

Glossaire

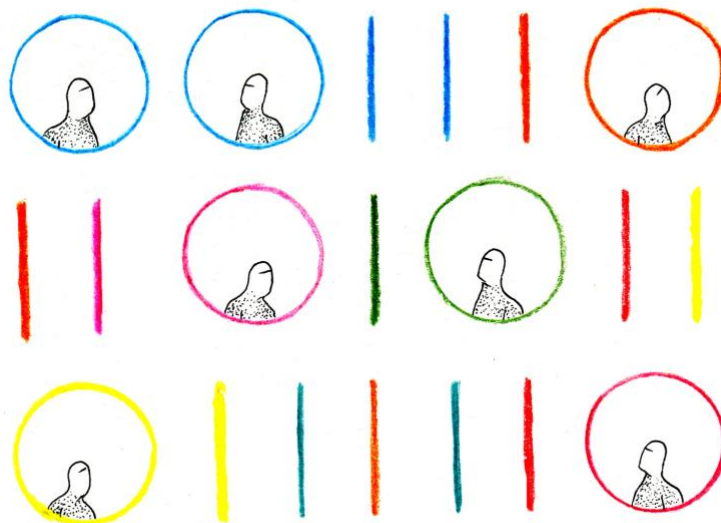
Introduction aux humanités numériques

Version du 13/12/2019

Ce glossaire explicite les termes scientifiques et techniques utilisés dans les projets relevant des humanités numériques. Les définitions proposées sont volontairement très généralistes pour permettre une première approche du terme. Des ressources internet sont proposées en fin de glossaire pour aller plus loin.

Ce document a été conçu au départ des glossaires de la plateforme d'édition de manuscrits et d'archives EMAN et du projet Foucault fiches de lectures. Des entrées ont été fusionnées, simplifiées ou complétées, d'autres ajoutées par Marie-Laure Massot, coordinatrice du projet, pour aboutir à une première version initialement préparée dans le cadre du [Parcours Humanités numériques de l'École normale supérieure](#) en 2018-2019. Ce glossaire a vocation à évoluer au fil du temps, s'enrichissant des contributions des personnes souhaitant s'associer au projet. Ce projet bénéficie de l'expertise des membres du [Groupe de travail humanités numériques](#) de l'[EUR Translitterae](#).

Une version interactive sera proposée à partir de janvier 2020 sur le site de l'[Initiative Digit_Hum](#) dans la rubrique [Ressources](#).



Index

| | | | |
|-------------------------------------|---|---------------------------------|----|
| Accessibilité des données | 3 | Image..... | 9 |
| Agora Project | 3 | Information géographique..... | 9 |
| Alignement..... | 3 | Intelligence artificielle | 9 |
| Ancre..... | 3 | Interface..... | 10 |
| Annotation | 3 | Interopérabilité..... | 10 |
| API..... | 3 | ISSN..... | 10 |
| Application | 3 | JPG/JPEG..... | 10 |
| Archive ouverte..... | 3 | Librairie | 10 |
| Ayant droit | 3 | Licence / licence libre | 10 |
| Balise..... | 4 | Lien / Lien hypertexte | 11 |
| Base de données..... | 4 | Linked Data | 11 |
| Base de données relationnelle | 4 | Logiciel | 11 |
| Bibliothèque numérique..... | 4 | Mashup | 11 |
| Carte..... | 4 | Métadonnées..... | 11 |
| Cartographie | 4 | Module..... | 11 |
| Champ..... | 4 | Moissonnage..... | 11 |
| CMS..... | 5 | Mots-clés | 11 |
| Collaboratif | 5 | Nuage de tags | 11 |
| Commentaire | 5 | Numérisation | 11 |
| Consortium | 5 | OAI-PMH | 11 |
| Coordonnées géographiques..... | 5 | Omeka..... | 12 |
| Corpus..... | 5 | Onglet | 12 |
| Creative Commons..... | 5 | Open Acces | 12 |
| Crowdsourcing | 5 | Open Data | 12 |
| CSV | 5 | Open Source | 12 |
| DBpedia..... | 5 | Page web..... | 12 |
| Déploiement d'outil | 5 | PDF..... | 13 |
| Dépouillement | 6 | Plateforme | 13 |
| Désambiguïser | 6 | Plugin | 13 |
| Description de document | 6 | PNG..... | 13 |
| Document | 6 | Post-édition..... | 13 |
| Donnée..... | 6 | RDF..... | 13 |
| Données structurées..... | 6 | READ | 13 |
| Droit d'auteur..... | 6 | Recherche à facettes..... | 13 |
| Dublin Core | 6 | Recherche avancée | 13 |
| EAD..... | 7 | Recherche plein texte | 14 |
| Encodage..... | 7 | Recherche simple..... | 14 |
| Enregistrement | 7 | Référentiel | 14 |
| Enrichissement des données | 7 | Répertoire..... | 14 |
| Entité nommée | 7 | Requête..... | 14 |
| Entrepôt OAI-PMH..... | 7 | Retroconversion..... | 14 |
| Exposition virtuelle | 7 | Serveur..... | 14 |
| Fiche..... | 7 | SIG..... | 15 |
| Fichier (bibliothèque) | 7 | Signet | 15 |
| Fichier (informatique)..... | 7 | Site | 15 |
| Format libre/propriétaire..... | 7 | Système d'information..... | 15 |
| Format ouvert/fermé..... | 8 | Tags | 15 |
| Formulaire..... | 8 | TAL..... | 15 |
| Fouille de texte | 8 | TEI..... | 15 |
| Framework..... | 8 | Thème / Template..... | 15 |
| FRBR | 8 | Thésaurus..... | 16 |
| Génétique | 8 | TIFF..... | 16 |
| Géolocalisation | 8 | Traduction... .. | 16 |
| Géomatique | 8 | Traduction automatique... .. | 16 |
| GPS | 8 | Transcription | 16 |
| Graphe | 8 | Transkribus..... | 16 |
| HTML..... | 8 | Triple store | 16 |
| HTR..... | 9 | Unicode / UFT8 | 16 |
| Huma-Num..... | 9 | URI | 17 |
| IdRef..... | 9 | Valeur | 17 |
| IIIF..... | 9 | Visualisation de données..... | 17 |

| | | | |
|--------------------------------|----|--------------|----|
| Visualisation par graphe | 17 | Wysiwyg..... | 17 |
| Voie dorée..... | 17 | XML..... | 18 |
| Voie verte..... | 17 | Zotero..... | 18 |
| Web sémantique..... | 17 | | |

Pour simplifier le système de renvoi, chaque terme faisant l'objet d'une notice est indiqué en rouge dans le texte.

A

Accessibilité des données : Il s'agit de garantir l'accès et la pérennité des données. Cette accessibilité est généralement assurée par des développements informatiques réalisés à partir de logiciels **open source**, et à partir de **librairies** sans dépendances, ou que l'on peut facilement remplacer. Elle est aussi assurée par un choix de format d'**encodage** des données, format qui doit être ouvert, documenté et utilisé par une communauté.

Agora-Project : Outil de travail collaboratif permettant de créer un espace en ligne pour une équipe et partager des **fichiers**, un fil d'actualité, un agenda, des notes, etc. Pour communiquer facilement autour d'un projet commun. Voir <https://www.agora-project.net>.

Alignement : Fait de positionner un élément d'une certaine façon par rapport aux autres. En linguistique, on parle d'aligner des éléments textuels de version ou de langue différentes, pour repérer automatiquement les transformations. Par extension, on parle d'alignement pour la correspondance entre du texte dans une **image** et sa version textuelle transcrite. En informatique, l'alignement désigne la mise en correspondance et le lissage par rapport à un standard et en bibliothèque l'alignement peut aussi désigner le fait de faire correspondre les entrées de deux **référentiels** ou **thésaurus**.

Ancre : Une ancre est l'endroit de destination d'un **lien** au sein d'une **page web** ; cela permet de se rendre directement à un emplacement précis de la page.

Annotation : L'annotation est ce qui se rajoute au texte. Elle peut être de nature explicative ou critique. Elle commente et ne décrit pas. L'annotation est dans l'**encodage** une strate supplémentaire à la documentation.

API : Le rôle d'une API (*Application Programming Interface*) ou interface de programmation est de fournir une bibliothèque d'outils (fonctions, objets, programmes) permettant l'interaction entre d'autres entités informatiques (serveurs, programmes, services).

Application : En informatique, une application est un programme utilisé pour réaliser une tâche ou un ensemble de tâches. Le terme est synonyme de **logiciel**. Un éditeur de texte, un jeu vidéo, un navigateur web sont des applications ; elles ont besoin des services d'un système d'exploitation pour fonctionner.

Archive ouverte : Une archive ouverte est un réservoir où sont déposées des publications issues de la recherche scientifique et de l'enseignement dont l'accès est libre et gratuit. Elle peut être institutionnelle (ex. OATAO de l' [Université de Toulouse](#)), régionale (ex. [OpenAIRE](#) pour l'Europe), nationale ([HAL](#) pour la France) ou disciplinaire (ex. [arXiv](#) en Physique, [RePEC](#) en Economie).

Ayant droit : L'ayant droit est une personne bénéficiant d'un droit sur un document en raison de sa situation juridique, fiscale, financière, ou d'un lien familial avec le bénéficiaire direct de ce droit. Ainsi un fonds peut être la propriété matérielle d'une bibliothèque (ex. le fonds Foucault conservé

au département des Manuscrits de la BnF) mais la propriété intellectuelle reste aux ayants droit d'un auteur. Dans le cadre de projets d'édition numérique, il faut donc demander l'autorisation de diffuser le fonds sous forme de reproduction numérique aux ayants droits et au lieu de conservation. Voir **Droit d'auteur**.

B

Balise : La balise est une marque que l'on utilise dans les langages informatiques afin de signaler une spécificité descriptive (exemple : italique, gras...) ou pour commander une action à un logiciel. Le langage **HTML** utilise des balises pour mettre en forme le texte, le **XML** fait de même mais avec une couche interprétative de ce contenu. La balise se matérialise par des chevrons ouvrants et fermants, elle est souvent double avec une balise ouvrante et une balise fermante (exemple : `<p>texte</p>`) ou se compose d'une balise unique qui souvent se termine par un / (exemple : `
` pour le retour à la ligne qui est par définition unique). Les balises ne sont pas visibles à l'écran mais elles sont consultables en demandant au navigateur l'affichage du code source de la page.

Base de données : Une base de données est une collection d'informations destinée à stocker des données de façon structurée. Le logiciel de gestion de bases de données (ou SGBD) permet de gérer les données qu'elle contient (insertion, suppression, modification). L'architecture traditionnelle qui s'est imposée dans les années 1970 est celle des bases de données relationnelles : les données y sont organisées en tableaux interreliés (correspondant à des entités distinctes), avec les colonnes représentant des types de données (**Champs**) et les lignes des ensembles cohérents de données (**Enregistrements**). Ce modèle reste encore aujourd'hui dominant, même si d'autres approches, plus souples et plus modulaires, émergent depuis 2010 (NoSQL notamment).

Base de données relationnelle : Une base de données relationnelle est une **base de données** structurée selon un modèle permettant de définir des liens entre les différents objets décrits par cette base de données. Exemple : une Commune appartient à un Département, un Département appartient à une Région.

Bibliothèque numérique : Une bibliothèque numérique (virtuelle ou en ligne ou électronique) est une collection de ressources (textes, **images**, sons) numériques (c'est-à-dire numérisées ou nées numériques) accessibles à distance (en particulier via Internet), proposant différentes modalités d'accès et de consultation à l'information à des publics différents. La bibliothèque virtuelle regroupe donc un ensemble de ressources et de services dématérialisés.

C

Carte : « La carte est une représentation géométrique conventionnelle, généralement plane, en positions relatives, de phénomènes concrets ou abstraits, localisables dans l'espace ; c'est aussi un document portant cette représentation ou une partie de cette représentation sous forme d'une figure manuscrite, imprimée ou réalisée par tout autre moyen. » Comité français de Cartographie (CFC), 1990.

Cartographie : Ensemble des études et opérations scientifiques, artistiques et techniques, intervenant dans l'élaboration d'une **carte**, d'un plan ou autre mode d'expression, à partir des résultats d'observations directes ou de l'exploitation d'une documentation, ainsi que dans leur utilisation.

Champ : Un champ est l'information élémentaire d'une **base de données**, il équivaut à la colonne d'un tableau. On peut également le définir comme la propriété d'un objet.

CMS (Content Management System) : Le CMS est un système d'interfaces qui permet de gérer la conception et la gestion d'un **site** sans avoir besoin de trop de connaissances en informatique. Par contre, la mise en forme du **site**, appelée souvent **thème**, demande des compétences en **HTML** et en d'autres langages internet. Le CMS sert essentiellement à diffuser du contenu rédactionnel sous forme de billets ou d'actualités mais ce n'est ni une **base de données** permettant l'exploitation des données ni un éditeur numérique permettant de faire de l'édition scientifique poussée. WordPress est actuellement le plus connu des CMS.

Collaboratif : Dispositif qui vise à faciliter la collaboration de différents participants ou publics grâce à des outils informatiques adaptés au partage et à l'échange d'information. On parle aussi de système contributif ou participatif et en anglais de *crowdsourcing*. La grande problématique de ce type de dispositif est la modération des contenus proposés.

Commentaire : Explication, interprétation ou analyse d'un texte ; notes et éclaircissements destinés à faciliter l'intelligence d'un texte. Voir **Annotation**.

Consortium : Association d'institutions, de structures ou de partenaires individuels, constituée dans le but de réaliser un projet commun.

Coordonnées géographiques : Couple de valeurs exprimées en degrés sexagésimaux ou décimaux, ou en grades, appelées longitude et latitude, exprimant la position d'un point situé à la surface de la Terre.

Corpus : Recueil de documents relatifs à une discipline ou une thématique, réunis en vue de leur conservation, leur édition ou leur exploitation.

Creative Commons : Publiées dès 2002, les **licences Creative Commons** (CC) proposent une solution légale aux personnes souhaitant offrir une autorisation non exclusive de reproduire, distribuer et communiquer une œuvre au public à titre gratuit. Elles permettent de faire apparaître clairement au public les conditions de la **licence** de distribution et de réutilisation de cette création. Voir : <https://creativecommons.org/share-your-work/>

Crowdsourcing : Voir **Collaboratif**.

CSV (Comma-separated values) : Le CSV est un format informatique stockant des données sous forme de **valeurs** séparées par un symbole. Le plus souvent, ce séparateur est la virgule ou le point-virgule. La représentation usuelle de ces données mémorisées en CSV est le tableau. Ce format ne permet pas d'**enrichissement** typographique (gras, italique, etc.), il conserve du texte brut.

D

DBpedia : DBpedia est un projet universitaire et communautaire d'exploration et d'extraction automatiques de données dérivées de Wikipédia. Son principe est de proposer une version structurée et sous forme de **données** normalisées au format du web sémantique des contenus de chaque fiche encyclopédique. DBpedia vise aussi à relier à Wikipédia (et inversement) des ensembles d'autres données ouvertes provenant du Web des données. Voir **Web sémantique & Open Data**.

Déploiement d'outil : Le déploiement consiste à faire passer un prototype à une version pérenne accessible en ligne et une fois cette version stabilisée, à y intégrer les **données** traitées ou à venir.

Dépouillement : Première étape d'une analyse documentaire : repérage et sélection d'informations contenues dans un document en fonction de critères prédéterminés. Le dépouillement commence par la sélection des parties composantes (articles, chapitres, **images**, séquences, etc.) qui seront décrites et analysées en fonction de la politique documentaire.

Désambiguïser : Faire disparaître l'ambiguïté d'un mot, d'une phrase en ne retenant qu'un seul sens et/ou en donnant des formes ou des **annotations** spécifiques aux différents sens.

Description de document : La description est formelle et ne concerne pas l'analyse ou toute interprétation du contenu. Le premier niveau de description d'un document est constitué par l'ensemble des **métadonnées** qui permettent aux utilisateurs et aux moteurs de recherche de retrouver le document. Le deuxième niveau de description d'un document est le plan structurel (structuration logique de son contenu).

Document : Un document renvoie à un ensemble formé par un support et une information, celle-ci enregistrée de manière persistante. Il a une valeur explicative, descriptive ou de preuve.

Donnée : Représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement et sa communication.

Données structurées : Dans les langages de formatage de type **XML** ou **HTML**, les données structurées sont des informations encadrées par des **balises** spécifiques dans les sources des **pages** et qui permettent à des outils d'édition ou d'exploitation, comme les moteurs de recherche, de les interpréter d'une certaine manière. Les données structurées répondent à un schéma d'utilisation (une norme) avec des règles de **balisage**.

Droit d'auteur : Le droit d'auteur se décompose en deux ensembles :

- a) Les droits moraux, inaliénables, qui concernent l'intégrité de l'œuvre, le droit de retrait et de repentance. Ils ne peuvent pas être cédés.
- b) Les droits patrimoniaux qui concernent l'exploitation de l'œuvre (reproduction, représentation). Ils peuvent être cédés par l'**ayant-droit**, à l'image d'un écrivain qui cède les droits de reproduction à son éditeur. Ces droits sont possédés par l'auteur de l'œuvre protégée. Il les transmet à ses héritiers – ou ayants-droit – à sa mort.

Si, en France, la durée de protection des droits patrimoniaux est de soixante-dix ans après la mort de l'auteur (sauf dérogation : les auteurs morts pour la France pendant les Guerres mondiales bénéficient de trente ans de protection supplémentaire), les droits moraux sont eux inaliénables et perdurent dans le temps.

Dublin Core : Le Dublin Core (dublincore.org) est un modèle de description de tout type de ressource numérique (audio, vidéo, livre, objet) qui propose un socle de quinze éléments. Ces quinze **champs** sont les suivants : Titre, Créateur, Sujet, Description, Source, Éditeur, Date, Couverture, Relation, Format, Langue, Type, Identifiant, Contributeur, Droit. Devenu standard international, il permet donc l'échange de **données** entre différents projets ou outils et facilite l'**interopérabilité** des **données**. Pour plus de précisions sur l'application du Dublin Core, voir la présentation très précise sur le site de la BnF :

http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html

Le Dublin Core n'impose pas de compléter tous les **champs** mais pour une meilleure diffusion et pérennité des données, il est important d'en remplir le plus possible. Ces champs sont tous répétables, sans aucune limite. Voir **Interopérabilité**.

E

EAD (Encoded Archival Description) : Standard de description archivistique qui permet d'encoder en **XML** un inventaire d'archives.

Encodage : L'encodage est l'action de structuration d'un texte avec des **balises** de différents formats (**TEI**, **HTML**, **EAD**, etc.). Chaque format possédant son propre langage mais également ses propres finalités. L'encodage peut concerner aussi bien des spécifications de mise en forme que des indications de structure ou des interprétations sémantiques.

Enregistrement : Terme consacré dans les **bases de données**, c'est l'ensemble des valeurs d'une ligne d'un tableau contenant lui-même un certain nombre de colonnes (que sont les **champs**). C'est équivalent à une **notice** ou à une **fiche**.

Enrichissement des données : Ajout d'informations complémentaires pour aider la lecture, comme la normalisation des noms propres, les **annotations**, etc. Différents types d'enrichissements :

- Enrichissement par complétion : ajout de formes normalisées de noms propres et ajout de références bibliographiques (identifiants de notices d'autorités de catalogues en ligne).
- Enrichissement par annotation : ajout de commentaires sur le sens du texte rédigé, ajout de rapprochements avec d'autres documents ou avec des œuvres dites et écrites, ajout de références complémentaires, etc.

Entité nommée : Une entité nommée est une expression linguistique référentielle, souvent associée aux noms propres et aux descriptions définies, qui a émergé avec le besoin d'applications de recherche d'information. Les entités nommées peuvent être objet de traitements à divers degrés de finesse (détection, extraction, reconnaissance, liaison). Elles sont l'objet d'une tâche du **traitement automatique du langage naturel** appelée reconnaissance d'entités nommées.

Entrepôt OAI-PMH : C'est un **répertoire** de **serveur** web sur lequel les fournisseurs de données peuvent déposer leurs **métadonnées** en attendant qu'un robot vienne les « moissonner » afin de les intégrer à son propre catalogue. Pour cela il faut utiliser le protocole **OAI-PMH**.

Exposition virtuelle : L'exposition virtuelle est un moyen de plus en plus utilisé par les musées, les détenteurs de fonds culturels, ainsi que les artistes pour diffuser leurs œuvres sur le web. Elle permet de créer des parcours thématiques à partir des données publiées.

F

Fiche : Une fiche rassemble tous les **champs** saisis dans un **formulaire**.

Fichier (bibliothèque) : **Répertoire** de fiches (unité organique et ordre thématique). Avant l'informatisation des catalogues des bibliothèques, ces derniers se présentaient sous forme d'ensemble de fiches, généralement appelé catalogue ou fichier papier. Depuis les années 90, la majorité des catalogues ont été rétroconvertis pour être consultables de manière informatique. Il est toutefois fréquent que les bibliothèques conservent leurs fichiers papier. Voir **Rétroconversion**.

Fichier (informatique) : Un fichier informatique est un ensemble de données numériques réunies sous un même nom, enregistrées en un seul contenant sur un support de stockage permanent. Un fichier informatique a un **format** numérique symbolisé par une extension finale généralement en trois ou quatre lettres.

Format libre/propriétaire : Le format libre est un format qui n'est la propriété de personne et donc exploitable par tous (ex. : **PNG**, **JPEG**). Le format propriétaire est un format rattaché à un brevet. L'utilisation du format propriétaire n'est pas forcément payante mais seule l'entreprise détentrice du brevet en a le monopole (ex. : **PDF**, **TIFF**). Les formats propriétaires peuvent être des formats ouverts ou fermés.

Format ouvert/fermé : Chaque format de **fichier** possède sa façon de coder l'information. Les spécifications techniques du fichier peuvent être rendues publiques ou non. Un format ouvert est un format dont les spécifications sont publiées (ex. : **TIFF**, ODF, **PDF**, etc). À l'inverse, un format fermé est un format dont on ne connaît pas les spécifications techniques (ex. : Excel, Word, etc.). L'ouverture ou la documentation d'un format n'empêche pas que celui-ci soit un format propriétaire.

Formulaire : On désigne par formulaire, une interface permettant de remplir des **champs** ou de spécifier des actions à réaliser. Le principe du formulaire est de saisir les données puis les sauvegarder à travers un bouton « Enregistrer / Sauvegarder ».

Fouille de texte (Text-mining) : Extraction de connaissances à l'aide de mesures statistiques ou de repérage d'unités textuelles dans un ou plusieurs textes. La fouille de texte permet ainsi d'extraire les **entités nommées**.

Framework : Un *framework* propose une infrastructure de développement pour un informaticien afin qu'il puisse créer une **application** (des lignes de code permettant de réaliser des actions).

FRBR : Modèle conceptuel de description bibliographique utilisé en bibliothèque. Il décompose la notice bibliographique comme un ensemble d'informations correspondant à 4 niveaux d'analyse : Item, Manifestation, Expression, Œuvre. Voir :

http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html

G

Génétique : Science qui vise à analyser et à éditer tous les états d'un texte.

Géomatique : Le mot « géomatique » vient de la contraction des termes « géographie » et « informatique ». Il s'agit du domaine informatique ayant pour objet la gestion, de façon numérique, de l'**information géographique**. « Discipline ayant pour objet la gestion des données à référence spatiale [géoréférencées, c.a.d. localisables de façon géographique] et qui fait appel aux sciences et technologies reliées à leur acquisition, leur stockage, leur traitement et leur diffusion. » (Marcel Bergeron, 1992, Vocabulaire de la Géomatique). Voir **SIG**.

GPS (Global Positioning System) : Système permettant de déterminer les **coordonnées géographiques d'un point du globe** à partir de l'observation des signaux radio émis par des satellites en orbite autour de la terre. En fonction de la méthode d'utilisation, du nombre et du type de récepteurs exploités, la précision obtenue sur les coordonnées varie de 100 mètres à quelques centimètres. Ces satellites ont été mis en orbite par les États-Unis d'Amérique.

Graphe : voir **Visualisation**.

H

HTML (HyperText Markup Language) : Le HTML est un langage qui permet d'encoder les pages web. C'est un **encodage** formel totalement différent de l'encodage structurel ou sémantique. HTML permet également de structurer *a minima* et de mettre en forme le contenu des pages, d'inclure des ressources multimédias dont des **images**, des **formulaire**s de saisie et des programmes informatiques. Il permet de créer des documents interopérables avec des équipements très variés de manière conforme aux exigences de l'**accessibilité des données** du web. En résumé, il permet de rendre visible un contenu proposé par une **page web** sur un navigateur web.

HTR (Handwritten Text Recognition) : Technologie de reconnaissance automatique d'écriture manuscrite. Voir **Transkribus**.

Huma-Num : Une très grande infrastructure de recherche consacrée au développement du numérique au sein des SHS et proposant plusieurs services ou outils aux acteurs des SHS en France.

I

IdRef (Identifiants et Référentiels pour l'Enseignement supérieure et la Recherche) : Application Web développée et maintenue par l'ABES (Agence bibliographique de l'enseignement supérieur, www.abes.fr/) qui permet, à des structures ou des usagers, d'interroger, consulter, créer et enrichir des notices d'autorité (<https://www.idref.fr>). Les catalogues Calames, SUDOC, theses.fr ou Persée l'utilisent pour leurs notices d'autorité.

Voir <http://documentation.abes.fr/aideidrefutilisateur/index.html>

IIIF (International Image Interoperability Framework, <https://iiif.io>) : Le consortium IIIF est né du constat que les bibliothèques numériques se sont développées sans concertation technique, notamment pour la production des **images**, ce qui rend aujourd'hui difficile leur partage. Si les protocoles d'échange des données (**OAI-PMH** par exemple) sont aujourd'hui très développés, rien n'existait pour les images. L'objectif du consortium est de développer un cadre d'**interopérabilité** pour la diffusion d'images haute résolution grâce à des API développées par le consortium. Une application possible serait de consulter sur une même **interface**, des images numérisées de plusieurs bibliothèques numériques. Le visualiseur Mirador a été développé à cette fin : <https://demos.bibliissima.fr/mirador/#8405964d-ac0f-4aac-b87d-731ecc1d535a>

Voir aussi : <https://doc.bibliissima.fr/iiif>

Image numérique : Image acquise, créée, traitée et stockée sous forme binaire, c.a.d. acquise par des convertisseurs analogique-numérique situés dans des dispositifs comme les scanners, les appareils photo, etc. Les formats d'images numériques les plus fréquents sont le .jpg (compression destructrice, poids de l'image réduit), le .GIF (peut être animé, format léger) .png (compression sans perte, bon compromis), le .pdf (possibilité de contenir des images en pixels et des données vectorielles, conserve la mise en page), le .TIFF (utilisé par les imprimeurs). Les formats recommandés :

Images à mettre sur le web : .jpg

Si besoin d'une compression non destructrice : .png pour informatique et .tiff pour l'impression

Pour un CV, mémoire, ou autres documents : .pdf

Voir **JPG/JPEG ; PDF ; PNG ; TIF**

Information géographique : Information qui est reliée à une localisation sur la Terre, exprimée par rapport à un système de référence. Une information géographique est une information que l'on peut situer sur un plan, une **carte**, directement par des coordonnées ou indirectement par relation à une autre information géographique. C'est en particulier, l'information sur les objets ou phénomènes naturels, les ressources culturelles, humaines ou économiques.

Intelligence artificielle (I. A.) : L'Intelligence artificielle (ou « *AI* » en anglais, pour *Artificial Intelligence*) est définie par l'un de ses créateurs comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique » (M. Lee Minsky). Il s'agit en quelque sorte de l'utilisation des ordinateurs ou de processus électroniques dans le but d'imiter le comportement humain, dans les domaines du raisonnement (jeux ou pratique des mathématiques), de la compréhension des langues naturelles, dans la commande d'un robot, etc.

Interface : Jonction entre deux matériels ou **logiciels** leur permettant d'échanger des informations par l'adoption de règles communes ; **module** matériel ou logiciel permettant la communication d'un système avec l'extérieur.

Interopérabilité : Possibilité de communication entre deux ou plusieurs systèmes, appareils ou éléments informatiques. Voir **Dublin Core** ; **HTML** ; **IIIF**

ISSN (International Standard Serial Number) : Identifiant bibliographique unique et standardisé qui s'applique aux publications en série (revues, magazines, journaux, bulletins, etc.), en cours de parution ou ayant cessé de paraître, quel qu'en soit le support, gratuites ou non, en accès libre ou non.

J

JPEG (*Joint Photographic Experts Group*) : Format standardisé d'**images** numériques qui permet une **visualisation** dans la plupart des logiciels d'images ainsi que sur les navigateurs internet. C'est le format image par excellence pour internet. Ce format ouvert et libre est associé à une méthode de compression avec pertes qui permet de diffuser des images au moindre coût taille/qualité. Mais la perte d'informations due à la compression est irréversible. Si la compression est trop forte, l'image s'affichera dans une qualité médiocre, voire de façon pixellisée. On parle aussi d'« image brûlée ». Les préconisations sont de numériser dans des formats non compressés et de haute qualité (TIFF généralement) puis de faire une copie JPEG pour la visualisation. Il existe désormais un nouveau format JPEG : le JPEG 2000 dont la méthode de compression est nettement moins destructive. À ratio équivalent, la compression JPEG 2000 est de meilleure qualité. Le JPEG 2000 peut également être utilisé sans compression, ce qui le rend très intéressant en termes de taille/qualité, notamment pour des **fichiers** de haute qualité. La BnF l'a adopté comme format de **numérisation** haute résolution et d'archivage pérenne.

L

Librairie (*Library*) / bibliothèque logicielle ou de programmes : Ensemble de fonctions utilitaires, regroupées et mises à disposition sous forme de routines ou modules préprogrammés afin de pouvoir être utilisées sans avoir à les réécrire.

Licence / licence libre : Un auteur peut décider de placer son œuvre sous licence libre, c'est-à-dire qu'il donne l'autorisation gratuite, à tous et par avance, d'utiliser son œuvre dans les conditions fixées dans la licence. Il existe plusieurs standards de licences libres (par exemple, pour les logiciels, les licences CeCILL, GNU GPL). Pour les créations, il existe les licences en **Creative Commons** (CC) ou encore **Art Libre**. Par exemple, les licences CC permettent à l'auteur, par le biais de variantes, d'indiquer aux utilisateurs de quelles libertés ils disposent sur l'œuvre et quelles

sont leurs obligations. Les 6 licences CC type autorisent toujours la libre diffusion de l'œuvre, mais peuvent interdire l'utilisation commerciale (NC) et les modifications (ND) ou encore imposer le maintien de la licence pour les œuvres dérivées (SA). Elles imposent toutes la mention du nom de l'auteur (BY). Par conséquent, lorsqu'une œuvre est placée sous ce type de licence, à condition d'en respecter les termes, toute personne peut utiliser l'œuvre sans avoir à solliciter une autorisation spéciale auprès de l'auteur.

Voir [Creative Commons](#), [Open Source](#) et <https://www.economie.gouv.fr/apie/propriete-intellectuelle-publications/contenus-sous-licences-libres>

Lien / Lien hypertexte : Lien opéré par un code **HTML** qui relie deux **pages** ou deux endroits accessibles par le **protocole HTTP** (web), il permet de passer automatiquement d'un document à un autre. Quand on clique sur un lien hypertexte, le navigateur nous envoie à une autre page internet (du même **site** ou d'un autre **site**) ou à un autre endroit dans la même page (lien interne avec une **ancree**), soit dans la même fenêtre, soit dans un nouvel **onglet**.

Linked Data : Notion introduite en 2006, par Tim Berners-Lee. Il s'agit d'une méthode de publication de données structurées, de manière à ce qu'on puisse établir efficacement des relations (**liens**) entre les données. Cela permet la création d'un réseau global d'informations et le découplage des données. Voir [Web sémantique](#).

Logiciel : Voir [Application](#).

M

Mashup / Application composite : Application qui permet sur un site web d'agréger ou retravailler de l'information en provenance d'une ou plusieurs sources extérieures. (Récupération automatique de données).

Métadonnées : Les métadonnées permettent la description et le traitement des ressources numériques (ou papier), elles sont généralement standardisées et à l'extérieur ou en entête du texte ou du document qu'elles décrivent. On distingue plusieurs types de métadonnées, descriptives ([EAD](#), [Dublin Core](#), MODS), techniques (EXIF, MIX-NISO, etc.), de structure (ALTO, METS, [TEI](#)).

Module : Voir [Plugin](#).

Moissonnage : Voir [OAI-PMH](#).

Mots-clés : Voir [Tags](#).

N

Nuage de tags : Voir [Tags](#).

Numérisation : Processus qui consiste à convertir des informations d'un support (texte, image, audio, vidéo) ou d'un signal électrique en données numériques. Pour la numérisation des **images**, voir [JPEG](#).

O

OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) : L'OAI-PMH est un protocole informatique développé afin d'échanger des **métadonnées** – uniquement celles-ci et non les ressources elles-mêmes. L'OAI-PMH définit deux types d'acteurs : les fournisseurs de données, qui déposent leurs **métadonnées** sur un **serveur** web appelé « **entrepôt** », et les fournisseurs de service qui collectent (qui « moissonnent ») ces données. Le moissonnage s'effectue à partir de **requêtes** formalisées à l'adresse de l'entrepôt, les résultats sont alors intégrés dans l'index ou le **répertoire** du **site** moissonneur. La liste des requêtes peut être consultée par exemple sur http://www.bnf.fr/documents/intro_oaipmh.pdf. Le standard de base d'échange du protocole OAI-PMH est le Dublin Core mais d'autres formats de **métadonnées** peuvent être exposés (EAD, RDF, etc.).

Ce protocole est utilisé notamment par les Archives Ouvertes et les **entrepôts** institutionnels, il s'est aujourd'hui largement répandu dans les institutions patrimoniales et notamment les bibliothèques. Il permet entre autres de construire des **sites** portails thématiques avec uniquement le résultat de **requêtes** sur les entrepôts repérés sur cette thématique. *Europeana*, la bibliothèque numérique européenne, est alimentée via le protocole OAI-PMH (<https://www.europeana.eu/portal/fr>).

Omeka : Logiciel de gestion de bibliothèque numérique mis à disposition sous **licence libre** (GNU – General Public License). De conception modulaire, l'outil permet à chaque **site** d'adapter les fonctionnalités proposées à l'aide de **plugins** et de **thèmes**. L'outil est développé aux États-Unis par le [Roy Rosenzweig Center for History and New Media](#) (CHNM) de l'Université George Mason qui est aussi à l'origine du logiciel de gestion bibliographique **Zotero**.

Onglet : Un onglet est, dans un **fichier** ou autre système de rangement, une petite excroissance visuelle porteuse d'une étiquette (typiquement, alphabétique) permettant un accès direct aisé aux documents ou une partie du document. Par analogie, sur les navigateurs internet, il permet d'avoir accès à plusieurs **sites** ou pages sur une seule fenêtre et de pouvoir passer rapidement d'une page à une autre. Cette interface riche sur une seule fenêtre provoque le risque d'avoir trop d'onglets ouverts...

Open Access : L'*Open Access* (ou aussi « libre accès », ou encore « accès ouvert ») à la littérature scientifique est un mode de diffusion des articles de recherche sous forme numérique, gratuite et dans le respect du droit d'auteur. Cette notion recouvre l'accès ouvert (*gratis open access*), les données diffusées en ligne gratuitement et le libre accès (*libre open access*), données diffusées en ligne gratuitement et librement, c.a.d. soumises à une **licence** d'utilisation dite libre (ex. **Creative Commons**). On distingue plusieurs modèles ou voies de l'*open access* : la **voie verte** et la **voie dorée**. Voir <https://openaccess.couperin.org/comment-definir-lopen-access/>

Open Data : Données ouvertes, dont l'accès est public et libre de droit, tout comme leur exploitation.

Open Source : Cela désigne le « code source ouvert » et s'applique aux logiciels dont la **licence** respecte les possibilités d'accès au code source du logiciel, de libre redistribution de ce code et de possibilités de travaux dérivés à partir de celui-ci. On peut ainsi adapter le code source d'un logiciel open source à ses propres besoins. La principale **licence** pour le logiciel open source est la **licence** GNU (General Public License). De plus en plus, on étend la définition de « l'open source » aux données et plus uniquement au code informatique, on parle alors d'*Open Data*.

P

Page web : C'est l'unité de base du web. Elle est conçue pour être consultée par un navigateur web et elle est identifiée par une adresse web. Elle est généralement constituée d'une structure en **HTML**, avec du texte et souvent d'**images**, de feuilles de style ou de scripts permettant l'affichage de données venant de bases de données. Elle est fabriquée à partir d'un éditeur **HTML** et localisée sur un **serveur** web (hébergement). Son affichage peut être paramétré pour s'adapter aux conditions locales de consultation (écran d'ordinateur fixe ou portable, écran de télévision, téléphone mobile, etc.) à travers une feuille de style.

PDF (Portable Document Format) : Le PDF est un format qui préserve la mise en forme d'un document – polices de caractère, **images**, objets graphiques, etc. – telle qu'elle a été définie par son auteur – et cela quels que soient le logiciel, le système d'exploitation et l'ordinateur utilisés pour l'imprimer ou le visualiser (au contraire des formats des traitements de texte). Il s'est très vite imposé comme format d'échange et d'archivage. Le format PDF n'est pas figé : il peut avoir des options personnalisées (compression des images et des textes, interdiction d'impression ou de modification...). Il peut surtout être uniquement graphique (vous ne pouvez pas copier le texte que vous voyez, il s'agit généralement d'une image qu'on a transformée en PDF) ou avoir une structure textuelle (vous pouvez copier le texte que vous voyez dans le **fichier** PDF). Voir **Image**.

Plateforme : Une plateforme informatique est un espace de travail virtuel qui permet d'utiliser un ensemble de logiciels, de stocker et de diffuser des données, et enfin de travailler à plusieurs. Elle se confond souvent avec un **site** internet (qui n'est basé que sur une seule technologie) ou avec une **bibliothèque numérique** (qui rassemble des contenus ayant un lien entre eux).

Plugin : En informatique, un *plugin* ou *plug-in*, aussi nommé **module** greffon ou plugiciel (ou extension dans les CMS ou **Omeka**), est un paquet structuré de codes informatiques qui complète un logiciel hôte pour lui apporter de nouvelles fonctionnalités.

PNG (Portable Network Graphics) : Format d'**image** ouvert standardisé par le W3C. Il a été conçu pour contourner le format GIF devenu semi-propriétaire et restrictif (nombre de couleurs notamment). Il s'agit d'une alternative intéressante par rapport au format **JPEG** car sa compression n'est pas destructive, ce qui implique aussi que son poids sera plus important. Malgré de nombreux avantages (gestion de la transparence notamment), il reste encore peu utilisé car souvent associé à un usage web uniquement. Voir **Image**.

Post-édition : La post-édition désigne l'activité qui consiste à repasser derrière un texte pré-traduit automatiquement pour le rendre humainement intelligible. Le langagier chargé d'effectuer cet exercice, à savoir le post-éditeur, a donc pour tâche de compléter, modifier, corriger, remanier, réviser et relire ce texte brut.

Voir **Traduction** ; **Traduction automatique** et <https://journals.openedition.org/traduire/460#tocto1n2>

R

RDF (Resource Description Framework)_: Modèle simplifié de description de données dont le principe de base consiste à transformer l'information des ressources afin qu'elles puissent être lisible par les machines et permettre, par conséquent, la création de liens à partir des valeurs des relations. Sa « grammaire » est constituée de *triplets* de trois éléments : sujet, prédicat et objet. Les données RDF sont stockées dans un **triple store**. Voir **Web sémantique**.

READ (Recognition and Enrichment of Archival Documents) : Plateforme de transcription et outil de reconnaissance automatique d'écriture manuscrite (**HTR**). Voir aussi **Transkribus**.

Recherche à facettes : La recherche à facettes est basée sur une classification préalable des données qui fonctionne à la manière d'un crible : les facettes proposent un résultat en fonction de l'indexation des données à l'intérieur de la classification. Elle se distingue de la **recherche avancée** en ce qu'elle ne permet pas de construire des **requêtes** personnalisées, par exemple en ajoutant des opérateurs booléens (« et », « ou », « sauf »).

Recherche avancée : Recherche par **requête** ou multi-critères. La recherche avancée peut aussi porter sur des **métadonnées** ou des **annotations**.

Recherche plein texte : La recherche plein texte consiste en une technique de recherche au sein d'un document électronique ou d'une **base de données** textuelles, dans laquelle le moteur de recherche examine tous les mots (chaînes ou suites de caractères) de chaque document enregistré.

Recherche simple : La recherche simple se concentre sur un seul champs de recherche (mot, auteur, titre), à l'inverse de la **recherche avancée** qui permet de croiser plusieurs critères de recherche (titre et auteur ; titre, auteur et date de publication, etc.).

Référentiel : Ensemble d'informations servant de références, parce qu'elles font autorité, ou parce qu'elles représentent un point de vue privilégié ou offrent une description stable d'une réalité. Un dictionnaire, une nomenclature, un système de coordonnées sont des référentiels. Certains référentiels sont constitués de données structurées selon des schémas et/ou des vocabulaires standardisés afin de pouvoir être mis en commun d'un **système d'information** à un autre. Plus généralement on appelle souvent référentiel un **thésaurus** vérifié et contrôlé permettant d'enrichir des données au sein d'un système d'information. AuréHal (<https://aurehal.archives-ouvertes.fr/>) donne accès par exemple à l'ensemble des référentiels utilisés par la base de données de l'archive ouverte HAL, sous forme de thésaurus contrôlés. Certains de ces thésaurus sont ouverts (celui des auteurs peut être enrichi par les interventions des usagers de l'archive), d'autres sont fermés (le thésaurus des domaines de recherche associés aux publications par exemple). Pour produire et exposer des données de bonne qualité, la plupart des instruments numériques s'appuient sur de tels référentiels, ouverts ou fermés : l'un des plus impressionnant est Rameau qui tient lieu de méta-référentiel pour les données des catalogues de la BNF et les données d'autorité (voir <https://www.bnf.fr/fr/indexation-sujet-les-referentiels-utilises-par-la-bnf#bnf-rameau> et <https://isidore.science/vocabularies>). Voir **Base de données** ; **Interopérabilité**.

Répertoire : Inventaire méthodique (énumération, liste, table, etc.) où les informations sont classées dans un ordre qui permet de les retrouver facilement, support d'informations. En informatique, un répertoire (dossier ou *folder*) est une liste de descriptions de **fichiers**. L'endroit de rangement de nos fichiers informatiques.

Requête : En informatique, le terme requête peut prendre plusieurs sens. Il peut s'agir d'une expression saisie dans un navigateur internet pour interroger un moteur de recherche afin de trouver l'adresse d'un site. Il désigne également l'URL d'une page web, saisie dans la barre d'adresse du navigateur web pour atteindre cette page. Dans le monde des **bases de données**, une requête SQL est un ordre d'exécution de traitement sur les données (extraction ou modification de données, par exemple).

Rétroconversion : Informatisation d'un catalogue papier afin de le rendre consultable via un catalogue en ligne. L'informatisation des catalogues permet aux lecteurs de pouvoir effectuer des **requêtes** plus ou moins complexes (**recherche simple** / **recherche avancée**) sur les collections, et offre de nouveaux services tels que l'affinage des résultats par facettes (**recherche à facettes**). On

parle aussi de plus en plus de **rétroconversion** pour tout processus de mise à disposition numérique d'un contenu auparavant imprimé (pour des anciens numéros de revues par ex).

S

Serveur : Un serveur informatique est un ordinateur qui offre des services à un ou plusieurs clients (parfois des milliers). Généralement, on parle de serveur pour désigner la machine qui héberge et diffuse des **sites** internet.

SIG : Un **système d'information** géographique est un logiciel informatique conçu pour acquérir, stocker, traiter et diffuser de l'information géographique, sous forme de plans et de **cartes**. Voir **Géomatique**.

Signet : Voir **Onglet**.

Site : Un site web, ou simplement site, est un ensemble de **pages web** et de ressources reliées par des **liens hypertextes** ; il est défini et accessible par une adresse web. Un site est hébergé sur un **serveur** web accessible via le réseau mondial internet ou via un intranet local. L'ensemble des sites web constituent le World Wide Web.

Système d'information : Système constitué des ressources humaines (le personnel), des ressources matérielles (l'équipement) et des procédures permettant d'acquérir, de stocker, de traiter et de diffuser les éléments d'information pertinents pour le fonctionnement d'une entreprise ou d'une organisation.

T

Tags : Terme anglais qu'on traduit par « étiquette » ou « mot-clé », il décrit une caractéristique de l'objet décrit et sert à faire des regroupements faciles des objets ayant les mêmes mots-clés. C'est l'unité de base d'une indexation, et c'est aussi la **métadonnée** la plus complexe à spécifier : il faut éviter toute redite par rapport aux autres **métadonnées** et choisir un nombre très limité de mots-clés, de préférence thématique, pour que l'indexation soit pertinente. Les **thésaurus** ou **référentiels** sont une manière d'indexer en limitant le nombre de tags.

Une **visualisation** par nuage de tags est une représentation visuelle des tags les plus utilisés sur un **site** web : généralement, les tags s'affichent dans des tailles et des polices de caractères d'autant plus visibles qu'ils sont utilisés ou populaires.

TAL (Traitement automatique des langues) : Le TAL est un domaine de recherche pluridisciplinaire au carrefour de la linguistique, de l'informatique et de l'**intelligence artificielle** (en particulier l'apprentissage artificiel). Il entretient aussi des liens privilégiés avec d'autres domaines, comme la didactique ou les sciences cognitives. Le TAL vise à modéliser le langage humain à des fins d'automatisation au moyen de méthodes symboliques et statistiques ; récemment, les approches neuronales (apprentissage profond) ont permis des avancées importantes. Le TAL utilise et produit des outils logiciels et des ressources linguistiques.

Le TAL peut permettre de repérer les **entités nommées** dans un texte, identifier des concepts, des acteurs et leurs relations ou encore regrouper les documents similaires dans un corpus (clusterisation). La traduction automatique ou la recherche d'informations multilingues (par exemple dans des bases de brevets) sont des applications phares du TAL. À un niveau plus théorique, le TAL a aussi permis des avancées importantes dans des domaines comme l'acquisition du langage ou la modélisation de l'évolution des langues (diachronie).

TEI (*Text Encoding Initiative*) : La TEI est un consortium fondé pour normaliser l'usage du langage **XML** pour l'**encodage** sémantique de textes historiques et littéraires. Par extension, on appelle TEI l'ensemble des **balises** et leurs règles d'application telles que définies et régulièrement mises à jour par le consortium. Voir <https://tei-c.org/>

Thème / Template : Un *template* (ou « thème », « *layout* », etc.) désigne l'enveloppe graphique d'un **site** internet, indépendamment de son contenu. Il s'agit par exemple de la disposition des colonnes, du choix des caractères ou des couleurs, de la structure des différents éléments, etc. Un *template* propose plusieurs **pages** web de base et des feuilles de style. Cela permet de séparer le contenu (les données) et le contenant (le thème), celui-ci pouvant être changé facilement sur les différents **CMS**.

Thésaurus : **Répertoire** structuré de termes (mots clés) pour l'analyse de contenu et le classement de documents. Voir aussi <https://www.cnrtl.fr/definition/th%C3%A9saurus>

TIFF (Tag Image File Format) : Format image propriétaire mais pleinement documenté, il est maintenant maintenu par Adobe. Ce format se caractérise par un en-tête riche comportant des **métadonnées** de format EXIF, IPTC et XMP. Doté d'une version sans compression, c'est un format image largement utilisé pour la conservation pérenne de **numérisation** haute qualité et pour l'impression en couleurs. Voir **Image**.

Traduction : Traduire s'est transposer dans une langue cible un texte donné dans une langue source. La traduction doit rendre compte le plus fidèlement possible du texte d'origine tout en étant parfaitement intelligible et correct dans la langue cible. Elle ouvre une multitude de questions de nature linguistique, psychologique, voire philosophique, afin de s'interroger sur les caractéristiques d'une « bonne traduction ».

Traduction automatique : Aussi appelée, *Machine Translation* (**MT**) ou TAO (traduction automatique outillée), la traduction automatique est la traduction d'un texte effectuée par ordinateur, sans intervention humaine. Ses débuts remontent aux années 50. D'abord basée sur des dictionnaires et règles de transfert, puis sur l'analyse statistique de très grands corpus, elle s'appuie désormais sur l'apprentissage profond. La qualité des traductions réalisées automatiquement est une question largement débattue. Les mémoires de traduction (**bases de données** où les traducteurs peuvent trouver des exemples de traductions passées) et la post-édition par un correcteur humain permettent d'améliorer les qualités de la traduction automatique. Voir **Intelligence artificielle**.

Transcription : En paléographie, la transcription consiste à reproduire un texte manuscrit, en notant les particularités du texte et rétablissant (ou non) les erreurs ou les abréviations qu'il peut contenir ; on parle de transcription diplomatique quand tous les phénomènes visibles du texte sont reproduits (comme la reproduction des retours à la ligne). En édition numérique, il s'agit de reproduire sous forme textuelle un texte qui a été numérisé en mode image. La transcription en contexte numérique obéit aux mêmes principes et méthodologies que la transcription « classique », les problèmes étant les mêmes. Tout travail de transcription est basé sur des principes d'**annotation**. Voir **Transkribus**.

Transkribus : Plateforme de reconnaissance de caractère et de **transcription** de manuscrits ou d'imprimés. Voir : <https://transkribus.eu/Transkribus>

Triple store : Entrepôt de données conçu pour le stockage et la récupération de données structurées en **RDF**. Le langage de **requête** est SPARQL.

U

Unicode / UTF-8 : L'Unicode est un standard informatique international qui permet de décrire toutes les lettres des différentes langues. Il vise au codage du texte écrit en donnant à tout caractère de n'importe quel système d'écriture un identifiant numérique, et ce de manière unifiée, quelle que soit la plateforme informatique ou le logiciel utilisé (à la différence d'ANSI). Il est plus complet que le code ASCII qui ne possède pas de signe diacritique. Le code informatique d'Unicode est standardisé par l'UTF (Universal Character Set Transformation Format) ; nous sommes maintenant en UTF-8. Le fait que le caractère soit codé en UTF ne veut pas dire qu'il va s'afficher correctement : il faut ensuite disposer de la police de caractère adéquate mais le caractère sera bien interprété informatiquement.

URI (Uniform Resource Identifier – Identifiant Uniforme de Ressource) : Chaîne de caractères qui identifie de façon unique une ressource sur un réseau. L'adresse URI doit permettre d'identifier une ressource de manière permanente, même si la ressource est déplacée ou supprimée. Une norme gérée par le W3C gère la syntaxe des adresses URI. L'URL (Uniform Resource Locators) qui permet d'identifier la localisation d'une ressource et l'URN (Uniform Resource Names) qui permet d'identifier une ressource, mais pas de la localiser, sont des spécialisations d'URI.

Voir **Web sémantique**.

V

Valeur : La valeur est ce qui est donné dans un **champ** à un **enregistrement** (ex. « 1912 » est la valeur pour le **champ** Date pour tel document).

Visualisation de données : La visualisation de **données** désigne la représentation graphique d'informations et de données. À l'aide d'éléments visuels comme les graphiques et les **cartes**, une visualisation de données permet de voir et de comprendre des tendances ou des valeurs inhabituelles dans les données, de manière très accessible. Dans le monde du *Big Data*, les outils et technologies de visualisation de données sont indispensables pour analyser d'énormes volumes d'informations et prendre des décisions en s'appuyant sur les données. Voir : <https://www.tableau.com/fr-fr/learn/articles/data-visualization>

Visualisation par graphe : Un graphe est une représentation graphique avec un ensemble de points, dont certaines paires sont directement reliées par un ou plusieurs liens. Cette technique permet de visualiser de façon différente et précise à la fois les « processus » ou les relations établies entre des données : elle permet de créer un dispositif de représentation de celles-ci dans un ensemble beaucoup plus fin et visuel qu'une simple liste à puces. Mais ce type de visualisation est basé sur des relations ou des rapports entre les données. Voir <http://innovatives.cnrs.fr/IMG/pdf/s6-auber.pdf>

Voie dorée : La voie dorée ou *gold open access* concerne des revues ou ouvrages nativement en **Open Access**, dès leur publication. Voir <https://openaccess.couperin.org/la-voie-doree-2/>

Voie verte : La voie verte ou *green open access* est la voie de l'auto-archivage ou dépôt par l'auteur dans une **archive ouverte**. Voir <https://openaccess.couperin.org/la-voie-verte-2/>

W

Web sémantique : Le Web sémantique, appelé aussi Web de données, est le Web permettant d'échanger et d'utiliser des données, de publier et de lier des bases de données sur le Web. Succédant au Web documentaire, il s'appuie sur un standard du Web, l'**URI** (Uniform Resource

Identifier), qui identifie une ressource. Le modèle de données **RDF**, également standard du Web sémantique, permet quant à lui de décrire, représenter et relier des données. Voir **DBpedia**.

Wysiwyg : Wysiwyg est un acronyme anglais qui signifie « *what you see is what you get* » : « ce que vous voyez est ce que vous obtenez ». Cela désigne une interface graphique (le plus souvent par **formulaire** ou bouton) qui permet de composer visuellement le résultat attendu sans passer par l'écriture et donc l'apprentissage de codes informatiques. L'exemple classique est Word, logiciel de traitement de texte qui permet d'éditer un texte sans passer par **XML**.

X

XML (eXtensible Markup Language, « langage de balisage extensible ») : Pour pouvoir être lue et archivée, une ressource numérique demande un **encodage** qui respecte les exigences de son auteur mais qui soit aussi compréhensible par d'autres. De nombreux standards d'encodage existent. Mais pour la représentation et l'échange des informations contenues dans la ressource, le XML est devenu le langage de référence. Il est utilisé dans de nombreuses situations et a développé des initiatives dérivées qui permettent de répondre à de nombreux besoins (dont XML-**TEI**). Le **HTML** est un langage avec une liste fermée de **balises** qui ne s'occupent que de la mise en forme. Le XML propose une couche supplémentaire avec une liste non limitée de balises qui permettent de structurer son propre langage : elles concernent généralement la structure ou l'interprétation du contenu. Le XML a donc une structure ouverte, les balises ne sont pas limitées mais il y a des règles d'utilisation à respecter. En tête d'un document XML, il y a généralement les **métadonnées Dublin Core**.

Z

Zotero : C'est un logiciel de gestion de références bibliographiques gratuit et **open source**. Il permet de gérer des données bibliographiques et des documents de recherche (fichiers **PDF**, **images**, etc.). Il s'intègre au navigateur web et permet de synchroniser des données depuis plusieurs ordinateurs, ainsi que de faire de la génération de citations (notes et bibliographies). Le développement du logiciel est à l'initiative du [Roy Rosenzweig Center for History and New Media](#) (CHNM) de l'université George Mason, le même centre qui développe **Omeka**.

Annuaire des ressources

Cet annuaire, forcément incomplet, vous donne les adresses de ressources fréquemment citées ou utilisées pour ce glossaire. Le classement est alphabétique par titre de la ressource.

Ressources numériques :

- ArcGIS : Système complet qui permet de collecter, organiser, gérer, analyser, communiquer et diffuser des informations géographiques : <https://learn.arcgis.com/fr/>
- Biblissima : Observatoire du patrimoine écrit du Moyen Âge et de la renaissance : <https://projet.biblissima.fr>
- BnF pour les professionnels : <https://www.bnf.fr/fr/Signaler> ; <https://www.bnf.fr/fr/outils-de-la-numerisation>
- BVH : Bibliothèques Virtuelles Humanistes - <http://www.bvh.univ-tours.fr>
- CAHIER : Corpus d'auteurs pour les Humanités. Informatisation, édition recherche : <http://cahier.hypotheses.org>
- Calenda : Plateforme en ligne dédiée à l'actualité de la recherche en lettres et sciences humaines et sociales. Elle publie des annonces de colloques, les programmes de séminaires, les cycles de conférences, les propositions d'emploi et les appels à contribution. <https://calenda.org>
- Cortext : Plateforme d'analyse et de visualisation de réseaux : <https://www.cortext.net/>
- Data Bnf : La bibliothèque de France nous guide dans ses ressources en regroupant sur une même page toutes les informations issues de ses différents catalogues, ainsi que de sa bibliothèque numérique Gallica : <https://data.bnf.fr/>
- DeepL : Plateforme de traduction automatique multilingue : <https://www.deepl.com/fr/translator>
- Dissem : Dissem détecte les articles derrière les péages et aide leurs auteurs à les télécharger en un clic vers un dépôt ouvert : <https://dissem.in>
- Eman : Plateforme d'édition de manuscrits et de fonds d'archives modernes numérisées : <http://www.eman-archives.org>
- Epi-revue : La plateforme propose un outil complet pour la gestion de la revue, son hébergement et la diffusion de ses contenus : <https://www.ccsd.cnrs.fr/epi-revues>
- Gallica : Bibliothèque numérique de la Bibliothèque nationale de France : <https://gallica.bnf.fr>
- GéOInformations : Espace interministériel de l'information géographique (avec un glossaire de l'information géographique : <http://www.geoinformations.developpement-durable.gouv.fr/glossaire-de-l-information-geographique-a855.html>)
- Gephi : Logiciel libre d'analyse et de visualisation de réseaux : <https://gephi.org/>
- GeoNames : Référentiel de noms géographiques : <https://www.geonames.org/>
- GITHUB : Plateforme de partage de code : <https://github.com>
- Hal : Archives ouverte : <https://hal.archives-ouvertes.fr>
- Huma-Num : Très grande infrastructure de recherche des humanités numériques proposant différents services : <https://www.huma-num.fr/>
- Hypothèses : Carnets (blogs) de recherche : <https://fr.hypotheses.org>

- **IIIF** (*International Image Interoperability Framework*) : Communauté et spécificités techniques pour définir un cadre d'interopérabilité pour la diffusion d'images haute résolution sur le Web : <https://iiif.io> ; <https://doc.bibliissima.fr/iiif>
- **Initiative_Digit_Hum** : ateliers, portraits, enquêtes, portfolio et ressources : <https://digithum.huma-num.fr/>
- **Isidore** : Accès aux données et services numériques de SHS : <https://www.rechercheisidore.fr>
- **ISTEX** : Socle de la bibliothèque scientifique numérique nationale : <https://www.istex.fr>
- **Khartis** : Créer simplement des cartes thématiques : <https://www.sciencespo.fr/cartographie/khartis/>
- **Magrit** : Une solution pour créer des cartes thématiques : <http://magrit.cnrs.fr/>
- **OpenEdition** : Ressources électroniques en sciences humaines et sociales. OpenEdition est une infrastructure complète d'édition électronique au service de la communication scientifique en sciences humaines et sociales. Elle rassemble quatre plateformes complémentaires dédiées (Revues.org, OpenEdition Books, Hypothèses, Calenda). <https://www.openedition.org>
- **OpenEdition Books** : Collections de livres : <https://books.openedition.org>
- **Persée** : Structure de service ayant pour mission de valoriser le patrimoine documentaire au bénéfice de la recherche en assurant sa diffusion, son enrichissement et sa préservation : <https://www.persee.fr>
- **Rameau** : Méta-référentiel pour les données des catalogues de la BNF et les données d'autorité : <https://www.bnf.fr/fr/indexation-sujet-les-referentiels-utilises-par-la-bnf#bnf-rameau>
- **Scripta-PSL** : Programme IRIS de PSL « Histoire et pratiques de l'écrit » : <https://scripta.psl.eu/presentation/>
- **Transkribus** : Plateforme de reconnaissance de caractère et de transcription de manuscrits ou d'imprimés : <https://transkribus.eu/Transkribus>
- **VIAF** : Fichier d'autorité international de référence (noms de personne, collectivités, noms géographiques, œuvres et expressions : <http://viaf.org>
- **WordPress** : Créer un site Web en quelques minutes : <https://fr.wordpress.com>
- **W3C** : Consortium (W3C), communauté internationale pour développer les standards du Web : <https://www.w3.org/>
- **Zotero** : Logiciel de gestion de références bibliographiques gratuit et open source : <https://www.zotero.org/>

Contributeurs par ordre alphabétique :

Charlotte Dessaint Bibliothèques Ulm / ENS PSL
David Denéchaud CAPHÉS CNRS / ENS
Julie Giovacchini Centre Jean Pépin CNRS / ENS
Marie-Laure Massot CAPHÉS CNRS / ENS (coordinatrice du projet)
Clément Plancq LATTICE CNRS / ENS
Thierry Poibeau, LATTICE CNRS / ENS
Agnès Tricoche AOROC CNRS / ENS
Richard Walter ITEM CNRS / ENS

Partenaires :

