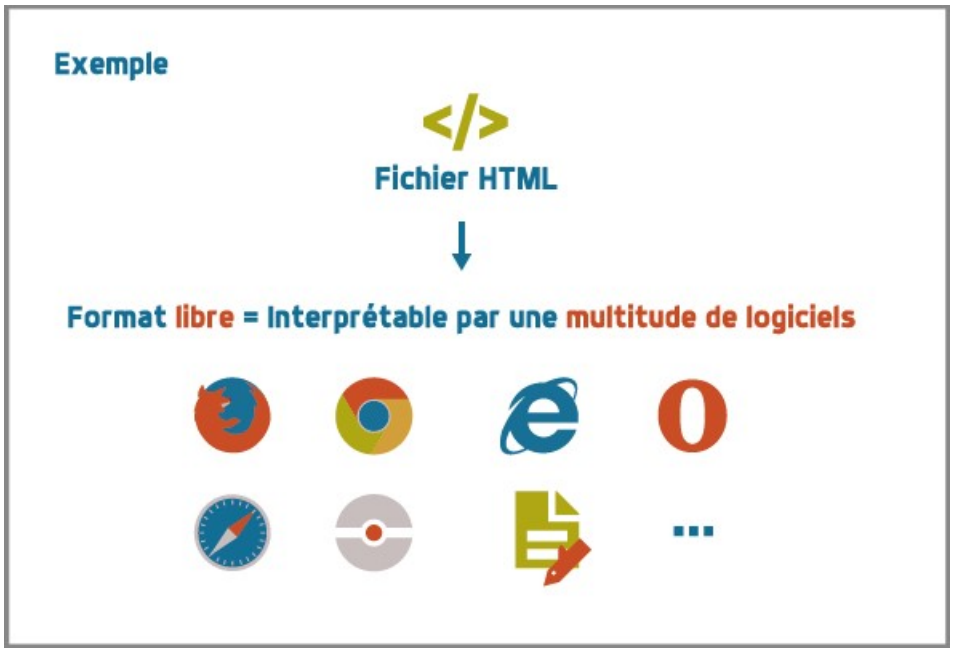


Ouvert ou propriétaire

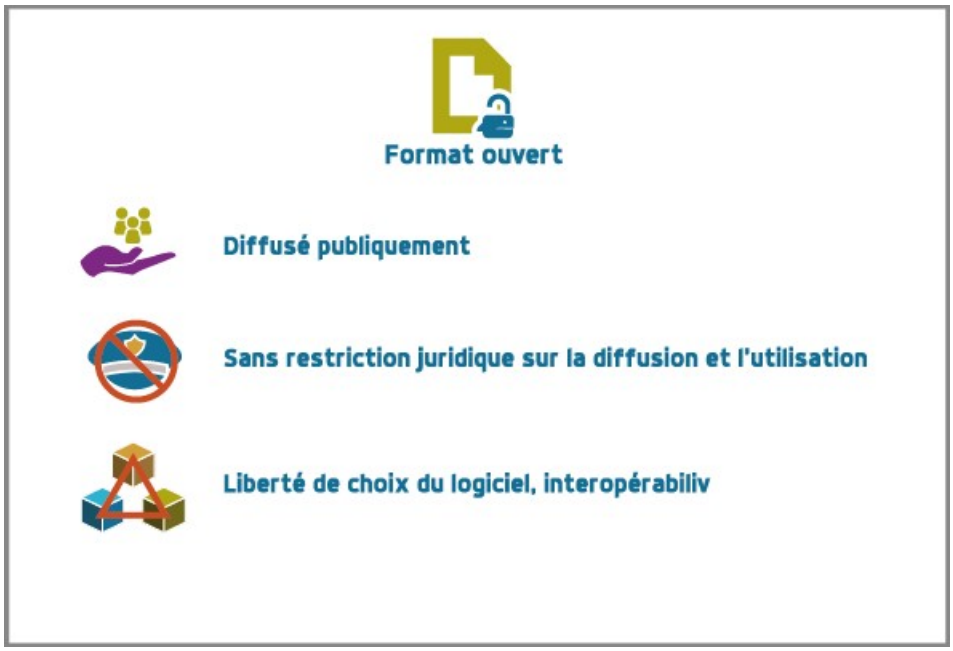
Le processus de choix de représentation et de définition d'un format est complexe et coûteux. Il peut être aussi stratégique d'un point commercial. Dès lors, les créateurs ont la possibilité de le rendre disponible pour tous librement ou non, de le cacher ou de le protége

Formats ouverts

On parle de **format ouvert** si le format est diffusé publiquement.



Par exemple, vous pouvez accéder librement sur le Web à la définition du format `HTML5` . De plus, aucune entrave légale n'accompag format ouvert et de ce fait, un format ouvert n'est pas lié à un logiciel. En effet, plusieurs logiciels différents peuvent librement lire ou informations représentées dans ce format. On facilite ainsi l'interopérabilité. Par exemple, le format `html` est utilisé par de nombreux delà du Web.



Formats fermés

On parle de **format fermé** ou propriétaire lorsque des restrictions d'accès et/ou d'utilisation s'appliquent.

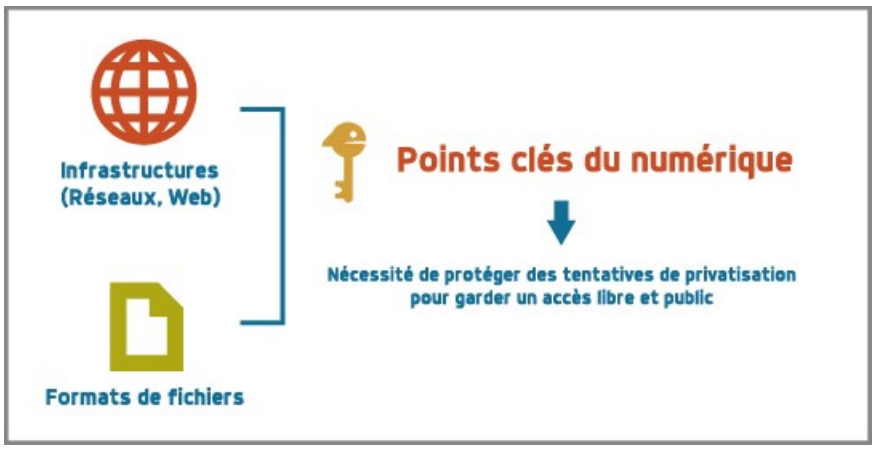
formats fermés exemple

Être propriétaire d'un format très répandu donne une puissance économique très importante dans notre monde numérique et une posit effet, la conversion étant impossible, une mise en concurrence est rendue très improbable et les utilisateurs sont alors contraints d'util associé. Si `HTML` avait été un format fermé, sans doute le web serait-il très différent de celui d'aujourd'hui ou n'existerait peut-être mê

formats fermés

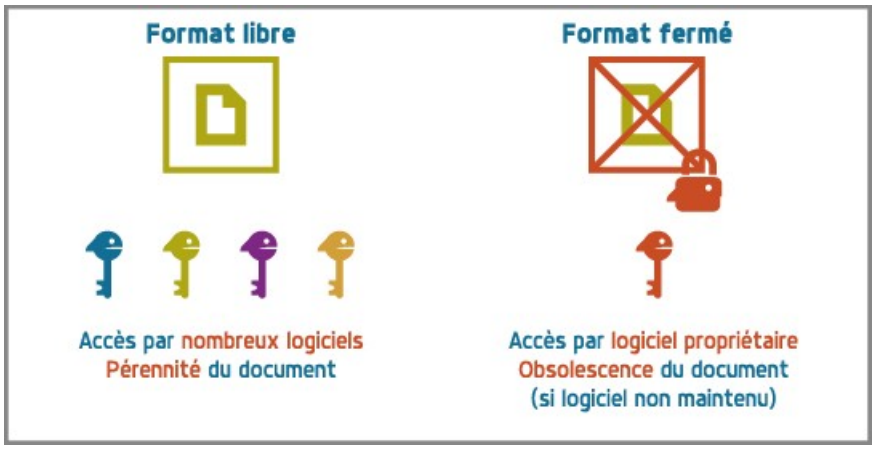
Une minute citoyenne

Le numérique est aujourd'hui un facteur de développement économique important. Ce développement repose en partie sur des infrast réseaux, le web, étudiés dans les semestres précédents. Les organisations publiques mondiales, pour ne pas freiner ce développement normes et étudient des garanties pour un accès neutre et de qualité à ces infrastructures. Les normes du W3C sont un exemple. Le dél neutralité du net est une autre illustration. Par le passé et encore aujourd'hui plusieurs entreprises, par des moyens techniques ou com d'accaparer ce que beaucoup pensent être soit un bien public soit des données personnelles. Mais ces infrastructures ne sont pas le seu numérique. La question des formats de représentation des données entre évidemment dans l'éventail des possibilités de contrôler l'éc



Lorsque vous enregistrez un document dans un certain format, c'est un peu comme si vous rangiez un objet dans une boite. Si le form protégé, alors cela signifie que lorsque vous voulez retrouver votre objet vous devez vous adresser à un tiers qui lui seul a l'autorisati La question de savoir si l'objet vous appartient toujours se pose donc, ou encore celle de la liberté d'utiliser cet objet.

Transposée dans le monde numérique, cette image signifie que limiter cet accès a de nombreuses conséquences. L'interopérabilité est : un document dans un format propriétaire, ne peut être librement utilisé dans un autre logiciel. La liberté des utilisateurs est égaleme échangeant avec un format propriétaire, vous forcez vos interlocuteurs à utiliser un logiciel précis.



Enfin, lorsqu'il s'agit de données sensibles ou devant être archivées pour une très longue durée, l'usage de formats propriétaires repos peuvent disparaître ou changer leur règles d'utilisation...

Comme pour les infrastructures, l'État et bien d'autres organisations sont conscientes de ces difficultés. Elle produisent souvent des d pour inciter à utiliser des formats ouverts et libres. Mais il est bien plus difficile de convaincre les utilisateurs souvent plus enclins à habitudes, résultant souvent de nombreux efforts d'apprentissage.

De votre côté, recevoir une formation indépendante des outils, donc plus fondamentale peut contribuer à être moins dépendant et mo monde numérique. Mais cela demande un effort particulier, une attente moins centrée sur l'immédiat et l'utilitaire, un peu moins pers conscience d'enjeux communautaires.