

Humanités Numériques

Cours 2 : Documents Numériques

Loïc Grobol

2021-09-21

Documents

Le *document* est une notion centrale en humanités Numériques

- Le champ voit le jour autour de la numérisation des documents
- Il concerne une large part de ses problématiques
 - Composition
 - Archivage
 - Traitement et exploitation
 - Diffusion

Mais c'est quoi un document ?

D'après vous ?

- → 10 minutes en collectif : arriver à une définition ou des définitions de *document*
- Vous pouvez commencer par lister des exemples de documents, de cas typiques, de cas limites et de choses qui sont proches de documents mais n'en seraient pas

Première tentative

Du latin *documentum* (« modèle, exemple »), dérivé de *docere* (« montrer, faire voir, instruire »)

1. Écrit qui sert de preuve ou de renseignement. (8ème dictionnaire de l'Académie)
2. (Par extension) Source de renseignement ou de preuve, quel que soit le support matériel (film, photographie, disquette, etc.)
(Wiktionnaire : « document »)

Est-ce satisfaisant ?

Texte et document

- Plusieurs définitions variables selon les époques et selon les disciplines et les domaines : en droit, en histoire, en informatique, en documentation...
- Historiquement la conjonction de deux éléments : un **support** matériel et un **contenu** informationnel.
- Définition actualisée : un « ensemble cohérent, stable et fini d'informations structurées et lisibles, à usage défini, quel qu'en soit le support » (Le Coadic et al., 2008)

Le document numérique

- Une forme de représentation de l'information consultable à l'écran d'un appareil électronique.
- Statut de l'affichage : document ou interface ?

...

- À l'inverse du document sur papier, il y a séparation de la présentation (mise en page, affichage) et de l'information (composition de texte, données).

Pour J-M Salaün et le RTP-doc

Un document est une **trace** permettant d'**interpréter** un **événement passé** à partir d'un **contrat de lecture**. (Salaün, 2012)

Cette définition issue de « Vu, lu, su. Les architectes de l'information face au monopole du web » (Salaün, 2012) s'appuie sur les travaux du réseau RTP-doc (Pédauque, 2003 ; Pédauque, 2004 ; Pédauque, 2006).

Elle repose sur l'idée qu'un document peut se voir suivant **trois** perspectives.

Vu : la forme

Le **support** du document :

- Un *codex* : le livre relié
- Une feuille ou une liasse
- Un journal, un magazine

Voir un document audiovisuel

- Pour un document audiovisuel (son, vidéo...), le support n'est pas directement utilisable, mais passe par une interprétation.
- Un document audiovisuel doit être reconstruit par une machine.
- Implicitement : la reproduction est censée être aussi fidèle que possible à l'original.

Voir un document numérique

Lien plus complexe entre ce qui est stocké et ce qui est perçu : à partir d'un même fichier, on peut générer plusieurs vues

Par exemple ces slides :

- Source
- Pdf

...

Un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs (Pédauque, 2003)

Lu : le contenu

Ce que dit le document, tel qu'il est interprété par ses destinataires

- Informations
- Contexte
- Représentations mentales

Contexte et construction du sens

L'auteurice du document doit fournir suffisamment de contexte pour son interprétation et ses destinataires intégrer ce contexte pour l'interpréter de la façon prévue.

On ne reçoit pas un document administratif comme un ouvrage documentaire ou une œuvre de fiction.

...

Le numérique apporte une nouvelle facette : l'interactivité. En faisant varier les vues (support) sur le document, on peut accéder à différentes interprétations.

Su : la médiation

La production et l'accès à un document est un acte *social*, qui porte une relation entre les créateurices d'un document et ses destinataires ainsi que les éventuels intermédiaires.

Cette mise en relation permet la transposition d'évènements passés (connaissance, réflexion imagination, émotion...) sur le présent (les états mentaux des destinataires).

Éléments supplémentaires

Cette définition (voir la source pour plus de détails) a l'intérêt d'être très générale et applicable à une grande variété de documents.

Elle peut en revanche être imprécise sur certains points qui recourent souvent à la fois le vu, le lu et le su.

Structuration des documents

Lien forme-contenu qui conditionne implicitement la transmission pourvu que les conventions soient intégrées.

- La structure *physique* : apparence visuelle (colonnes, justification, mise en page mais aussi espaces (!), ponctuation...)
- La structure *logique* : liée au contenu intellectuel (sections, chapitres... avec titres, thèmes, mots-clés)
- La structure *générique* : liée au genre, à la convention de communication. Ainsi un article contient toujours une introduction, des sections suivant un plan plus ou moins stéréotypé, une conclusion...

Métadonnées

Des documents qui décrivent (et éventuellement identifient) des documents :

- Auteurices
- Date

- Publication
- Genre
- ...

→ Mêmes impératifs que les documents qu'elles décrivent.

Conditionnent la transmission, l'archivage et la réception.

Formats

Comment est représenté le document en pratique :

- Standard —interopérable : par qui et comment le document est-il lisible ? Est-il portable ? Est-ce pérenne ?
 - Standardisations officielles (ISO, Afnor, W3C, Consortium Unicode)
 - Ou de facto (doc, gif), qui peut devenir officiel (JPEG, PDF)
- Ouvert : les spécifications sont-elles publiées intégralement et librement utilisables ? Sont elles librement utilisables ? Gratuitement ?

Pas uniquement le stockage :

Connaissez-vous ces formats : pdf, OpenDocument, HTML, XML, TIFF, PNG, JPEG, gzip, bzip, xz, tar, zip, 7z ?

Pas uniquement le stockage :

- Conserver la présentation : PDF, OpenDocument
- Conserver la structure : HTML, XML
- Conserver une image
 - En intégralité : TIFF, PNG
 - En limitant la taille au prix de l'information (compression destructive) : JPEG
- Réduire la taille (en général) : gz, bzip, xz
- Agréger : tar, zip, 7z
- Encodage vs. conteneur : AV1 vs. Matroska

Bibliographie

Yves-François Le Coadic, Paul-Dominique Pomard, et Éric Sutter. 2008. *Dictionnaire encyclopédique de l'information et de la documentation*. Armand Colin, Paris, 3 éditions, mars.

Roger T. Pédaque. 2003. Document : forme, signe et médium, les re-formulations du numérique., juillet.

Roger T. Pédaque. 2004. Document et texte. Permanence et transformations. *Document en discussion au sein du RTP-DOC*, juin.

Roger T. Pédaque. 2006. Document et modernités., mars.

Jean-Michel Salaiün. 2012. *Vu, lu, su. Les architectes de l'information face au monopole du web*. La Découverte, éditions.