Le tournant numérique des sciences humaines et sociales

olivier.baude@huma-num.fr

Un tournant ? Des tournants ?

- Qu’est-ce que le « tournant numérique » des SHS ?
- Y-a-t-il un ou plusieurs tournants ?
- Le tournant numérique des SHS ou la naissance des humanités numériques ?
- Quel rôle pour les données de la recherche ?

MANIFESTE DES DIGITAL HUMANITIES (2010)

Contexte

Nous, acteurs ou observateurs des digital humanities (humanités numériques) nous nous sommes réunis à Paris lors du ThatCamp des 18 et 19 mai 2010. Au cours de ces deux journées, nous avons discuté, échangé, réfléchi ensemble à ce que sont les digital humanities et tenté d'imaginer et d'inventer ce qu'elles pourraient devenir. À l'issue de ces deux jours qui ne sont qu'une étape, nous proposons aux communautés de recherche et à tous ceux qui participent à la création, à l'édition, à la valorisation ou à la conservation des savoirs un manifeste des digital humanities.

I. Définition

1. Le tournant numérique pris par la société modifie et interroge les conditions de production et de diffusion des savoirs.

3. Les digital humanities désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales.

IV. Orientations

9. Nous lançons un appel pour l'accès libre aux données et aux métadonnées. Celles-ci doivent être documentées et interopérables, autant techniquement que conceptuellement.

10. Nous sommes favorables à la diffusion, à la circulation et au libre enrichissement des méthodes, du code, des formats et des résultats de la recherche.

IV. Orientations

14. Nous appelons à la construction de cyberinfrastructures évolutives répondant à des besoins réels.

Les SHS et la production et diffusion des savoirs

Données de la recherche :
« ensemble d'informations factuelles enregistrées sur des supports, produites ou collectées, selon divers procédés au cours d'un processus de recherche" *(N. Reymonet, M. Moysan, A. Cartier, R. Délémontez. Réaliser un plan de gestion de données « FAIR » : modèle , 2018.)*

Publications scientifiques :
La notion de **littérature scientifique** désigne plus largement l'ensemble des publications scientifiques. Les publications scientifiques permettent de diffuser des informations scientifiques et techniques, produites et utilisées notamment par les chercheurs.

1. Le document :
Un document est une pièce informative et/ou faisant office de preuve, enregistré de manière persistante, quel que soit le support matériel (film, photographie, disquette, etc.)

2. La collection de documents : corpus, archives…

3. Les bibliothèques (et archives) pour conserver et rendre accessibles les documents

Le tournant numérique des bibliothèques

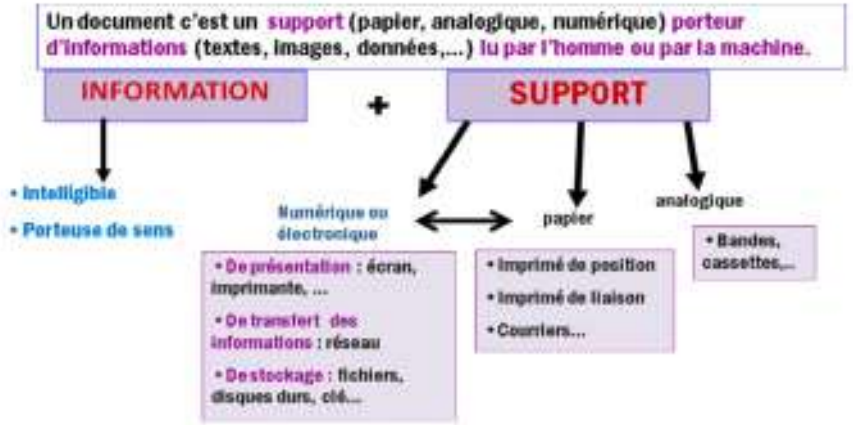
(2300 ans d'histoire en 45 secondes)

Le tournant numérique des bibliothèques

Les missions fondamentales traditionnelles des bibliothèques se transposent dans le champ numérique :

- constituer une collection
  - en numérisant les collections physiques
  - en acquérant des contenus numériques auprès des éditeurs
  - en signalant des ressources externes
- communiquer les documents au(x) public(s)
  - en développant des moteurs et des interfaces de recherche efficaces
  - en développant (ou en réutilisant) des outils de visualisation confortables
- conserver les documents
- valoriser et promouvoir la collection et les services qui l'accompagnent

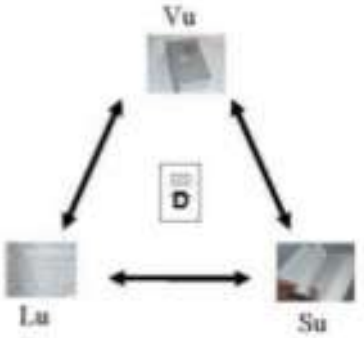
Le tournant numérique du « document »

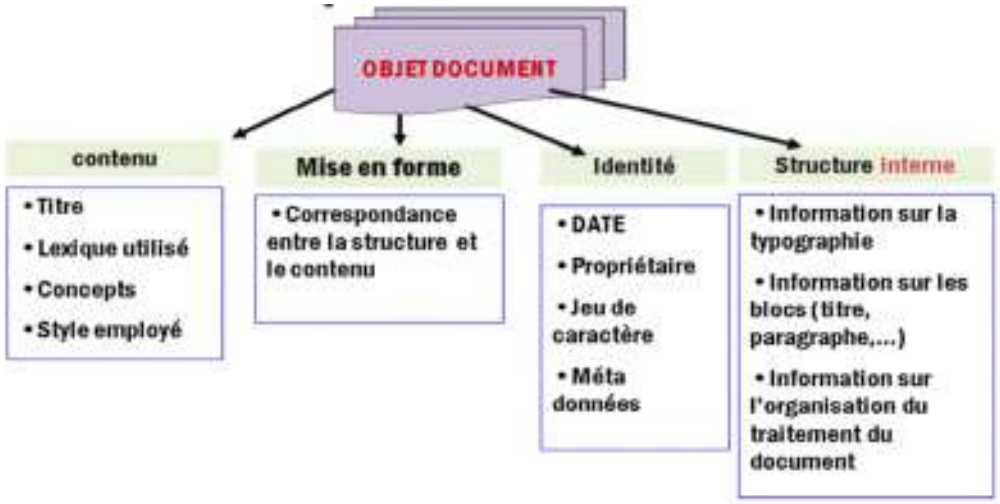


Le document numérique

Un **document numérique** est une forme de représentation de l'information consultable à l'écran d'un appareil **électronique**.

À l'inverse du document sur papier, qu'il soit manuscrit ou imprimé, le document numérique permet de séparer la présentation (les techniques de mise en page) de l'information (composition de texte, données).





Les deux principes fondamentaux des documents structurés

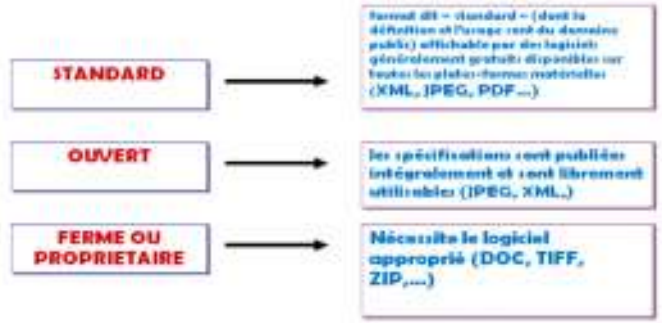
L'approche des documents structurés est basée sur deux principes:

1) Un document numérique est un fichier texte (en format texte, e.g., ASCII), auquel on superpose des conventions additionnelles qui permettent de représenter une structure hiérarchique (en arbre).

2) La structure hiérarchique du document numérique doit correspondre le mieux et le plus explicitement possible à la nature et à la structure de "l'information" qui doit être véhiculée par le document.

On appelle "document structuré" tout document numérique (donc, concrètement, un fichier, ou une séquence d'octets stockable sur un support quelconque) créé dans le respect de ces deux principes.

Formats



Le tournant informatique

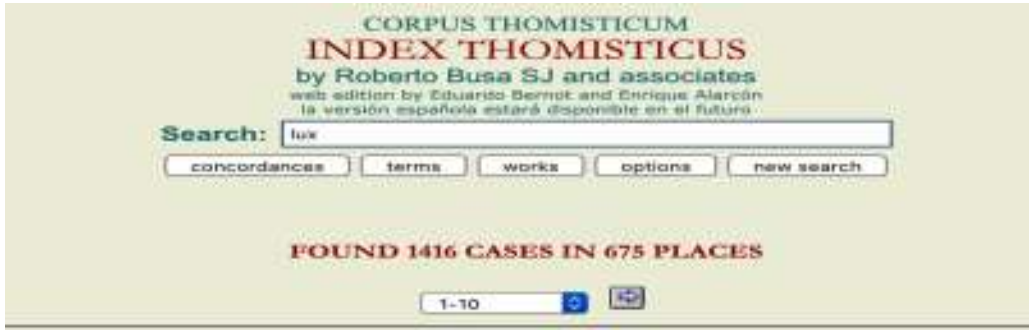
Roberto Busa **1946 : L'*Index Thomisticum***

Intérêt de l'informatique pour le projet d'Index des 118 textes de Thomas d'Aquin.

Rencontre avec T.J. Watson, fondateur d'IBM.

Taper sur des cartes perforées tout l'ouvrage pour générer automatiquement un index de chaque mot du corpus.

Projet de 30 ans, 56 volumes.



la concordance automatique

> « *Le lecteur ne doit pas simplement attacher ses propres significations aux mots qu'il lit, mais il doit aussi rechercher les significations que les mots ont pour l'auteur lui-même. Par ailleurs tous les mots fonctionnels et grammaticaux (qui dans mon esprit ne sont pas vides du tout mais au contraire très riche sur le plan philosophique) manifestent la logique la plus profonde de l'être qui génère les structures de base du discours humain* ». [Busa,1980]
>
> **Conduire l'analyse à un niveau jusque là inaccessible. Il ne s'agit plus d'interpréter le texte mais de l'observer comme objet.**

Le tournant informatique

- Numériser, compter :
  - Les années 1970 Lexicométrie et textométrie

  Exemple d'analyse d'un discours d'un homme politique (fréquence et distances sémantiques) :

- Les grands corpus 1970 - 1980
  - Le terme de *corpus linguistics*, la linguistique des corpus, est apparu dans les années 80, bien après la création des premiers corpus informatisés, pour désigner une tendance nouvelle en linguistique, l'étude de grandes quantités de textes par le biais d'outils informatiques.
  - 1980 : British National Corpus est un corpus de texte de 100 millions de mots d'échantillons d'anglais écrit et parlé provenant d'un large éventail de sources.
    - Un échantillon d'un million de mots issus de différents textes publiés en 1961 dans différents domaines aux Etats-Unis. Ce corpus peut ensuite être mobilisé comme mètre-étalon pour analyser des œuvres ou des textes particuliers encmesurant la manière dont ils se différencient du corpus de référence.
  - FRANTEXT : 4250 textes littéraires et philosophiques, mais aussi scientifiques et techniques numérisés (19e-20e s.).
  - ESLO : Grand corpus oral numérique (1970- 2010- XXXX) par le LLL

- *Explorer l'information : le document structuré et l'exploration outillée*
- *Explorer des gisements d'informations : les corpus, collections de données et archives*

Le tournant informatique

Outiller l'exploration

Traiter un texte c'est plus que le lire, plus que l'annoter, plus que l'associer avec d'autres textes… Il s'agit d'exposer sa structuration afin de permettre un monde distribué, où *« les livres dans la bibliothèques savent se parler entre eux ».*

La Text Encoding Initiative (abrégé en TEI, en français « initiative pour l'encodage du texte ») est un format de balisage et une communauté académique internationale dans le champ des humanités numériques visant à définir des recommandations pour l'encodage de ressources numériques, et plus particulièrement de documents textuels.

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

Outiller les bibliothèques

La normalisation des règles de catalogage garantit l'homogénéité des catalogues et permet l'échange des notices.

Principe

premier

Confére

ratifie la

normes

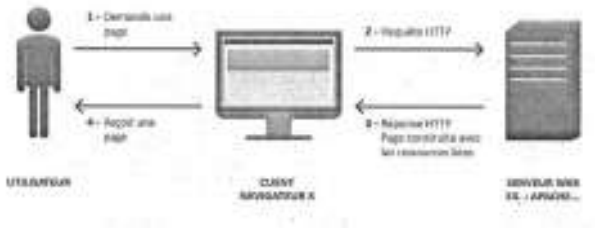
Le tournant informatique et le Web

- **Internet** est le réseau informatique mondial accessible au public.
- **World Wide Web :** un système hypertexte public fonctionnant sur Internet. Le Web permet de consulter, avec un navigateur, des pages accessibles sur des sites.
- **Les 3 piliers du web :**

  Le langage HTML (HyperText Markup Language) fournit la structure d'une page web. HTML permet d'écrire des documents hypertextuels liant les différentes ressources d'Internet avec des hyperliens

  Les URL (Uniform Resource Locators) constituent des adresses pour les ressources présentes sur le web.

  Le protocole HTTP (Hyper-Text Transfer Protocol) règle les échanges entre le programme client qu'est un navigateur et le programme serveur dont il attend des ressources.



- Web 1.0 web statique (Internet des pages)
- Web 2.0 web collaboratif (Internet des personnes)
- Web 3.0 web sémantique (Internet des données)

Le tournant des humanité numériques

Les « Humanités numériques » :

Journal officiel du 09/07/2019

Définition : *Domaine de recherche et d'enseignement au croisement de l'informatique et des lettres, des arts, des sciences humaines et des sciences sociales, visant à produire et à partager des savoirs, des méthodes et de nouveaux objets de connaissance à partir d'un corpus de données numériques.*

Équivalent étranger : digital humanities (en)

Existe-t-il un tournant des cadres théoriques des sciences humaines et sociales ?

Le tournant des Humanités numériques : 1. La textualité rayonnante

The William Blake Archive

Une archive numérique :

œuvres numérisées
notices, publications, …

« Le travail de numérisation et de mise en ligne des oeuvres de William Blake fait figure de pionnier dans ce domaine.

La William Blake Archive est considérée aujourd'hui comme une sorte de modèle pour les projets de numérisation de corpus en humanités numériques, en raison de la qualité du travail d'édition dont il rend compte, aussi bien en termes de choix des matériaux qu'en ce qui concerne l'encodage du corpus (les textes ont été encodés très tôt en SGML puis en XML) ou la conception des interfaces de navigation et de consultation »

MOUNIER, Pierre. *Ce que l'ordinateur apporte aux humanités* In : *Les humanités numériques : Une histoire critique* [en ligne]. Paris : Éditions de la Maison des sciences de l'homme, 2018 (généré le 15 novembre 2020). Disponible sur Internet : <http://books.openedition.org/editionsmsh/12036>. ISBN : 9782735124831. DOI : https://doi.org/10.4000/books.editionsmsh.12036.

Le tournant des Humanités numériques : 1. La textualité rayonnante

Constituer des archives de collection de documents numériques apporte de nouvelles méthodologies et de nouveaux cadres théoriques :

Jerome McGann et la textualité radiante

« Le caractère ouvert et non clos de l'archive numérique bouleverse les pratiques de monographie et d'éditions savantes, elles mêmes transformées par les mediums »

« Les technologies numériques qu'utilisent les chercheurs en sciences humaines se sont focalisées quasiment exclusivement sur les méthodes de tri, d'accès et de diffusion de larges corpus de sources et sur certains problèmes spécialisés de stylistique et de linguistique computationnelle.

De ce point de vue, le travail traite rarement des questions d'interprétation et de réflexivité qui sont les problèmes de premier plan qui occupent la plupart des chercheurs et enseignants des humanités. Les technologies numériques sont restées au stade de l'instrument en étant au service des tâches techniques et pré-critiques des bibliothécaires, des archivistes et des éditeurs scientifiques.

Mais le champ de recherche et d'enseignement dans les humanités en général ne prendra pas au sérieux l'usage des technologies numériques tant que ne sera pas faite la démonstration que ces outils améliorent la manière dont nous explorons et expliquons les œuvres esthétiques, tant qu'elles n'étendront pas le champ d'application de nos procédures interprétatives»

L'archive permet ainsi d'un côté de rassembler des matériaux hétérogènes autour de l'œuvre elle-même, rendant compte non seulement de l'œuvre originelle (il ne s'agit pas seulement de donner accès à une transcription qui n'est que le texte dématérialisé, mais d'abord au fac-similé de l'édition princeps qui en restitue autant que possible la matérialité), ***mais aussi de l'intertextualité qui la lie à d'autres œuvres, textuelles, picturales, matérielles – autrement dit, sa « radiance » même.***

MOUNIER, Pierre. *Ce que l'ordinateur apporte aux humanités* In : *Les humanités numériques : Une histoire critique* [en ligne]. Paris : Éditions de la Maison des sciences de l'homme, 2018 (généré le 15 novembre 2020). Disponible sur Internet : <http://books.openedition.org/editionsmsh/12036>. ISBN : 9782735124831. DOI : https://doi.org/10.4000/books.editionsmsh.12036.

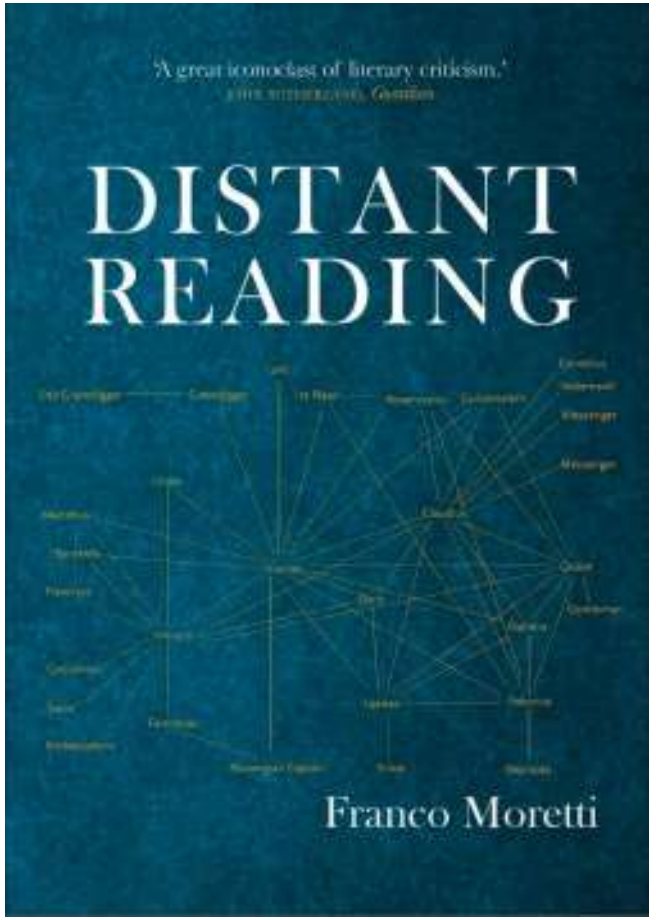
Le tournant des Humanités numériques : 1. La textualité rayonnante

La création des Centres des humanités numériques : 3 caractéristiques

- Des centres pour de la recherche collective :
  - Chercheurs en sciences humaines et sociales et en littérature
  - Informaticiens
  - Bibliothécaires et documentalistes
- L'interdisciplinarité
- Une critique des humanités classiques

Le tournant des Humanités numériques : 2. La lecture à distance *(Distant reading)*

- **L'étude littéraire sur de vastes corpus nécessite « la montée en généralité » le repérage de « patterns »**
- De nouvelles méthodes pour bouleverser les études littéraires (opposition close reading et distant reading)
- « Je dirais simplement que nous avons beaucoup à apprendre des méthodes des sciences sociales et des sciences naturelles » [Moretti, 2003]



Le distant reading est un concept proposé par Franco Moretti pour décrire une approche en analyse littéraire. Littéralement « lecture à distance », le distant reading s'oppose à la « lecture de près », le close reading, la méthode classique utilisée en analyse littéraire.

En close reading, l'attention est portée sur l'œuvre qui est lue et analysée en détails.

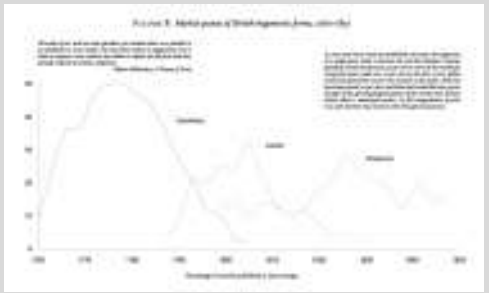
Moretti argumente que le close reading est incapable de saisir toute l'envergure de la littérature, car même si une telle approche est possible pour deux cents livres d'un siècle donné, que savons-nous réellement sur la littérature du siècle en question si celle-ci se compte en centaines de milliers d'œuvres ?

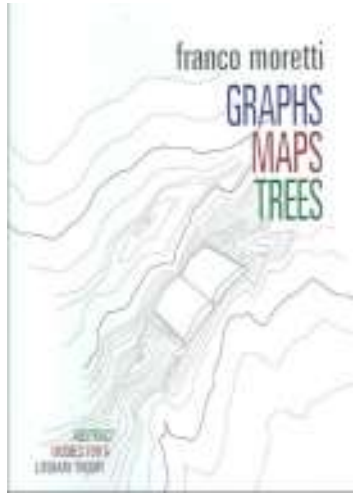
MOUNIER, Pierre, *Ce que l'ordinateur apporte aux humanités* In : *Les humanités numériques : Une histoire critique* [en ligne]. Paris : Éditions de la Maison des sciences de l'homme, 2018 (généré le 15 novembre 2020). Disponible sur Internet : <http://books.openedition.org/editionsmsh/12036>. ISBN : 9782735124831. DOI : https://doi.org/10.4000/books.editionsmsh.12036.

Le tournant des Humanités numériques : 2. La lecture à distance *(Distant reading)*

Franco Moretti : la lecture à distance

- Graphs, Maps, Trees : modèles abstraits pour l'histoire littéraire
  - Graphiques pour analyser les évolutions
  - Cartes pour la localisation et les relations
  - Arbres pour la structure

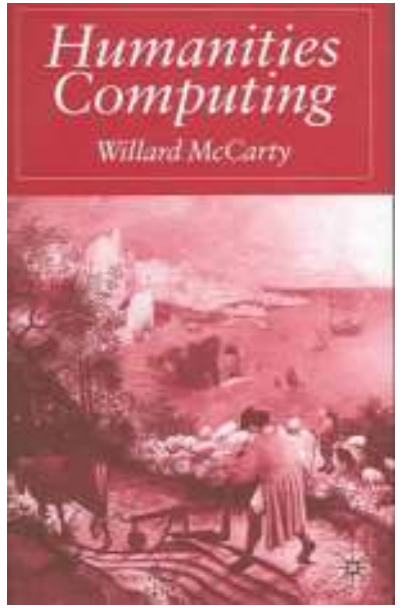




Le tournant des Humanités numériques : 3. Formaliser la connaissance

Williard McCarty : L'informatique et les humanités

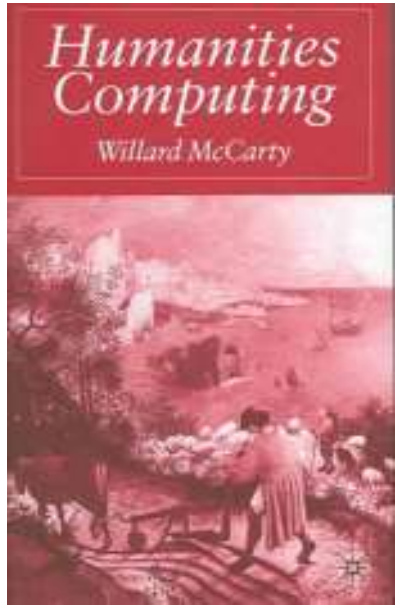
- Intérêt de la contrainte imposée au chercheur par le traitement numérique
- Pour McCarty, ce n'est pas la puissance de calcul de l'ordinateur qui est véritablement intéressante ; c'est plutôt la contrainte imposée par le traitement numérique de l'information au chercheur qui le conduit à devoir formaliser tout à la fois le matériau sur lequel il travaille (ce qui le transforme en données) et le raisonnement qu'il mobilise pour traiter ces données.
- Cela nécessite de formaliser la matériau (le corpus en data) mais aussi de formaliser le raisonnement pour traiter ces données



Le tournant des Humanités numériques : 3. Formaliser la connaissance

Williard McCarty : L'informatique et les humanités

- La modélisation ne permet pas de produire des connaissances nouvelles mais aide le chercheur à se poser de nouvelles questions (notamment ce qui ne fonctionne pas avec le modèle).
- « L'ordinateur est un télescope pour l'esprit. C'est l'instrument qui permet d'observer d'une autre manière et plus loin ».
- Démarche expérimentale (non pour valider des hypothèses mais pour générer de nouvelles questions par des résultats négatifs).

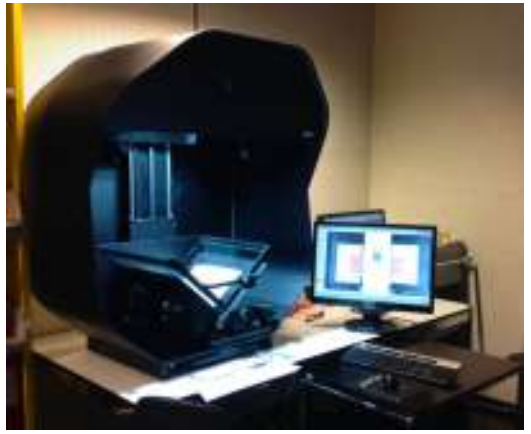


Le tournant des Humanités numériques :

Le tournant des humanités numériques est principalement celui de la place octroyée aux données de la recherche.

La place des « données de la recherche » comme *pratique* et comme *production* scientifique

Définition : « *Enregistrements factuels (chiffres, textes, images, sons) utilisés comme source principale pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche* ». OCDE 2007



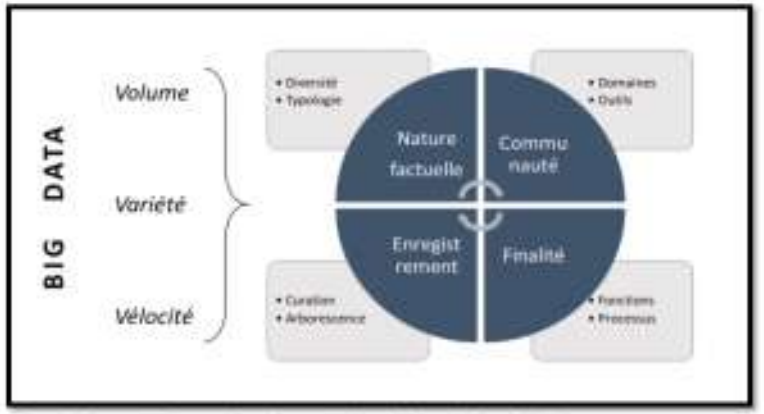


Figure 1. Les quatre dimensions du concept des données de la recherche

Joachim Schöpfel

HAL Id: hal-01979798https://hal.archives-ouvertes.fr/hal-01979798

La place des « données de la recherche » comme *pratique* et comme *production* scientifique

Gérer le cycle de vie des données de la recherche

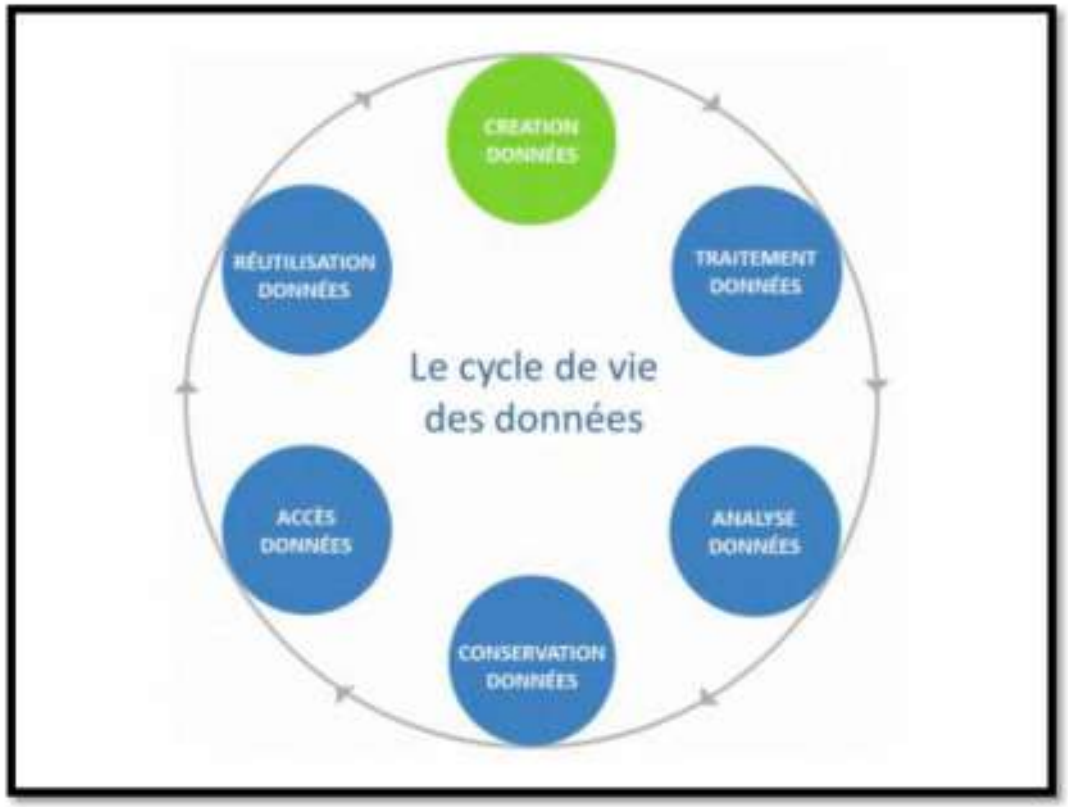


Figure 2. Cycle de vie des données (source : inist.fr)

La place des « données de la recherche » comme *pratique* et comme *production* scientifique

Un cadre juridique

Article 30 de la LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique

1. Publications

L'article 30 donne un nouveau droit aux auteurs, celui de publier en accès ouvert leurs articles après un embargo de 6 à 12 mois, pour des recherches financées au moins à 50% sur fonds public.

Il crée donc un droit.

2. Données

L'article 6 définit le principe d'ouverture par défaut pour les données administratives : données construites dans le cadre des missions de service public, auxquelles les données de recherche appartiennent.

Il crée donc une obligation

La place des « données de la recherche » comme *pratique* et comme *production* scientifique

Des pratiques scientifiques

Plan S (Europe) : Le plan S acte l'obligation que les publications de recherches scientifiques, financées par des subventions publiques obtenus des organismes de financement nationaux et européens, soient publiées en libre accès dans des revues ou sur des plateformes à partir du 1er janvier 2020. Il est adossé à une coalition de financeurs de la recherche, notamment l'ANR en France.

- **Le mouvement pour la science ouverte**

- Les chercheurs emploient aujourd'hui les outils numériques dans toutes les facettes de leur activité. Ils innovent, développent de nouveaux outils, créent et gèrent des plateformes numériques ainsi que de précieuses données. Leurs publications sont nativement numériques.
- Cependant, ces productions scientifiques numériques de la recherche sont souvent difficilement accessibles : ressources non-publiques ou non-documentées, nécessité d'être abonné ou d'avoir acheté la ressource, nécessité de s'identifier de manière variable selon la ressource, formats incompatibles voire incohérents etc. L'utilisation pleinement efficace du numérique est entravée. La science ouverte vise à corriger cette insuffisance.
- La science ouverte favorise une approche transversale de l'accès au travail scientifique, des visées et du partage des résultats de la science mais aussi une nouvelle façon de faire de la science, en ouvrant les processus, les codes, les méthodes ou encore les protocoles. L'accès ouvert rend possible le mouvement de diffusion de la production scientifique sur internet. L'accès ouvert lève les barrières de l'accès en maintenant toutes les protections du droit d'auteur.
- Aussi ouvert que possible, aussi fermé que nécessaire (Rapport commission européenne 2016)

https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaquette_Science-Ouverte_18112019.pdf

https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Plan_S.pdf

https://www.science-ouverte.cnrs.fr/le-mouvement-pour-la-science-ouverte/

https://www.cnrs.fr/sites/default/files/pdf/Plaquette_PlanDDOR_Nov20.pdf

La place des « données de la recherche » comme *pratique* et comme *production* scientifique

« La science ouverte est la diffusion sans entrave des publications et des données de la recherche »

La France s'engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans payement.

Axe 1 Généraliser l'accès ouvert aux publications scientifiques

L'ouverture des publications scientifiques doit devenir la pratique par défaut aussi vite que possible. Pour engager cette dynamique, les publications issues de recherches financées au moyen d'appels à projets sur fonds publics seront obligatoirement mises à disposition en accès ouvert, que ce soit par la publication dans des revues ou ouvrages nativement en accès ouvert, soit par dépôt dans une archive ouverte publique comme HAL.

Axe 2 Structurer et ouvrir les données de recherche

Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes.

Les données « FAIR »

Constat :

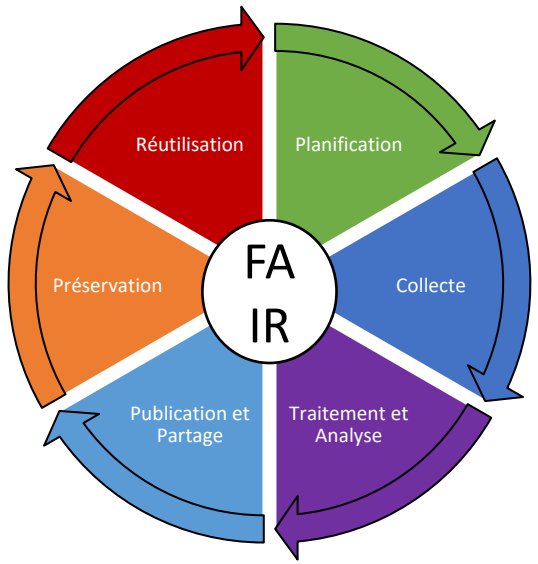
- Le taux de réutilisation des données (scientifiques) est extrêmement faible ;
- Objectifs :

- Améliorer les infrastructures pour permettre aux machines de retrouver les données, et aux chercheurs de les réutiliser ;

- Promouvoir des principes concis et mesurables – bonnes pratiques pour la gestion des données.

Chacun des 13 principes se décline en un ensemble de caractéristiques que doivent présenter les données et les métadonnées pour faciliter leur découverte et leur utilisation par les hommes mais aussi par les machines

- Trois types d'entités:
  - Données (ou tout objet numérique),
  - Métadonnées (informations sur l'objet numérique),
  - L'infrastructure de support (ex. moteur de recherche, plateforme de stockage, etc.).



Un outil pour la science ouverte. Les données au cœur d'un process scientifique

O. Rouchon 2020 (CINES)

Les données « FAIR »

Faciliter la découverte des données par les humains et les systèmes informatiques et requiert une description et une indexation des données et des métadonnées.

- **F1** Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne.
- **F2** Les métadonnées décrivant les données sont riches.
- **F3** Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher.
- **F4** Les métadonnées spécifient l'identifiant de la donnée.

Stocker durablement les données et les métadonnées et à faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint) et d'utilisation (licence).

- **A1** Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé.
  - **A1.1** Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle.
  - **A1.2** Le protocole utilisé permet l'authentification et l'autorisation si besoin.
- **A2** Les métadonnées sont accessibles même quand les données ne le sont plus.

Intégrer les données téléchargeables, utilisables, intelligibles, et combinables avec d'autres données, par des humains et des machines.

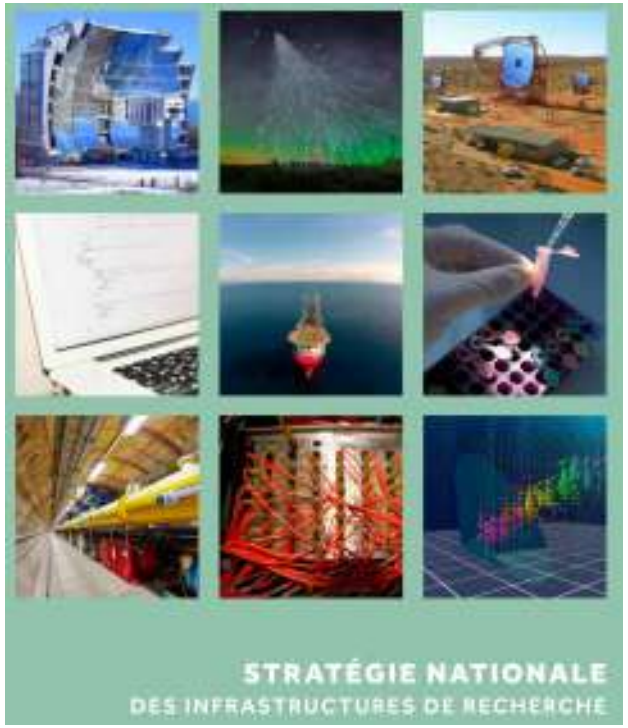
- **I1** Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.
- **I2** Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR.
- **I3** Les données et les métadonnées incluent des liens vers d'autres (méta)données.

Rendre les données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction et/ou transparence de la science).

- **R1** Les données et les métadonnées ont des attributs multiples et pertinents.
  - **R1.1** Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible.
  - **R1.2** Les données et les métadonnées sont associées à leur provenance.
  - **R1.3** Les données et les métadonnées correspondent aux standards des communautés indiquées.

O. Rouchon 2020 (CINES)

Le tournant des Infrastructures de recherche



« les IR doivent s'inscrire dans une dynamique de science ouverte et de maîtrise du cycle de vie des données qu'elle produit et utilise »

Le tournant des Infrastructures de recherche

OpenEdition

OpenEdition est un portail de ressources électroniques en sciences humaines et sociales.

OpenEdition Journals

OpenEdition Books

hypotheses

calenda

Le tournant des Infrastructures de recherche

- **Archive ouverte HAL**
- L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion d'articles scientifiques de niveau recherche, publiés ou non, et de thèses, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le tournant des Infrastructures de recherche

Persée permet la consultation et l'exploitation libres et gratuites de collections complètes de publications scientifiques. Créé par le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, il a été mis en ligne en 20051.

Aujourd'hui, Persée est un service d'appui à la recherche dont les missions sont la valorisation numérique du patrimoine scientifique, la recherche et l'innovation en matière d'outils et de méthodes pour développer des corpus numériques, la diffusion de compétences et d'expertise dans ce domaine.

Le portail permet un accès ouvert à la connaissance : accès libre aux contenus, développement open source, standards ouverts, interopérabilité et exploitation transversale des données.

https://www.persee.fr/

https://fr.wikipedia.org/wiki/Pers%C3%A9e_(portail)

Le tournant des Infrastructures de recherche

La TGIR Huma-Num

Les missions

Accompagner l'évolution des communautés SHS dans le contexte du numérique et de la science ouverte.

1. Mettre en œuvre une infrastructure pour les données de la recherche SHS (dans la logique des *FAIR principles : Findable, Accessible, Interoperable, and Reusable*) et en tenant compte des pratiques des SHS.

1. Participer à la construction des infrastructures internationales, en accompagnant les communautés SHS nationales, dans le contexte de l'European Open Science Cloud.

Huma-Num

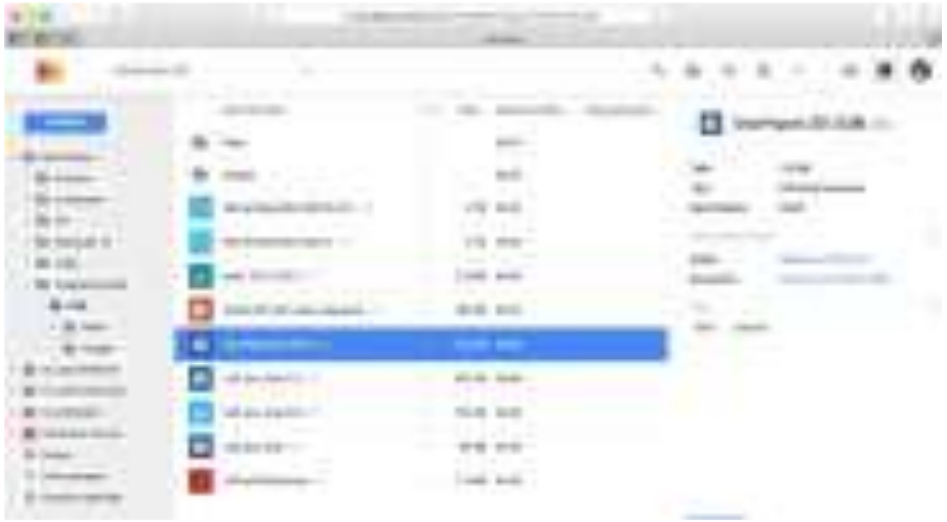
Des services pour les données en SHS

Des services classiques …

ShareDocs

stockage « outillé » pour les SHS

- Stockage des données en mode « dropbox » (mais sur les serveurs d'Huma-Num, en France)
- Partage, édition, traitement des données
- Un espace de travail collaboratif, sécurisé.



Stockage et hébergement de données, de sites, d'applications,…

Des services classiques …

Infrastructure physique

- Cluster Web, VM
- Puissance de calcul (CPU, GPU,…)

Accompagner les traitements des données :

- OCR-HCR
- Licences (Oxygen, Maxqda, Sphinx, Qgis, Arcgis, Photoscan,…)
- Serveur R avec environnement logiciel R Studio Pro
- Serveur FileMaker Pro
- Gitlab
- …

…aux services pour les données à forte valeur ajoutée pour préserver, signaler publier

nakala

isidore

Préservation à long terme

…aux services pour les données à forte valeur ajoutée :

L'entrepôt

NAKALA permet à des équipes de recherche de déposer tous types de données numériques (fichiers textes, sons, images, vidéos, objets 3D, etc.) documentées dans un entrepôt sécurisé afin de les partager.

Principes clés

- Donnée publique (publication) / Donnée privée (dépôt)
- Dépôt simple / Dépôt par lot | Donnée unitaire / Donnée multiple
  - chaque objet est identifié par un DOI : collection, donnée, version, fichier
  - embargo sur les fichiers
  - métadonnées obligatoires + facultatives (modèle étendable)
- Regroupement de n'importe quelles données au sein de Collections
- Partage de droits sur les données et/ou collections
- Consultation, API : REST (in/out), SPARQL, IIIF et OAI (out), Intégration (OEmbed)

…aux services pour les données à forte valeur ajoutée :

ISIDORE pour signaler, accéder et enrichir les données

ISIDORE est un moteur de recherche permettant d'accéder en ligne aux documents et aux données numériques en sciences humaines et sociales.

- Publications
- Événements
- Corpus, jeux de données
- Fonds d'archive,…

+ de 6 millions de données signalées et reliées entre elles avec des référentiels scientifiques

+ d'1,3 million d'utilisateurs/an

+ de 7000 « sources » de données collectées dans le monde

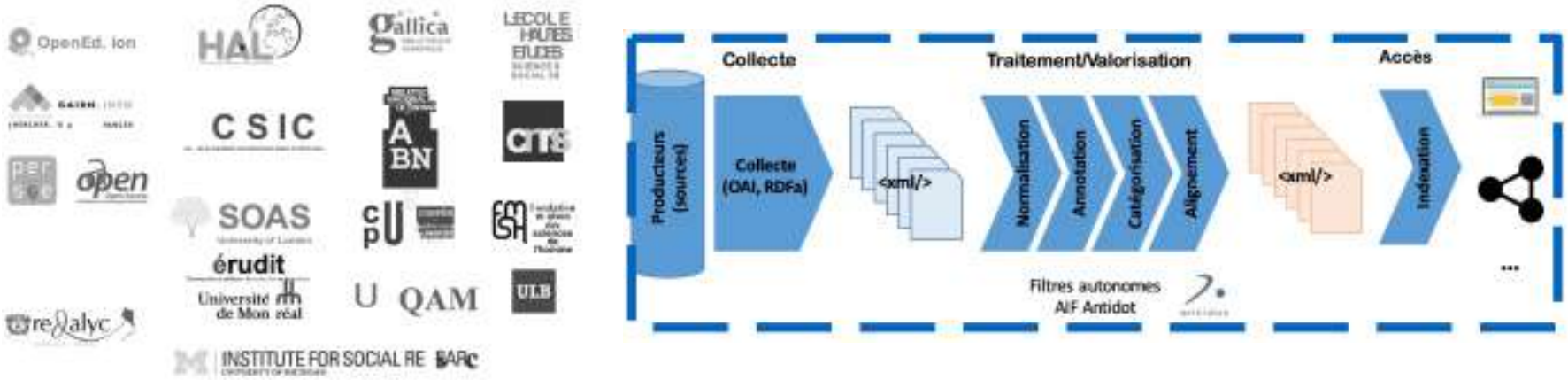
isidore

Un assistant de recherche fondé sur les technologies du web sémantique

Collecter
Traiter

Enrichir
Classer
Catégoriser

Accéder



Un service de préservation à long terme des données

THE FRENCH INFRASTRUCTURES SUPPORTING OPEN SCHOLARLY COMMUNICATION

Huma-Num

per se

métopes

CCSD

OpenEdition

La MSH Val de Loire

L'ATELIER NUMÉRIQUE

- Communication
- Information géographique
- Web
- Audiovisuel
- Dispositif photographique mobile
- Drone
- Huma-Num Box/Stockage
- VisualGraphs
- Serveur de traitement de données

Conclusion

- Qu’est-ce que le « tournant numérique » des SHS ?
- Y-a-t-il un ou plusieurs tournants ?
- Le tournant numérique des SHS ou la naissance des humanités numériques ?
- Quel rôle pour les données de la recherche ?