



Schweizerische Gesellschaft für Geschichte
Société suisse d'histoire
Società svizzera di storia
Societad svizra d'istorgia

Standards pour les données numériques de la recherche en histoire

Un Livre blanc de la SSH

Version : 10.11.2023, 1.0

Ce document est issu d'un modèle élaboré par le groupe de travail « Pérennisation des données » du Département « [Recherche fondamentale](#) » de la [Société suisse d'histoire \(SSH\)](#). En faisaient partie (pour certains à des moments différents) Simon Gabay, Tobias Hodel, Moritz Mähr, Stefan Nellen, Barbara Roth-Lochner, Pascale Sutter, Andrea Voellmin et Karin von Wartburg. Le rapport a été développé et discuté au cours de divers workshops et a bénéficié de la collaboration d'experts extérieurs, dans le cadre d'un *open peer review*.

Nous remercions chaleureusement l'ensemble des participant·e·s pour leur engagement et pour ces discussions constructives.

Le présent document a été adopté par le Département « Recherche fondamentale » le 2.7.2022 et a été avalisé par le Comité directeur de la SSH le 22.11.2022.

Principes fondamentaux

Le document est interactif et renvoie, via des hyperliens, aux titres traitant de sujets connexes.

Public cible

Le public cible de ce rapport est constitué d'**historien·ne·s** qui travaillent avec des (bases de) données, qui éditent et souhaitent publier en ligne des documents, mais aussi d'**informaticien·ne·s** qui participent à des projets scientifiques en histoire.

Objectif

Le document propose un **glossaire** et un **catalogue** qui répertorient les standards les plus importants permettant l'interopérabilité, afin que les projets d'histoire numériques puissent facilement être exploités à nouveau par des tiers et mis en réseau. Le présent rapport se focalise sur des standards techniques.

La notion d'**interopérabilité** désigne la possibilité d'échanger et de relier des informations et données entre elles par le biais de divers systèmes et prestataires de services. Grâce à l'application de standards et à la transmission de données élaborées par des chercheurs individuels, la réutilisation se voit promue et encouragée.

En outre, le respect des standards rend possible une meilleure cohérence et connectivité avec les travaux de recherche internationaux et transnationaux ; en effet, les données et les résultats des recherches deviennent réutilisables dans divers contextes. Simultanément, l'utilisation de standards facilite l'archivage des bases de données¹.

Le document veut rendre compte de ces standards dans une perspective historique (dans le prolongement de perspectives issues du domaine plus large des humanités) et neutre, sans jugement de valeur sur les technologies présentées.

Le rapport s'articule autour d'une introduction qui présente les notions fondamentales, d'un glossaire qui explicite les concepts les plus courants, et d'un catalogue décrivant les standards actuellement en usage. En raison des évolutions technologiques, ni le glossaire ni le catalogue n'ont de prétention à l'exhaustivité ou à l'immutabilité. Cette version, datée et publiée, est en effet un produit de son temps. Par conséquent, des révisions seront assurément nécessaires afin d'adapter les formats et les technologies décrites.

Le caractère partiel du présent document rappelle, en parallèle, que seule une utilisation active des données et des méthodes numériques rend possibles les discussions productives qui permettent, en définitive, d'aboutir aux progrès de la théorie et de la pratique. En conséquence, nous soutenons logiquement la publication ouverte, offensive, proactive et libre de toutes les données (même « sales »), de même qu'une utilisation critique de celles-ci.

¹ Nous utilisons la notion de « standard » en la distinguant bien de celle de « norme ». Le concept de norme induit pour nous une manière relativement uniforme ou unifiée, reconnue comme telle et génériquement appliquée (ou pour le moins tendant à l'être), visant à établir ou réaliser quelque chose, cette manière de faire normée s'étant imposée par rapport à d'autres usages qui pouvaient être appliqués jusqu'à l'imposition de ladite norme. Une norme se définit donc comme une règle usuelle (et parfois légale) qu'une procédure de normalisation a permis de convertir en règle, généralement applicable et officiellement publiée, régissant un état de fait. Une norme est adoptée et publiée après que toutes les instances impliquées dans le processus de normalisation se sont prononcées. Suivant la Liste des normes EN de l'Union européenne et l'EN 45020, une norme est « un document établi par consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné ». D'autres objectifs peuvent être liés à la volonté d'une normalisation, comme la rationalisation, l'homogénéisation, la compatibilité, la facilité d'usage ou la sécurité : <https://fr.wikipedia.org/wiki/Norme> (état au 17.04.2023).

La notion de standard est plus large et peut être issue d'une régulation formelle ou informelle (au sein d'une norme ou d'une voire plusieurs règles), mais elle peut également apparaître de manière non concertée. De ce point de vue, le concept de standard est le plus courant dans les domaines de la technique et de la méthode. En langue allemande, une confusion est apparue au cours des dernières années, qui tend à comprendre le terme de standard de la même façon que la notion anglaise de *standard*, synonyme à celle de norme. De l'anglais provient également le terme *de-jure-Standard*, qui recouvre la notion alémanique de norme. On parlera de *standards De-facto* ou de *quasi-standards* lorsque, au contraire de la normalisation, des méthodes et des règles s'établissent par la pratique, et non suite à des lois, ordonnances ou accords : https://fr.wikipedia.org/wiki/Norme_et_standard_techniques (état au 17.04.2023).

Introduction

Les pages qui suivent font office d'indications générales à destination de projets d'histoire comportant des aspects digitaux, recourant à des chaînes de traitement orientées vers une question de recherche. Ce guide se présente en deux parties et offre des informations sur les standards et sur les approches pour préparer, gérer et analyser des données. Nous souhaitons ainsi faciliter le dialogue entre historien·ne·s, spécialistes de la gestion des données et informaticien·ne·s.

La première partie explicite des technologies et approches fondamentales pour les humanités numériques en listant diverses définitions. Sous forme de catalogue, la deuxième partie décrit les standards techniques à disposition pour lancer un projet numérique.

Le présent document traite de **standards techniques d'utilisation de données** en relation avec des projets en sciences historiques. Le terme « standard » désigne ici une manière relativement uniforme ou unifiée, largement reconnue et répandue pour produire ou réaliser quelque chose (ou du moins y aspire), qui s'est imposée par rapport à d'autres manières de faire. À ce titre, le fait que ces standards se soient établis *de jure* ou *de facto* importe peu.

Données

Au pluriel, la notion de données dérive du latin *datum*, que l'on peut traduire par « donné ». Ce concept désigne ici des représentations numériques, codées de manière strictement réglementée, au moyen de (chaînes de) caractères qu'on nomme la syntaxe. Comme cette notion est utilisée sous diverses formes et dans divers contextes, il est nécessaire d'apporter quelques précisions.

Par le biais d'enrichissements sémantiques, on transforme les données en informations. La mise en réseau de ces informations conduit à un savoir (historicisable). Les données sont produites et préparées. Par conséquent, la production et la préparation de données sont un processus personnel et subjectif. De ce point de vue, les données sont davantage des *capta* (quelque chose de perçu subjectivement), ou même des *facta* (des constructions), que des *data* (des donnés)².

Le processus de l'enrichissement peut être assimilé à une pyramide du savoir :

- Au moyen de la **syntaxe**, des **signes** deviennent un **énoncé** → Cela crée une **donnée numérisée**
- Une ou plusieurs significations est/sont attribué·e·s aux **données numériques** par ou plusieurs **élément·s sémantique·s** → Cela crée des **informations**
- Les informations sont **reliées** et **évaluées** → Cela donne du **savoir**

Les **données de recherche** sont une forme spécifique de données. Sous cette notion, nous comprenons toute donnée établie au cours d'un processus de recherche, par exemple lors

² Drucker, Johanna : « Humanities Approaches to Graphical Display », in : *Digital Humanities Quarterly* 5 (1), 2011. En ligne : <<http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>>, état : 11.03.2016.

d'une digitalisation, d'un dépouillement de sources, d'une expérience, d'une mesure, d'un inventaire ou d'un questionnaire. La disponibilité et la réutilisation d'informations numériques incluent un accès aussi aisé et gratuit que possible à ces données de recherche. Logiquement, aussi bien les données primaires que les métadonnées sont à inclure de cette catégorie.

Le terme de **données primaires** n'est pas courant dans les sciences historiques, puisqu'il y est plus commun d'y parler de sources, peu importe que celles-ci soient numériques ou analogiques. Par souci de cohérence, les pages suivantes assimilent le terme de données primaires – ou de données source – à la notion de « sources ». Par conséquent, sous cette dernière notion peuvent être compris une image autant qu'une carte, un objet ou un texte. Fondamentalement, il n'existe aucune forme matérielle qui ne puisse être considérée comme source et, par conséquent, tout peut constituer une donnée primaire.

Il est néanmoins important de différencier les sources analogiques des sources numériques. Les données nées numériques (*born-digital*) sont des sources historiques qui ont été produites sous une forme numérique dès l'origine, et dont l'exploitation, dans le cadre d'un projet de recherche, exigera forcément des méthodes numériques³. Au contraire, les données analogiques sont nées sous une forme matérielle, et ne deviennent exploitables électroniquement qu'après avoir fait l'objet d'une conversion numérique (scan, traitement d'image, reconnaissance textuelle).

Les **métadonnées** sont le complément des données primaires. Selon notre définition, elles sont plus larges que les simples métadonnées techniques de documents numériques. Nous entendons, par métadonnées, des descriptions qui peuvent également s'appliquer à des documents analogiques⁴. L'utilisation de standards est ici centrale, car elle rend possible un échange fluide entre plateformes et institutions. Les standards de métadonnées sont des spécifications qui décrivent dans le détail les termes des métadonnées. Ils permettent de définir des éléments, leur structure et leur signification, rendant possible une interopérabilité sémantique entre les diverses applications utilisatrices desdits standards. L'objectif de l'établissement de standards de métadonnées est d'offrir des descriptions uniformes de ressources faciles à intégrer dans une application logicielle, ainsi que de permettre la transmission de ces données entre plateformes et institutions⁵.

Pour mettre en réseau les données et les ressources, les institutions de l'information documentaire (bibliothèques, centres d'archives etc.) emploient des **données descriptives standardisées**. Une donnée standardisée est ainsi une forme de vocabulaire contrôlé définie en amont de l'indexation. En anglais, ces données standardisées sont considérées comme des « notices d'autorité » (*authority files*) et elles servent à identifier une entité (personne, lieu, etc.).

³ Exemples de données nées numériques : jeux vidéos ; procès-verbaux enregistrés sous forme de documents word ; bases de données de personnes recherchées.

⁴ A contrario, la perception des métadonnées dans DaSCH. Un ensemble d'informations de base sur un projet y est entendu comme métadonnées.

⁵ Hartmann, Sarah : « Metadatenstandards ». *Teil 2 der Einführung in die Interoperabilität von Metadaten und Metadatenformaten*, Göttingen 05.11.2010. En ligne : <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Schulungen/InteroperabilitaetVonMetadaten/teil21.pdf?__blob=publicationFile>, état : 11.05.2017. Exemples de standards de métadonnées : → [METS/MODS](#), → [MARCXML](#), → [EAD](#).

Les données peuvent être préparées, traitées, enrichies ou structurées et modélisées. Par **préparation des données**, on entend la description, la transcription ou toute autre forme d'attribution de signification. Sont également inclus dans cette catégorie **l'enrichissement ou la structuration de données**, accomplis en tenant (idéalement) compte des standards. Cette étape a pour but d'enrichir les données par des informations et elle est souvent accomplie en parallèle à la préparation des données.

Puisque les données représentent une composante capitale des projets numériques, les formats de données revêtent une immense importance en vue d'assurer la connectivité et l'échange. Au début de l'ère digitale, les formats de données étaient un obstacle, car on utilisait des formats propriétaires. Depuis quelques années, le monde de la recherche a opté pour des formats libres et/ou multimodaux (PDF → [XML](#)). Ce document décrit principalement des formats et systèmes libres et flexibles, tels que le XML et le RDF.⁶

Choisir des formats de données libres facilite la conservation à long terme et simplifie la réutilisation des données.

Les données ouvertes et structurées sont un pas de plus vers un *World Wide Web* plus intelligent pouvant être qualifié, à la suite de Tim Berners-Lee, de **Web sémantique**. Dans ce réseau, les informations doivent être structurées plus efficacement que jusqu'à présent, afin de permettre une meilleure exploitation numérique des informations. Le traitement automatique des informations doit ainsi être simplifié et amélioré. → RDF, → [RDFS](#) et → [OWL](#) représentent des pans fondamentaux de ce Web sémantique. Les technologies du Web sémantique promettent d'extraire des informations davantage que des extraits de textes ou de parties de textes. Par conséquent, de meilleurs mécanismes d'interrogation et d'évaluation des données pourront être envisageables et avec une puissance supérieure à celle d'aujourd'hui, fondée sur les bases de données relationnelles⁷.

Préparation et (re)présentation des données

La distinction entre la préparation des données (structuration et modélisation des données et saisie) et leur publication est désormais bien établie. Comme les attentes en matière d'affichage de données évoluent constamment, leur présentation nécessite une adaptation constante, tout en ne changeant rien aux méthodes d'alimentation et de structuration des données. À l'inverse, des adaptations du modèle de données ne doivent pas exercer d'influence sur leur restitution.

La séparation stricte entre les données et leur restitution permet de produire des présentations variées, adressées à des publics divers, tout en puisant dans la même base de données⁸.

⁶ Les standards proposés/décrits ne correspondent pas aux prescriptions du [CECO](#) pour une conservation durable des données.

⁷ Kuczera, Andreas : « Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi », *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 05.05.2015, <<http://mittelalter.hypotheses.org/5995>>. Les interrogations sur le mode *Friend of A Friend* (FOAF) ne sont possibles qu'avec des efforts considérables, et sont, en réalité, très rares. Kuczera, Andreas : « Das Deutsche Textarchiv in der Graphenwelt », *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 04.04.2017, <<http://mittelalter.hypotheses.org/10025>>.

⁸ Similaire à : Sahlé, Patrick : *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. Teil 2: Befunde, Theorie und Methodik., Bd. 2 / 3, Norderstedt 2013 (Schriften des IDE 8). En ligne : <<http://kups.ub.uni-koeln.de/5352/>>. Il s'agit là d'éditions numériques :

À l'ère du livre, l'accent était mis sur le rendu visuel des informations (mise en page, tableaux, etc.). Les données numériques se distinguent toutefois par leurs interconnexions, qui sont impossibles à restituer visuellement dans leur globalité : chaque présentation n'est qu'une forme possible et partielle de l'ensemble des données.

Malgré toutes les recommandations en faveur de la standardisation et de la structuration, nous insistons sur la nécessité d'une utilisation critique des données numériques et de leurs diverses infrastructures. Les standards et les infrastructures reflètent des hypothèses épistémologiques fondamentales et prévoient des méthodes et pratiques (implicites). Ainsi, il est nécessaire d'interroger les structures de pouvoir susceptibles d'être renforcées ou soutenues par les données, la structuration des données ou par leur transmission. Actuellement, la recherche intersectionnelle au sein des Humanités numériques met en lumière ces problématiques de façon critique⁹.

Glossaire

Ce glossaire définit brièvement les notions principales. Leur pertinence pour la réalisation de projets en histoire est indiquée en italique dans la colonne de droite. Les exemples éventuels sont placés à la suite, surlignés en gris.

pour chaque édition, nous sommes en présence d'un unique mode de représentation/présentation possible.

⁹ Voir particulièrement D'Ignazio, Catherine; Klein, Lauren F.: *Data Feminism*, Cambridge, Massachusetts 2020. En ligne : <<https://mitpressonpubpub.mitpress.mit.edu/data-feminism>>, état : 18.03.2022. Jarrett, Kylie : « The Digital Housewife. Feminism, Labour and Digital Media, 2016 ». Losh, Elizabeth; Wernimont, Jacqueline : *Bodies of Information: Intersectional Feminism and the Digital Humanities*, 2018 (Debates in the Digital Humanities). En ligne : <<https://dhdebates.gc.cuny.edu/projects/bodies-of-information>>, état : 18.03.2022. Russell, Legacy : *Glitch Feminism. A Manifesto*, 2020 und Ortolja-Baird, Alexandra ; Nyhan, Julianne : « Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive », in : *Digital Scholarship in the Humanities*, 19.10.2021, S. fqab065. En ligne : <<https://doi.org/10.1093/lc/fqab065>>, état : 18.03.2022.

<h2>Définitions</h2>	
<h3>Données</h3>	
<p>Les données correspondent à des faits isolés, à des moments, à des indications chronologiques, etc., pouvant être saisies <i>via</i> l'observation, la mesure ou toute autre façon d'enregistrer un fait. Ces données sont représentées par des signes correspondant à des informations.</p> <p>Les données peuvent être saisies sous différentes formes et niveaux de structuration. La définition du terme de structuration peut fortement varier et il n'existe par conséquent pas de séparation stricte entre les divers degrés de structuration des données (typiquement, entre données non-structurées, semi-structurées ou structurées).</p>	
<h3>Données « sales/bruitées » et données « propres »</h3>	
<p>Les données n'émergent pas par génération spontanée, car elles sont produites. Les résultats de tels processus sont souvent ce qu'on désigne par le terme de <i>dirty/noisy data</i>, soit des données « sales » ou « bruitées ». Ces (groupes de) données peuvent aussi s'avérer précieux et exploitables pour la recherche. Des procédés de nettoyage peuvent être utilisés pour convertir les données collectées dans une forme adéquate et pertinente, propice à répondre à une question de recherche. Ces données sont alors appelées des <i>clean data</i>, des « données propres ». Les procédés de nettoyage des données conduisent néanmoins systématiquement à des pertes et à des transformations, ce dont la recherche discute activement¹⁰.</p>	
<h3>Données non-structurées</h3>	
<p>Les données non-structurées se présentent sous une forme non formalisée, telle que des textes issus de la numérisation et de la reconnaissance textuelle, des programmes de traitement de texte ainsi que des données audio, vidéo ou iconographiques « brutes ». Ces données ne disposent pas d'un modèle de métadonnées techniques (création et traitement du fichier). Dans le cadre de la → Préparation des données, différents moyens peuvent</p>	<p><i>Les données non-structurées proviennent souvent de vastes processus de numérisation et forment de riches réservoirs pour la détection</i></p>

¹⁰ Voir particulièrement Rawson, Katie; Muñoz, Trevor: « Chapitre 23 Against Cleaning », in : Gold, Matthew K.; Klein, Lauren F. (éd.) : *Debates in the Digital Humanities* 2019. En ligne : <https://doi.org/10.5749/j.ctvg251hk>, état : 21.03.2022.

<p>être mis en œuvre pour tenter d'octroyer une ou des structures aux données récoltées. Des procédés automatisés ou semi-automatisés (souvent avec un arrière-plan de linguistique informatique) peuvent être utilisés pour faciliter cette structuration.</p>	<p><i>d'images ou de textes. La principale caractéristique de cette forme de données consiste en la quantité, au détriment de la qualité.</i></p>
<p>Les textes lus par OCR se présentent sous une forme peu, voire pas structurée. Les fichiers TXT sont également des extraits de textes non-structurés. Un autre exemple de données non-structurées réside dans les textes issus de Google Books.</p>	
<p>Données semi-structurées ou faiblement structurées</p>	
<p>Les données semi-structurées ne sont pas structurées mais, contrairement aux données non-structurées, elles peuvent contenir soit des métadonnées liées à des images (lieu, date de la prise de vue, etc.) ou à des courriels (destinataire, date, etc.), soit des balises, comme des textes codés en XML ou des données JSON.</p> <p>Entre les données structurées et non-structurées, il existe un vaste champ d'expression. La structure des données peut être définie seulement partiellement, par exemple par des associations clé/valeur, c'est-à-dire que seule la signification de la clé est connue ou alors, dans un tableau, celle de la colonne. Le contenu ne l'est pas, car il se présente sous la forme de texte pur. De telles données sont typiquement des textes encodés en format XML, des tableaux ou des métadonnées de ressources numériques.</p>	
<p>Données semi-structurées : édition des comptes annuels de la ville de Bâle entre 1535 et 1610 (https://gams.uni-graz.at/context:srbas, état : 23.12.2021)</p>	
<p>Données structurées</p>	
<p>Les données structurées sont indépendantes de leur cadre structurel descriptif ; elles relient des entités identifiées de la manière la plus précise possible. Il s'agit typiquement de bases de données relationnelles et de <i>linked data</i> (partiellement basées sur des thesaurus et des → Ontologies).</p>	<p><i>Les données structurées sont faciles à évaluer et à exploiter. La documentation quant à la structuration est fondamentale pour les données structurées, afin de comprendre les choix de structuration et les hypothèses qui les sous-tendent.</i></p>
<p>Données hautement structurées : Répertoire officiel des communes (https://www.bfs.admin.ch/bfs/de/home/grundlagen/agvch.html, état : 23.12.2021)</p>	

<p>Jeu de caractères/Unicode</p>	
<p>Unicode est un standard international qui définit, sur le long terme, un code numérique pour toute police de caractère ou élément de texte, et ce pour toutes les cultures écrites et tous les systèmes de caractères connus. Son but est de surmonter l'obstacle de l'utilisation de codifications diverses et donc incompatibles entre pays et cultures. Unicode est constamment complété avec des caractères provenant d'autres systèmes d'écriture¹¹. Afin de reproduire les caractères spéciaux, tels que les signes combinés, on utilise l'<i>Unicode Normalization Form D</i> (NFD).</p>	<p><i>Une restitution correcte et durable des signes est centrale pour l'édition de texte. Pour permettre les échanges, il convient d'appliquer des standards (universels).</i></p>
<p>Des signes spéciaux manquants peuvent par exemple être encodés avec la MUFI (<i>Medieval Unicode Font Initiative</i>). Il existe beaucoup de polices librement accessibles comme <i>Andron Scriptor</i> https://mufi.info/m.php?p=mufi&i=968. D'autres polices complètes avec une large palette de caractères ainsi que d'autres outils pour l'adaptation des configurations de clavier sont disponibles sur sil.org.</p>	

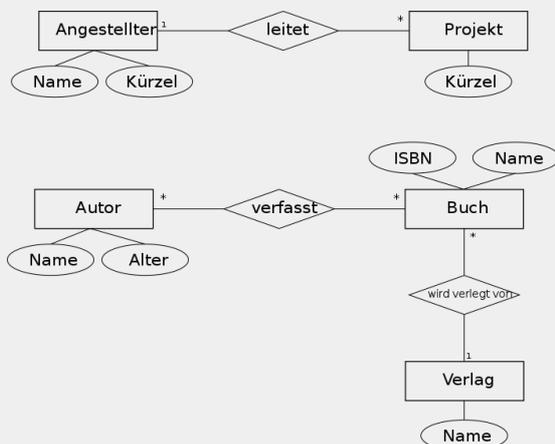
¹¹ Unicode, in : Wikipedia, <<https://fr.wikipedia.org/wiki/Unicode>>, état : 28.04.2017.

Formes d'assignation et de structuration

Modèle de données

Un modèle de données est un modèle formel comportant des données à décrire et à exploiter (par exemple une base de données, une édition ou, dans le cadre d'un projet de recherche, toutes les données pertinentes à collecter) dans un cadre d'utilisation spécifique, les données traitées étant toutes mises en relation¹². Les modèles de données permettent des opérations complexes sur lesdites données. Si elles respectent des normes, ces données permettent un échange ou la mise en relation avec d'autres données ; elles contribuent ainsi à la pérennisation du savoir¹³.

Les modèles de données rendent les questions de recherche opérationnelles. Ils permettent de traiter numériquement les questionnements de la recherche et assurent une haute qualité de données, en définissant les exigences auxquelles les données doivent répondre lors de leur saisie par un tiers.



Source: Körner, Peter: « Zwei (kleine) Beispiele für Entity Relationship Diagramme als SVG », en ligne : <https://de.wikipedia.org/wiki/Datei:Er-diagramm.svg>.

Base de données

Les bases de données et les systèmes de gestion des bases de données sont des systèmes d'administration des données. Suivant les types d'utilisation, leurs fonctions peuvent varier comme la présentation et la sauvegarde efficaces, durables des données. On citera les bases de données relationnelles et les bases de données orientées graphe. Beaucoup de systèmes de données garantissent certaines propriétés (*Atomicity, Consistency, Isolation, Durability*) pour les manipulations de base (*Create, Read, Update, Delete*).

¹² « Modèle de données », in : Wikipedia, , <https://fr.wikipedia.org/wiki/Modèle_de_données>, état : 25.04.2023.

¹³ Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (éd.): *Digital Humanities. Eine Einführung*, Stuttgart 2017, p. 100.

<h2>Banques de données relationnelles</h2>	
<p>Les bases de données relationnelles se caractérisent par le fait qu'elles se composent de tableaux bidimensionnels reliés. Les bases de données relationnelles sont stables et éprouvées depuis plusieurs années mais, contrairement aux bases de données orientées graphe, elles sont moins flexibles lorsqu'un élargissement du modèle est envisagé.</p> <p>Les bases de données relationnelles présentent un fort taux de cohérence des données. Les bases de données relationnelles normalisées permettent une conservation fiable et univoque des données et contribuent ce faisant à la persistance des données. Les bases de données relationnelles offrent en outre des fonctions permettant de définir et de contrôler l'intégrité des données. De nos jours la gestion électronique des données est dominée par le modèle de base de données relationnelle.</p>	<p><i>Les bases de données de recherche sont typiquement orientées vers des questionnements concrets et permettent de renseigner des champs clairement définis. Les capacités de connexion et d'interopérabilité doivent être tout particulièrement prises en compte lors de la mise en place.</i></p>
<p>Le <i>Repertorium Academicum Germanium</i> est un projet de banque de données au long cours (http://www.rag-online.org/, état : 23.12.2021), qui offre de complexes possibilités d'interrogation des données.</p>	
<h2>Banques de données orientées graphe</h2>	
<p>Les banques de données orientées graphe se définissent par la mise en réseau d'objets isolés (et pas par le biais de tableaux).</p> <p>Les informations peuvent être stockées sous forme de graphes, par exemple sous forme de triplets en RDF, c'est-à-dire sous la forme d'équations simples composées de trois parties : sujet-prédicat-objet.</p> <p>Un graphe est donc un ensemble d'énoncés sous forme de liens permettant d'exprimer plusieurs propriétés ou relations avec un objet ou entre des objets.</p> <p>Le RDF (<i>Resource Description Framework</i>) est un modèle adopté par le W3C pour une représentation flexible des données sous forme de triplet (trois unités), composé du sujet (ressource à décrire), du prédicat (propriété) et de l'objet (valeur).¹⁴. Les relations simples entre ressources (liens) peuvent être représentées au moyen du RDF. Les données représentées par le RDF constituent des réseaux. Pour définir et représenter des → vocabulaires et → ontologies, → RDFS (schéma RDF pour des cas peu complexes) ou le langage ontologique du Web (→ OWL pour constituer des vocabulaires complexes) sont disponibles.</p>	<p><i>Le regroupement en graphes permet de réaliser simultanément plusieurs déclarations sur les entités, de manière similaire aux feuilles de données contenues dans les diverses bases de données.</i></p> <p><i>Le RDF permet de représenter de manière flexible des structures de connaissances sans qu'il soit nécessaire de procéder à des normalisations/adaptations à des structures de données prédéfinies (comme dans les bases de données relationnelles).</i></p>

¹⁴ Voir la page web du W3C: <<https://www.w3.org/RDF/>>, état : 21.03.2022. Actuellement, le W3C poursuit le travail sur RDF-star et SPARQL-star. Ainsi, les énoncés peuvent prendre la forme de triplets, pour préciser par exemple les provenances d'informations ou d'annotations. Voir https://w3c.github.io/rdf-star/cg-spec/editors_draft.html, <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-star/>, état : 21.03.2022.

<p>On appelle « triplestore » une base de données conçue pour le stockage des triplets (RDF). Les bases de données orientées graphe remplissent des buts similaires, mais elles ne sont pas spécifiquement destinées aux technologies RDF. La livraison des triplets s'effectue via des langages de requête (par ex. → SPARQL).</p> <p>Les bases de données orientées graphes sont moins performantes que les bases de données relationnelles optimisées et sont par conséquent plus lentes. La modélisation se réalise de manière explicite et est ainsi plus complexe.</p>	
<p>Sujet–prédicat–objet : Personne X – est auteur de – NN. La phrase affirmative est formée ainsi : « Max Frisch » (sujet) « est l'auteur de » (prédicat) « <i>Homo Faber</i> » (objet). Un projet éditorial fondé sur la technologie RDF : Edition Euler de la littérature apocryphe chrétienne : <https://www.unil.ch/irsb/fr/home/menuinst/recherche/editer-la-litterature-apocryphe-chretienne-1.html>, état : 18.03.2022.</p>	
<p>XML (eXtended Markup Language)/Balisage de texte</p>	
<p>XML est un langage de balisage utilisé pour faciliter l'échange de données. Dans le XML, les données peuvent prendre différentes formes, comme des métadonnées (par ex. des informations bibliographiques) mais aussi des données elles-mêmes (par. ex. des éditions).</p> <p>Par balisage de texte, on entend un balisage structuré (souvent selon des schémas) de textes digitaux, généralement sous forme de XML. XML (<i>eXtensible Markup Language</i>) est un métalangage qui permet de définir des langages de balisage et en particulier leurs éléments, leurs attributs et leurs relations respectives.</p> <p>Le format XML permet de structurer des textes, de les enrichir d'informations et de les rendre lisibles automatiquement ; les données restent néanmoins toujours lisibles par des êtres humains. Si le balisage est réalisé selon des standards définis, les documents annotés peuvent être échangés et mis en réseau.</p> <p>Les schémas XML définissent formellement les langages de balisage. Les langages de balisage issus de XML (ou des documents réalisés au moyen de tels langages) sont des hiérarchies ordonnées (structures arborescentes). XML est un langage ouvert, raison pour laquelle les données ne peuvent être comprises et réutilisées que par des schémas ou par l'utilisation de standards.</p>	
<p>Les éditions de texte sont aujourd'hui systématiquement converties en XML et tagués selon les directives de certains langages de balisage (→ TEI, → CEI, → MEI).</p> <p>Structure d'un fichier XML typique encodé en TEI :</p>	

<pre> <TEI> <teiHeader> <!-- Métadonnées--> </teiHeader> <body> <text> <!-- Texte d'édition--> <p> <orgName> <!-- entité balisée--></orgName> </p> </text> </body> </TEI> </pre>	
<p>Schéma</p>	
<p>On appelle schéma la description formelle d'une structure de données. Les schémas définissent les aspects d'une donnée formée dans un langage et comment interagissent les diverses parties d'une → ontologie définie au préalable. « Un schéma permet [...], lors [...] de la saisie de données, d'utiliser des règles communes et d'assurer ainsi que certaines informations soient elles-aussi saisies, ou alors que des structures particulières soient exclues. »¹⁵</p> <p>Les schémas permettent l'→ interopérabilité.</p> <p>Les schémas peuvent être appliqués avec XML et RDF.</p> <p>Les schémas confirment le fait que les données sont bien formées (<i>well formed</i>), mais pas forcément qu'elles soient valides. Cela revient à dire que le schéma ne permet pas de contrôler que les données sont conformes aux règles du modèle (par exemple la sémantique en usage dans le TEI).</p>	<p><i>Seule l'utilisation de schémas permet d'appréhender (de manière automatisée) quelles définitions et combinaisons de définitions sont permises et, partant, possibles.</i></p>
<p>Schéma XML</p>	
<p>Pour le contrôle et la validation des fichiers XML, il est possible de les relier à des schémas. Cela permet de faciliter l'échange et l'interprétation des données en n'autorisant certaines balises qu'en des points clairement définis. Les schémas sont restrictifs et constituent la syntaxe des codes XML. Les schémas facilitent ainsi la saisie de données, les éditeurs XML offrant des valeurs autorisées en fonction du contexte.</p>	

¹⁵ Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (éd.): *Digital Humanities. Eine Einführung*, Stuttgart 2017, p. 135.

<p>Il est nécessaire de distinguer les langages permettant la validation de la grammaire (Schéma XML, Relax NG) et le Schématron, ce dernier étant un complément de ladite grammaire.</p> <p>Dans le cas du TEI, le schéma est automatiquement généré à partir d'un ODD (<i>One Document Does it all</i>), permettant de lui associer une documentation rédigée en langage naturel.</p>	
<p>Schéma RDF</p>	
<p>Le RDFS (<i>RDF Schema</i>) est une extension du → RDF pour créer des → vocabulaires (simples) et des → ontologies. Des classes (types) de ressources peuvent ainsi être définies, ainsi que leurs caractéristiques potentielles. La notation est basée sur le RDF¹⁶.</p>	
<p>Dans le contexte des métadonnées, l'utilisation de schémas facilite l'échange automatique entre les systèmes avec un effort minimum¹⁷.</p>	
<p>Ontologie</p>	
<p>Les ontologies sont des données structurées qui, au moyen d'un modèle de données clairement défini et documenté, représentent le discours historique sur la réalité, c'est-à-dire des informations sur des entités (lieux, concepts, métiers, etc.), leurs caractéristiques et leurs relations (équivalences, hiérarchies, etc.), y compris les métadonnées relatives à leur définition et à leur création.</p> <p>Des informations/caractéristiques/rerelations isolées peuvent être basées sur des → vocabulaires ou des → Thesaurus.</p>	<p><i>Grâce aux ontologies, les informations sont explicites (sans hypothèses implicites), unifiées et exploitables numériquement.</i></p>
<p>Vocabulaire</p>	
<p>Un vocabulaire (contrôlé) est une liste de mots-clés contenant un certain nombre de mots déjà définis ou qui ont déjà été définis au moment du commencement du projet.</p>	
<p>Thesaurus</p>	

¹⁶ Voir la définition du W3C : <<https://www.w3.org/TR/rdf-schema/>>, état : 11.05.2017.

¹⁷ Hartmann, Sarah: *Metadatenstandards. Teil 2 der Einführung in die Interoperabilität von Metadaten und Metadatenformaten*, Göttingen 05.11.2010. En ligne : <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Schulungen/InteroperabilitaetVonMetadaten/teil21.pdf?__blob=publicationFile>, état : 11.05.2017.

<p>Les thesaurus sont des vocabulaires contrôlés structurés de telle sorte que les relations hiérarchiques, de même que les autres types de relations, puissent être exprimées. Un thesaurus doit décrire et représenter un domaine thématique.</p>	<p><i>Les thesaurus aident à l'indexation scientifique des textes et permettent la recherche et le balisage au moyen de concepts d'aujourd'hui.</i></p>
<p>Jurivoc : Thesaurus trilingue du Tribunal fédéral et de l'ancien Tribunal fédéral des assurances.</p>	
<h2>Adresse/Adressage</h2>	
<p>Pour interroger les données de manière ciblée, il faut les assortir d'une adresse. En particulier pour les utilisations liées au → Web sémantique, il est nécessaire de pouvoir interroger (durablement) des données. Sur cette base, il est possible d'extraire des contenus de données de manière répétée, pour en enrichir d'autres avec des contenus actualisés. Pour ce faire, les → identifiants, → URI/IRI ou → DOI sont nécessaires. Grâce à un → permalien, un ensemble de données peut être durablement interrogé.</p>	<p><i>Seule l'adressage permet de faire appel à des ressources sur le long terme et de manière ciblée.</i></p>
<p>Dans le cadre du projet → Metagrid, les données personnelles des personnes enregistrées peuvent être échangées entre partenaires grâce à l'adressage.</p>	
<h2>Identifiant/Identifiant pérenne</h2>	
<p>Une chaîne de caractère est considérée comme un instrument d'identification lorsqu'elle est associée à un objet de manière univoque. Afin d'éviter autant que possible les modifications des identifiants, il est recommandé d'utiliser une combinaison aléatoire de chiffres et de lettres, et d'éviter les noms personnels ou autres.</p> <p>Un identifiant pérenne (<i>persistent Identifier</i>) est un identifiant permanent garanti à long terme par l'institution qui le met à disposition.</p>	<p><i>Les informations sur des entités peuvent être adressées et consultées via des identifiants.</i></p>
<p>Le numéro d'identification d'une personne au sein d'une base de données est un identifiant.</p>	
<h2>URL</h2>	
<p>Une URL (<i>Uniform Resource Locator</i>) est une adresse internet qui identifie et localise une ressource telle qu'un site internet et qui en définit généralement également le protocole de contact (souvent → http/https).</p>	

URI/IRI	
Un URI/IRI (<i>Uniform Resource Identifier/Internationalized Resource Identifier</i>) est un identifiant unique pour une ressource. Un URI/IRI peut consister en une URL.	<i>L'attribution d'un URI permet l'adressage, par exemple au moyen du navigateur (browser).</i>
URN	
Un URN (<i>Uniform Resource Name</i>) est un URI de schéma URN. Il s'agit d'un marqueur, pérenne et indépendant de tout lieu, désignant une ressource numérique.	
ARK	
Les clés de ressources archivistiques (<i>Archival Resource Keys, ARK</i>) servent d'identifiants pérennes ou de références stables et fiables pour des objets d'information. Le système ARK est ouvert et toute institution peut attribuer des ARK après s'être enregistrée et avoir obtenu un <i>Name Assigning Authority Number (NAAN)</i> .	
DOI	
DOI (<i>Digital Document Object Identifier</i>) est un identifiant numérique durable et unique pour les objets physiques, numériques ou abstraits. Il est actuellement surtout utilisé pour désigner les articles en ligne de revues scientifiques et des données de publications (par exemple → ZENODO).	
Alors que les URL et URI peuvent être attribués librement par le webmaster concerné, les DOI ne peuvent être acquis que contre paiement ¹⁸ . En Suisse, les URN sont attribués par la Bibliothèque nationale.	
Lien/Permalien	
Un lien est la forme abrégée d'un hyperlien, soit une référence à un hypertexte (voir aussi → URL).	<i>Les permaliens garantissent qu'une information puisse être retrouvée dans la durée.</i>

<p>Un permalien désigne une → URL qui reste permanente. Un objet numérique peut ainsi être cité, car le lien renvoie toujours au document/à la ressource.</p>	
<h2>Échange de données/communication/mise en réseau/identification</h2>	
<h3>Extraction de données et interfaces</h3>	
<p>Les offres d'échanges de données sont fondamentales afin de mettre ses propres données, textes et documents à disposition du public. Différentes possibilités existent pour cela : par le moyen de sites internet, de téléchargement (FTP) ou, dans l'idéal, par le partage libre via des → interfaces (→ RESTful-API/ SPARQL, SPARQL Protocol and RDF Query Language).</p> <p>Le partage via interfaces permet les requêtes en direct. Lors de requêtes via des interfaces, les données ne peuvent être modifiées/adaptées que si des autorisations sont accordées.</p>	<p><i>Les interfaces standardisées permettent d'agrèger des données et des métadonnées provenant de différentes sources/de différents fournisseurs</i></p>
<h3>HTTP</h3>	
<p>Le <i>Hypertext-Transfer-Protocol</i> (HTTP) est un protocole d'échange des données ; il est utilisé principalement pour charger des pages web par le biais d'un navigateur web. Des données peuvent être envoyées via ce protocole. Le HTTPS consiste en une version cryptée du protocole afin d'assurer une circulation des données plus sécurisée.</p>	
<h3>API/RESTful-API</h3>	
<p>Une interface de programmation (en anglais <i>application programming interface</i>, soit « interface de programmation d'application »), plus souvent désignée par l'acronyme API, est une partie de programme mise à disposition par un système logiciel pour que d'autres programmes puissent s'y connecter¹⁹.</p> <p>Une RESTful-API (<i>Representational State Transfer Application Programming Interface</i>) est une interface basée sur les services web permettant l'accès à une base de données via une application telle qu'un navigateur web. Les requêtes à une RESTful-API se font via diverses méthodes, des http sur des URL avec des segments de chemin prédéfinis (GET, DELETE etc.). Ainsi,</p>	<p><i>REST permet l'accès et l'adaptation de ressources clairement identifiées (par exemples via des URI), mais aussi des requêtes pouvant être transmises via l'URL.</i></p>

¹⁹ Voir « Interface de programmation », dans Wikipedia, https://fr.wikipedia.org/wiki/Interface_de_programmation, état 19.04.2023.

<p>des données peuvent être lues, créées, modifiées ou supprimées. Les réponses peuvent être lues en JSON et → XML.²⁰</p> <p>Il est possible d'extraire les données/informations souhaitées en fonction des requêtes, à condition que les données soient préalablement structurées selon les directives et qu'elles puissent être transmises en conséquence et soumises à une évaluation.</p>	
<p>SPARQL</p>	
<p>SPARQL est un langage de requête pour les banques de données RDF où des requêtes peuvent être formulées auprès d'un → Triplestore (SPARQL-Endpoint) afin de rechercher de manière ciblée des triplets. SPARQL peut aussi être utilisé pour créer des données (triplets) dans un Triplestore (SPARUL: SPARQL/Update).²¹</p> <p>Les résultats de l'interrogation sont fournis sous forme de fichier XML ou Turtle avec une ligne par résultat.</p>	<p><i>SPARQL permet une extraction ciblée d'informations sous forme triplets.</i></p>
<p>Exportation et échange de données</p> <p>Open Access/Open Source/Open Data</p>	
<p>L'Open Access signifie l'accès électronique libre, permanent et gratuit aux publications scientifiques ou de tout autre type²².</p> <p>Pour assurer un accès durable et sans obstacle à la base de données (indépendamment du fait de savoir s'il s'agit de travaux scientifiques ou de logiciels), un grand nombre d'institutions renommées suivent des stratégies d'Open Access/Open Source²³.</p> <p>Les données disponibles et utilisables sans restriction sont nommées des Open Data. Le modèle de déploiement en cinq étoiles de Tim Berners-Lee définit les divers niveaux de données ouvertes : ils s'étalent de la première étoile (licence ouverte) aux cinq étoiles (licence ouverte, format structuré et ouvert/indépendant d'une propriété, données liées à des → URI)²⁴.</p>	<p><i>Permet la réutilisation libre de travaux de recherche et d'indexation.</i></p> <p><i>La publication du code et des textes autorise une réutilisation simple et correcte. Les développements subséquents peuvent se baser sur des savoirs/systèmes déjà existants.</i></p>

²⁰ Voir une introduction poussée : Rodriguez, Alex: « RESTful Web services. The basics », in : *IBM. Developer Works*, 09.02.2015, <<http://www.ibm.com/developerworks/library/ws-restful/index.html>>, état : 19.04.2017.

²¹ Voir pour davantage de documentation : W3C SPARQL Working Group: SPARQL 1.1 Overview, W3C, 21.03.2013, <<https://www.w3.org/TR/sparql11-overview/>>, état : 19.04.2017.

²² Selon l'ASSH, voir un glossaire sur cette thématique ici : <<https://www.sagw.ch/sagw/sagw/themen/wissenschaftskulturen/open-science/open-access/open-access-strategie-der-sagw>>, état : 18.03.2022.

²³ Voir également la déclaration de Berlin sur l'Open Access : *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, 22.10.2003, <<https://openaccess.mpg.de/Berliner-Erklaerung>>, état : 19.04.2017.

²⁴ Cf. <<http://5stardata.info/de/>>, état : 23.12.2021.

<p>Le mouvement <i>Open Data</i> (en Suisse : https://opendata.ch) estime qu'un accès libre aux données renforce la transparence (dans la recherche et l'administration) ainsi que l'efficacité et l'innovation²⁵.</p>	
<p>Open Data Le portail des données ouvertes de l'administration en Suisse, hébergé par les Archives fédérales : <https://opendata.swiss/>, état : 23.12.2021.</p>	
<p>Linked (Open) Data (LD/LOD)</p>	
<p>Le Linked Open Data (LOD) désigne des données structurées et mises en réseau, librement disponibles sur le web, pouvant être identifiées ou déréférencées au moyen de → URIs/IRIs²⁶. Aujourd'hui, le LOD est le plus souvent mis en œuvre au moyen de technologies du → Web Sémantique telles que le → RDF (à condition d'utiliser des → Ontologies ; requêtes via → SPARQL endpoint).</p>	<p><i>Le Linked Open Data permet la mise en réseau avec des données déjà existantes et se trouvant ailleurs, autorisant ainsi l'utilisation d'un savoir déjà constitué.</i></p>
<p>swissTopo propose des informations géographiques sur les communes ainsi que d'autres caractéristiques géographiques sous forme de <i>Linked Open Data</i> : <https://www.geo.admin.ch/en/geo-services/geo-services/linkedata.html>, état : 23.12.2021.</p>	
<p>Wikidata</p>	
<p><i>Wikidata</i> est le nom d'une base de données libre et collaborative qui a notamment pour but de soutenir <i>Wikipedia</i>. Le projet a été lancé par <i>Wikimedia</i> Allemagne et fournit certains types de données comme source commune destinée à d'autres projets <i>Wikimedia</i>, tels que les dates de naissance ou toute autre donnée d'intérêt général, potentiellement réutilisable dans tous les articles liés à des projets <i>Wikimedia</i>²⁷.</p>	<p><i>Wikidata met à disposition des Linked Open Data gérées par la communauté Wikipedia.</i></p>
<p>Henker (trad. « bourreau ») comme nom de famille ou comme concept, voir <https://www.wikidata.org/wiki/Q37547302>, état : 23.12.2021.</p>	
<p>Données (et code) pérennes</p>	
<p>Souvent, les données (et codes) numériques ne sont « que » des outils de travail ou des résultats intermédiaires, étant de ce fait de nature temporaire. Au contraire du format papier, de nombreux formats numériques ne sont pas lisibles ou reconstituables sans prendre en compte l'environnement dans lequel ils ont été créés. Afin d'être durablement réutilisables, les données numériques et leurs restitutions, doivent être enregistrées sous des formats libres et ouverts.</p>	
<p>Des principes pour des données durables ont par exemple été développées dans le cadre du Projet Endings.</p>	

²⁵ Voir le site internet : <<https://opendata.ch/>>, état : 12.5.2017.

²⁶ Voir le site internet : <<http://linkeddata.org/>>, état : 12.05.2017.

²⁷ Wikidata, dans Wikipedia, o.D., < <https://fr.wikipedia.org/wiki/Wikidata> >, état : 19.04.2023.

Catalogue des standards utilisés dans les sciences historiques

Les institutions mémorielles ont défini des standards pour la saisie de (méta)données afin de promouvoir l'échange de données. La connaissance des principaux standards facilite l'interprétation des métadonnées des institutions mémorielles, en particulier lorsque les données sont réutilisées (sous une forme numérique).

Il existe des standards dans les domaines de la description formelle, des entrées de recherche normalisées (autorité, fichiers de normes) et des formats de données.

Règles pour la description formelle de documents et d'autres ressources au sein des institutions mémorielles.

<p>Archives</p> <p>Depuis 2005, l'Association des archivistes suisses (AAS-VSA) possède son propre groupe de travail « Normes et Standards », qui se confronte intensivement à ces thématiques. Son activité est documentée sur le site internet du groupe de travail.</p> <p>Des groupes d'experts issus de l'organisation non-gouvernementale « Conseil International des Archives (ICA/CIA) » s'engagent en faveur des standards en matière d'archivage.</p> <p>Malgré ces efforts, il n'existe pas encore de standards contraignants en Suisse pour les descriptions d'archives, ce qui crée des difficultés pour la comparaison d'informations ou leur réutilisation par des machines.</p>	
<p>ISAD(G) = International Standard Archival Description (General)</p>	
<p>ISAD(G) est la norme générale et internationale de description archivistiques du CIA avec des champs obligatoires uniformisés ainsi qu'une hiérarchie sur plusieurs niveaux.</p> <p><i>Comme d'autres standards archivistiques, est en cours de remplacement par le → RIC.</i></p>	<p><i>Reposant sur des pratiques propres à chaque centre d'archives, ISAD(G) n'est que partiellement utile dans l'échange de données.</i></p>
<p>ISAAR(CPF) = International Standard Archival Authority Record for Corporate Bodies, Persons, and Families</p>	
<p>ISAAR(CPF) est le standard international pour les notices d'autorité pour les archives d'organisations ou d'institutions, de personnes et de familles. Il permet une saisie standardisée des données sur les producteurs de documents, décrivant le contexte de production (provenance). Son</p>	<p><i>ISAAR(CPF) permet l'accès à l'ensemble des documents versés par un producteur d'archives.</i></p>

<p>utilisation est moins répandue que celle d'ISAD(G), car la saisie demande davantage de temps. <i>Est actuellement en cours de remplacement, comme d'autres standards archivistiques, par le → RIC.</i></p>	
<p>RiC = Records in Context</p>	
<p>RiC en cours de développement. Ce standard se veut une nouvelle norme pour la description globale des archives et pour les relations qu'elles entretiennent entre elles. RiC entreprend de résumer en <i>un seul</i> standard l'ensemble des standards archivistiques (ISAD[G], ISAAR[CPF], ISDF, ISDIAH), au moyen d'une ontologie.</p> <p>Voir aussi → RiC-O, → EAD.</p>	<p><i>Le RiC souhaite utiliser les possibilités offertes par le Linked Open Data et révéler ainsi diverses relations et liaisons entre archives.</i></p>
<p>Bibliothèques</p> <p>Bibliosuisse (https://bibliosuisse.ch/), anciennement Bibliothèque Information Suisse, est l'association nationale des bibliothèques, des centres d'information et de leurs collaborateurs. Au niveau international, l'IFLA (International Federation of Library Associations and Institutions) représente les intérêts des services d'information. L'IFLA a édité de nombreux standards dans le domaine des sciences de l'information.</p> <p>Les métadonnées des bibliothèques, en particulier les notices bibliographiques, sont souvent librement accessibles et peuvent être utilisées sans restriction. Les catalogues sont saisis au format standardisé MARC et contiennent souvent des références aux données normatives.</p>	
<p>Les métadonnées du catalogue numérique de la Bibliothèque nationale suisse, de <i>Helveticat</i> et de la Bibliographie de l'histoire suisse (BHS) sont disponibles au format MARC21 par le biais du logiciel Z39.50 et sur demande via OAI-PMH. Une sélection des données peut être téléchargée dans le portail opendata.swiss.</p>	
<p>ISBD</p>	
<p>L'ISBD (<i>International Standard Bibliographic Description</i>, « Description bibliographique internationale normalisée ») est un standard élaboré par l'IFLA. L'ISBD permet une description uniformisée de diverses ressources bibliothéconomiques et autorise l'échange de données.</p> <p>L'ISBD définit quels éléments doivent figurer dans la description d'un document. Il établit la séquence des éléments dans la description et définit leur ponctuation (codage). L'ISBD est en cours de révision. À l'avenir, il devra répondre aux exigences induites par le <i>Linked Data</i>.</p>	<p><i>La ponctuation entre les éléments tels que le titre principal et le sous-titre sont également fournis lors de l'exportation du fichier et doivent par conséquent être supprimés après chaque réutilisation.</i></p>
<p>Exemple d'une zone ISBD avec titre formel et indication de l'auteur :</p>	

<p>Titre formel : premier complément : complément suivant / indication de l'auteur Zwingli : widerständiger Geist mit politischem Instinkt / Franz Rueb ; Bilder : Hans Rueb La notice bibliographique dans la BHS se trouve ici.</p>	
AACR2	
<p>Les Anglo-American Cataloguing Rules sont des règles pour la description formelle d'œuvres. Elles ont été conçues dans les années 1960 et ont souvent été révisées. Elles formaient le corpus de règles le plus répandu sur le plan international, ayant été adoptées par plus de 30 États, dont la Suisse²⁸. AACR2 sont désormais obsolètes et remplacées par le →RDA.</p>	<p><i>Les règles de catalogage déterminent comment certains éléments formels, tels que le titre, les auteurs ou la collection doivent être décrits.</i></p>
<p>Pour les travaux comportant plus de 3 auteur·e·s, toutes les personnes à part la première sont supprimées. Pour l'indiquer, le texte « et al. » est ajouté. En outre, plusieurs notions itératives sont systématiquement raccourcies avec l'AACR2.</p>	
RDA	
<p>Le code <i>Resource Description and Access</i> (RDA) est le successeur de l'AACR2. Il a été introduit en 2010. Une édition révisée a été publiée en 2021. Le RDA se base sur le modèle de données → IFLA LRM et sur les principes internationaux de catalogage (ICP). Initialement conçu pour une utilisation par les bibliothèques, ce code est également adapté à un usage plus large dans des domaines tels que les musées ou les archives. Le texte de codage est régulièrement actualisé et exclusivement publié en ligne, dans le «RDA Toolkit». L'utilisation du «RDA Toolkit» est payante. Le RDA a été introduit dans l'espace francophone en 2010. Les principaux catalogues de bibliothèques suisses (SLSP, BN etc.) utilisent le RDA²⁹.</p>	
<p>Pour décrire chaque œuvre, la transcription constitue une caractéristique importante du RDA. Au contraire des AACR2, les informations contenues sont transmises par le RDA telles qu'elles figurent dans l'œuvre elle-même. Les indications relatives aux auteur·e·s, par exemple, – si elles figurent dans le document –, ainsi que les titres et les universités partenaires, de même que l'institut de rattachement sont relevés. Exemple d'une saisie mot à mot dans <i>Helveticat</i> : Prof. Dr. Andreas Furrer (RA, Universität Luzern), Prof. Dr. Markus Müller-Chen (RA, Universität St. Gallen).</p>	
RSWK	
<p>Au sein des bibliothèques scientifiques et universitaires, des bibliothèques publiques ou des bibliothèques spécialisées d'Allemagne, d'Autriche et de</p>	<p><i>De plus en plus de bibliothèques ont</i></p>

²⁸ Anglo-American Cataloguing Rules, dans : Wikipedia, < https://en.wikipedia.org/wiki/Anglo-American_Cataloguing_Rules>, état : 19.04.2023.

²⁹ Des informations complémentaires sous : <https://www.bnf.fr/fr/code-de-catalogage-rda> >, état : 19.04.2023.

<p>Suisse alémanique, le RSWK – <i>Regeln für den Schlagwortkatalog</i>, servent d’outil pour la description des contenus intellectuels. Elles définissent l’utilisation des mots-clés mobilisés en les tirant de la <i>Gemeinsame Normdatei</i> (GND), gérée par la Bibliothèque nationale allemande.</p> <p>Les RSWK sont un langage d’indexation précoordonné: Si un mot-clé ne suffit pas à la description d’un objet, les termes sont assemblés pour constituer une chaîne .</p> <p>En Suisse romande, c’est → RAMEAU, géré par la Bibliothèque nationale de France, qui sert de référence pour l’indexation.</p>	<p><i>abandonné l’utilisation de chaînes de mots-clés clés et utilisent les termes GND comme mots-clés individuels.</i></p>
<p>Le livre de Thomas Maissen <i>Verweigerte Erinnerung. Nachrichtenlose Vermögen und die Schweizer Weltkriegsdebatte 1989–2004</i> a été assorti des mots clé suivants : Suisse – neutralité – Guerre mondiale (1939–1945) – Traitement de la mémoire – Histoire 1989–2004.</p> <p>La dimension temporelle n’est pas interrogeable avec RSKW (Mot-clé temporel : « Geschichte 1989–2004 »). Si le sujet d’un travail est limité dans le temps, le mot-clé « histoire » est saisi avec la durée exacte.</p>	
<p>FRBR/FRBRoo</p>	
<p>Les <i>Functional Requirements for Bibliographic Records</i> (FRBR; français «Fonctionnalités requises des notices bibliographiques») sont un modèle de données dévolu aux métadonnées bibliographiques. Elles constituent aujourd’hui la plus importante base théorique pour l’établissement de règles de catalogage de bibliothèques³⁰. L’Ontologie FRBR se base sur le modèle de relations entre entités et définit une série de concepts fondamentaux pour le catalogage.</p> <p>Les RDA se basent sur le <i>IFLA Library Reference Model</i> (IFLA LRM). Le modèle FRBR a été remplacé par le → IFLA LRM.</p>	<p><i>Alors que les règles traditionnelles se concentrent surtout sur les détails de la description et que les concepts essentiels (p. ex. "œuvre", "édition", "exemplaire") sont considérés comme acquis, le FRBR tente précisément de définir ces concepts et de les mettre en relation.</i></p>
<p>Avec les FRBR, les diverses rééditions et traductions d’une œuvre sont indiquées, permettant un meilleur aperçu par les utilisateurs et utilisatrices. Exemple: Dürrenmatt, Friedrich : <i>Das Versprechen</i> (Übersetzung mit Autoritätseintrag für das Werk).</p>	
<p>IFLA LRM</p>	
<p>Le modèle de référence <i>IFLA Library Reference Model</i> (IFLA LRM) est un modèle conceptuel qui couvre tous les aspects de la saisie bibliographique. En outre, il favorise la mise en relation des données bibliographiques au moyen des <i>Linked Data</i>. Il développe les modèles précédents tels que les <i>Functional Requirements for Bibliographic Records</i> (→FRBR), <i>Functional Requirements for Authority Data</i> (FRAD) et <i>Functional Requirements for Subject Authority Data</i> (FRSAD).³¹</p>	

³⁰ Functional Requirements for Bibliographic Records, in : Wikipedia, <https://fr.wikipedia.org/wiki/Sp%C3%A9cifications_fonctionnelles_des_notices_bibliographiques>, état : 26.04.2023.

³¹ <<https://www.ifla.org/resources/?oPubId=11412>>, état : 21.03.2022.

Musées et collections	
En raison de l'hétérogénéité des collections et des musées, une uniformisation des standards est beaucoup plus complexe. Par conséquent, seuls quelques thesaurus, vocabulaires et standards de métadonnées seront nommés ici.	
Vocabulaires Getty : AAT, TGN, ULAN, CONA, IA	
<p>Les vocabulaires Getty décrivent divers aspects des objets, artistes et lieux pouvant typiquement être rencontrés dans des musées. Tous ces vocabulaires sont structurés et non exhaustifs.</p> <p>AAT = <i>Arts and Architecture Thesaurus</i> TGN = <i>Thesaurus of Geographic Names</i> ULAN = <i>Union List of Artist Names</i> CONA = <i>Cultural Objects Name Authority</i> IA = <i>Iconography Authority</i></p>	
LIDO = Lightweight Information Describing Objects	
LIDO est un format de métadonnées pour la mise à disposition de données d'objets de musées sur des portails Internet.	
CIDOC-CRM (CIDOC Conceptual Reference Model)	
<p>CIDOC-CRM est un outil théorique et pratique pour l'ajout d'informations dans le domaine du patrimoine culturel.</p> <p>Il est destiné à servir de langue commune aux spécialistes afin de formuler des requêtes auprès de leurs systèmes d'information, mais aussi en tant que fil rouge pour une bonne pratique quant à la modélisation conceptuelle des données. De cette manière, cet outil offre un « liant sémantique » nécessaire à la bonne transmission de diverses sources d'informations issues du patrimoine culturel, telles que les informations publiées par des musées, bibliothèques et archives.</p>	
<h2>Notices d'autorité</h2>	
<p>Les normes d'autorité poursuivent les objectifs suivants :</p> <ul style="list-style-type: none"> ● Identification univoque des entités, ● Fusion de synonymes et de variantes orthographiques, ● Différenciation des homonymes, ● Accès uniforme au moyen d'une forme de rattachement (labels/heading/descripteur). Les données normalisées sont des répertoires de termes normalisés utilisables pour une indexation formelle, mais aussi thématique. 	

Les fichiers de normes créés par des bibliothèques ne font figurer que les noms et les notions nécessaires à la description des œuvres disponibles dans les bibliothèques et les institutions d'archives (les *literary warrants*). Les fichiers normés ne sont, de ce fait, pas des thesaurus qui couvrent l'ensemble d'un champ de savoir.

Les fichiers de normes bibliothécaires les plus répandus en Suisse et les plus importants sur un plan international sont décrites ci-après.

GND	
<p>Le <i>Gemeinsame Normdatei</i> (GND) est un fichier normalisé pour des noms de famille ou de personnes, de collectivités de conférences, d'indications géographiques, de mots-clés et de titres servant en premier lieu au catalogage de la littérature dans les bibliothèques ; son usage au sein d'institutions archivistiques, de musées, de projets ou sites internet est en plein essor³². Le Lobid GND offre une recherche avancée. Il est géré en commun par la Bibliothèque nationale allemande, par toutes des associations de bibliothèques germanophones et d'autres institutions affiliées.</p> <p>Les différentes entités peuvent être interrogées séparément au moyen de l'identifiant de fonds partiel.</p> <p>Les règles générales pour les notices d'autorités reposent, d'une part, sur le <i>Resource Description and Access</i> (→RDA) pour les entités utilisées dans un but d'indexation formelle et de contenu, comme les personnes et les collectivités. D'autre part, les « règles allemandes pour le catalogage par mot-clés » (→ RSWK) définissent l'établissement de notices d'autorité pour les entités utilisées uniquement pour l'indexation de contenu, comme les mots-matières.</p> <p>Toute personne intéressée peut, grâce à la licence CC0 1.0, disposer gratuitement des notices d'autorité GND aux formats MARC 21 Authority, MARCXML et RDFxml par le biais de divers canaux.</p>	<p><i>La collaboration au sein du GND s'opère soit par le biais de réseaux, par la Bibliothèque nationale suisse ou après une concertation directe avec la Bibliothèque nationale allemande. Des projets numériques en histoire (tels que le DHS) peuvent, après une formation, être habilités à ajouter des entrées concernant des personnes.</i></p>
<p>Afin d'ajouter une nouvelle entrée, certains critères se doivent d'être respectés dans le but de réaliser une identification univoque. Dans le cas de noms de personnes, ces critères se composent du nom ainsi que des dates de vie, ou alors du domaine d'activité spécifique. Les notices d'autorité peuvent aussi, sans problème, être complétées à partir d'informations déjà structurées et liées dans le GND. Voir par exemple Albert Einstein ou Wilhelm Tell.</p>	
RAMEAU	
RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) est un langage d'indexation matière. Ce langage est utilisé en	

³²<https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html>, état : 18.03.2022.

France par la Bibliothèque nationale de France, par les bibliothèques universitaires, par de nombreuses bibliothèques publiques et spécialisées ainsi que par plusieurs institutions privées. Il est aussi utilisé en Suisse romande.	
Exemples dans RAMEAU: Albert Einstein , Guillaume Tell .	
LCSH	
Le fichier de normes Library of Congress Subject Headings (LCSH) est géré par la bibliothèque du Congrès à Washington. Il est utilisé pour l'indexation de contenus bibliographiques dans l'espace anglophone.	<i>Les nouvelles entrées et les modifications du LCSH ne peuvent que s'opérer par «proposals» uniques à la Policy and Standards Division (PSD).</i>
Exemples dans le LCSH: Albert Einstein .	
MACS	
Le projet <i>Multilingual Access to Subjects (MACS)</i> avait pour objectif l'établissement d'un vocabulaire plurilingue de recherche dans les notices d'autorité ³³ . En l'état actuel, les mots-clés du GND, de RAMEAU et du <i>Library of Congress Subject Headings</i> sont liés et interrogeables dans le GND. Pour l'instant, Le GND a renoncé à établir de nouveaux liens. <i>Le projet n'est plus poursuivi.</i>	<i>Seuls les mots-matières sont liés Les Mappings sont librement accessibles sous forme de Linked Data.</i>
GND-Mul	
La Bibliothèque nationale allemande est l'initiatrice du projet GND-mul . Son but est de relier le GND avec des notices d'autorités en d'autres langues, mais aussi avec des thesaurus de mots-matières. Le projet de base doit offrir une référence en matière de structures de données pour tous les types de concordances croisées avec le GND.	

³³ MACS. Multilingual Access to Subjects, in : DNB, 30.07.2020, <https://www.dnb.de/DE/Professionell/Metadatendienst/Metadaten/Voclink/voclink_node.html#doc58264bodyText2>, état : 21.12.2021.

Fichiers normalisés pour des entités isolées (fichiers d'autorité)	
Personnes	
ULAN	
Le Getty Research Institute met librement à disposition sa base de données des artistes : <i>Getty Union List of Artist Names</i> (ULAN).	
GND	
De nombreux noms de personnes (auteur·e·s, personnalités historiques) sont déjà saisis dans le GND . Le GND indique des personnes du monde entier et de toutes les régions linguistiques. Il ne se limite pas aux personnes actives dans l'espace germanophone. .	<i>Très présent dans la saisie de noms de personnes, le GND a déjà quelques doublons pour ce type d'entités. En fonction de l'utilisation des données, il est utile de se limiter au niveau de catégorisation Tp3 ou Tp1.</i>
ISNI (International Standard Name Identifier)	
ISNI est un méta-standard géré par l' <i>ISNI International Agency Ltd</i> . INSI alimente sa base de données à partir de systèmes d'identifiants préexistants et, d'autre part, à partir des diverses agences ISNI. Les données de normalisation des personnes proviennent actuellement de 52 sources, dont le GND. ISNI permet une identification univoque, au niveau mondial, des auteurs et de tous les acteurs (scientifiques, inventeurs, graphistes, artistes, interprètes, producteurs, éditeurs, traducteurs, etc.) impliqués dans la publication d'une œuvre. Les organisations, les groupes et les collectivités peuvent également obtenir un ISNI. Les variantes de noms sont également enregistrées. L'attribution des ISNI pour la Suisse est du ressort de MVB (https://german-isbn.de/isni/die-isni).	
ORCID (Open Researcher and Contributor ID)	
ORCID relie l'identité des scientifiques avec leurs travaux. Grâce à ORCID, des documents et des ensembles de données sont non seulement liées à un nom, mais également à un identifiant attribué à une personne. ORCID n'est pas destiné à l'identification de personnages historiques.	
data.bnf.fr	

<p>Data.bnf.fr est une base de données fédérant les données de la Bibliothèque nationale de France (BnF) au moyen de technologies issues du Web Semantique (→RDF) afin de faciliter leur réutilisation par des tiers. Comme les données bibliothéconomiques, la majorité des personnalités enregistrées sont des auteur·e·s pour lequel·le·s une entrée bibliographique existe. Des liens vers des ressources externes (telles que → VIAF ou wikidata) sont également établis.</p>	
<p>Services de mise en réseau pour les personnes</p>	
<p>VIAF</p>	
<p>Le projet <i>Virtual International Authority File</i> (VIAF) relie les notices d'autorité du GND et d'autres sources au moyen des concordances entre noms de familles afin de réaliser une <i>notice d'autorité virtuelle internationale</i>. Le VIAF n'est pas un fichier normalisé au sens propre, car il n'est pas possible de saisir des notices d'autorité dans le VIAF. Les données sont proposées en ligne et sont disponibles pour des recherches, pour l'échange et le partage de données. Dans un but de catalogage, chaque jeu de données se voit attribué un numéro comme identifiant, à partir duquel un URI peut être généré à fin d'une utilisation en <i>Linked Data</i>.</p>	<p><i>Afin de recevoir un numéro VIAF, une entité doit toujours être saisie dans un fichier normé national participant au programme.</i></p>
<p>Exemple dans VIAF: Albert Einstein.</p>	
<p>Metagrid</p>	
<p>Le service web suisse <i>Metagrid</i> permet la création, la gestion et l'analyse de liens entre des personnes identiques présentes dans différents sites web et bases de données³⁴. Metagrid n'est pas un fichier de normes. Grâce à la recherche Metagrid, il est possible d'effectuer des recherches dans les concordances enregistrées pour des personnes issues des bases de données participant à <i>Metagrid</i>. <i>Metagrid</i> est particulièrement utile lorsqu'aucune entrée GND n'a été créée pour les personnes. L'échange s'effectue de manière fédérée et non hiérarchique.</p>	<p><i>Ce service offre une contribution importante pour la recherche d'informations historiques fiables en l'absence de notice GND sur les personnalités concernées.</i></p>
<p>Quelqu'un souhaite savoir sur quels sites (d'institutions participantes) on peut trouver des informations sur une personnalité historique bien identifiée. Exemple: Friedrich Traugott Wahlen sur Metagrid, relié à (sélection) Documents diplomatiques suisses, Dictionnaire historique de la Suisse, Bibliographie d'histoire suisse, Archives de l'histoire rurale.</p>	

³⁴ <<https://www.metagrid.ch/>>, état : 23.12.2021.

Collectivités/organisations/institutions	
A notre connaissance, il n'existe à ce jour aucune compilation complète des collectivités, organisations et institutions historiques et de leurs fonctions.	
GND	
Plusieurs noms de collectivités sont saisis dans le GND .	<i>Le GND ne fait généralement pas la différence entre collectivité et lieu</i>
Lieux	
Getty Thesaurus of Geographic Names	
Le Thesaurus of Geographic Names (TGN) est une base de données de près de 1'000'000 descriptions d'environ 900'000 localités. Il est édité par le Getty Research Institute . Le thesaurus couvre l'ensemble du globe en hiérarchies actuelles, anciennes, mais aussi géophysiques et géopolitiques . Outre les descriptions, les coordonnées, les types de lieux et diverses autres informations sont enregistrées. Depuis 2014, le Thesaurus est disponible en Linked Open Data .	<i>Le TGN est une collection extrêmement dense, construite pour répondre à des besoins divers, ce qui induit une structuration qui n'est pas toujours satisfaisante.</i>
Geonames	
Geonames est une banque de données permettant l'identification et le référencement d'entités géographiques. Diverses personnes, issues de plusieurs pays, mettent leurs connaissances à disposition de <i>Geonames</i> .	<i>Geonames ne fournit qu'une typologie « plate » des lieux (codes de lettres)</i>
GND	
Le GND propose des noms géographiques. Les relations hiérarchiques entre ces termes géographiques ne sont pas saisies de manière exhaustive. En outre, les références aux lieux historiques indiquent rarement les entités territoriales qui les ont précédées ou qui leur ont succédé.	<i>Les noms géographiques sont marqués dans le GND avec code de sous-stockage Tg1.et peuvent ainsi être extraits. Il n'est parfois pas fait de distinction entre les collectivités</i>

	<i>(communauté bourgeoise) et les lieux, raison pour laquelle le GND ne doit être utilisé qu'avec prudence pour les lieux.</i>
Pleiades	
<i>Pleiades</i> est un répertoire et un graphique collaboratif de sites antiques. Il publie des informations essentielles sur les sites et les espaces antiques et propose des services pour la recherche, la représentation et la réutilisation de ces informations sous licence ouverte ³⁵ .	
data.bnf.fr	
Les ressources proposées par Data.bnf.fr (auteur·e·s, travaux, dates, représentations, périodiques) comportent également des indications sur les lieux. Les données proviennent de → RAMEAU et de la section responsable des cartes et plans. Outre les liens vers les données de Data.bnf.fr, des informations supplémentaires telles que les coordonnées géographiques sont disponibles.	
Ortsnamen.ch	
Le portail des recherches toponymiques en Suisse met à disposition du public les travaux de cet organisme et donne des indications sur l'état de la recherche toponymique en Suisse. La base de données en ligne permet de rechercher un nom de lieu et de l'indiquer sur une carte. Pour chaque occurrence, d'autres informations telles que le dialecte, la phonétique, la description du type de lieu, le canton, la commune, le lieu-dit, les hameaux, l'état des sources, les coordonnées sont fournies.	<i>Avec Ortsnamen.ch, des noms de lieux et de lieux-dits n'étant pas indiqués dans la carte nationale peuvent être localisés.</i>
TOPOterm	
TOPOterm est un service web permettant l'enrichissement des interrogations sur les lieux au moyen de synonymes historiques. Une recherche à l'aide de différentes orthographes produit ainsi des résultats. Un cache permet également de réaliser des requêtes directes dans le navigateur web.	<i>Par le biais d'interface, les données sont disponibles en Opendata.</i>

³⁵ <<https://pleiades.stoa.org/>>, état : 21.03.2022.

SwissNAMES3D/swisstopo

SwissNAMES3D est une banque de données de noms gérée par [swisstopo](#). Elle contient les noms de lieux, de montagnes, fleuves et d'autres localités indiqués sur la carte nationale à l'échelle 1:25'000 (et plus). Les entrées sont en outre regroupées en divers types de lieux tels que montagnes, communes, etc., de manière à pouvoir restreindre une recherche à ces critères.

Dans l'ensemble, la base de données *SwissNames* regroupe les entrées géoréférencées de plus de 190'000 localités. Pour chaque entrée, la position exacte, le canton, la commune et – si c'est opportun – l'altitude, sont indiqués.

Une interrogation à l'aide de *Linked Open Data* peut être réalisée par le biais d'un Endpoint SPARQL, par exemple. <<https://sparql.geo.admin.ch/sparql>>, état : 21.03.2022..

Les lieux peuvent être localisés sur la carte nationale grâce à SwissNAMES3D.

Concepts (mots-clés et unités lexicales)

GND

Le [GND](#) héberge de nombreux mots-clés/termes qui contiennent des dénominations et des relations divergentes avec des quasi-synonymes et des notions apparentées. Il existe en outre des liens vers les jeux de données normalisés équivalents de la → LCSH et de la → RAMEAU ainsi que vers les → notations de la DDC. Une systématique permet un accès à celle du GND. Celle-ci est élargie ou créée selon les besoins, c'est-à-dire qu'elle n'est pas exhaustive.

Les Mappings renvoyant aux mots-clés de LCSH et de RAMEAU ainsi qu'aux notations DDC sont librement accessibles comme Linked Data.

Les termes allemands de Henker ou Nachrichtler sont synonymes, dans le GND, du concept d'[exécuteur des hautes œuvres](#). Dans la systématique du GND, la notion appartient à la classe 7.7b « sentence criminelle » et 9.4ab « métiers, activités, fonctions ; appartenance professionnelle, vision du monde ». Un lien thématique vers la notion d'« exécution » est produit. Il est en outre fait référence à → [RAMEAU](#) « Exécutions capitales et exécuteurs », etc.

Classifications

Outre le catalogage d'objets, les bibliothèques font usage de classifications pour permettre un accès thématique à leurs collections. Les classifications sont généralement structurées hiérarchiquement afin de rendre explicites les relations.

<p>DDC</p> <p>La Classification décimale de Dewey (CDD) est la classification la plus répandue à l'international. Elle permet l'indexation de contenus de livres. Son accès n'est pas libre ; il est soumis à l'acquisition d'une licence.</p> <p>Le →GND indique des liens entre les mots-clés et la notation CDD qui leur correspond.</p>	<p><i>Les Mappings entre les notations CDD et les mots-clés du GND sont librement accessibles en tant que Linked Data.</i></p>
<p>CDU</p> <p>La classification décimale universelle (CDU) est établie à partir de la →CDD. Elle a été développée à la fin du 19e siècle en Belgique comme alternative à la CDD, très américano-centrée. Elle est essentiellement en usage dans les bibliothèques non-anglophones.</p>	<p><i>La CDU est par exemple utilisée par la Bibliothèque de l'EPF.</i></p>
<p>Regensburger Verbundklassifikation (RVK)</p> <p>La Regensburger Verbundklassifikation (RVK) est une classification très répandue en Allemagne, en Autriche, en Italie et en Suisse. Elle est originellement développée comme système de classement mais sert également d'instrument de catalogage.</p>	<p><i>La RVK est peu utilisée dans le milieu universitaire suisse.</i></p>
<p>Ouvrages de références, recueils de vocabulaire, normes linguistiques et dictionnaires</p>	
<p>Vocabulaire juridique (histoire du droit)</p>	
<p>À partir des mots-clés issus des Materien der Polizeyordnungen publiés en langue allemande, un vocabulaire historique des concepts juridiques en plusieurs langues est en cours d'élaboration avec le SKOS.</p>	

<p>CoNLL - Données linguistiques (Conference on Computational Natural Language Learning)</p>	
<p>CoNLL est un congrès annuel visant à traiter de <i>Shared Tasks</i> (« problèmes communs »). La forme des données proposées pour ces fonctions s'est imposée comme standard pour le traitement automatique du langage (NPL). Chaque ligne représente un <i>token</i>, dont les diverses annotations linguistiques sont strictement ordonnées et séparées par des tabulations qui constituent autant de colonnes : ID indique la position dans la phrase, FORM indique le token, LEMMA le lemme, POS la partie de phrase, FEAT sa morphologie, HEAD le parent syntaxique, etc. Les valeurs nulles sont indiquées par un trait de soulignement (<u> </u>). Il existe des recommandations pour la valeur des annotations. Des systèmes de référence et des jeux d'annotation standards existent, mais ils dépendent de la langue annotée.</p>	
<p>Dictionnaires nationaux en Suisse</p>	
<p>Les dictionnaires suivant sont édités sous l'égide de l'ASSH et mis à disposition en ligne, avec diverses fonctions de recherche :</p> <p>Allemand : Dictionnaire suisse-alsacien ou Schweizerisches Idiotikon. Pour les recherches sur la langue allemande, voir également le Wörterbuchnetz.</p> <p>Français : Le Glossaire des patois de la Suisse romande (GPSR) permet des recherches en ligne. Pour des recherches sur la langue française, voir également ATLIF et Lexilogos.</p> <p>Italien : Le Vocabolario dei dialetti della Svizzera italiana (VSI) publiera sa version en ligne à la fin 2022-début 2023.</p> <p>Romanche : Le Dicziunari Rumantsch Grischun (DRG) documente les langues rhéto-romanes depuis le 16e siècle et est consultable par le biais du DRG-online.</p>	
<p>Autres ouvrages de référence enregistrés par le DNB: Liste des ouvrages de référence du DNB</p>	

<h2>Formats de données</h2>	
<p>METS/MODS (Metadata Encoding and Transmission Standard, Metadata Object Description)</p>	
<p>Le Metadata Encoding & Transmission Standard (METS) est un format XML défini à partir d'un Schema XML servant à la description par métadonnées d'objets issus de collections digitales.</p> <p>METS est souvent utilisés en combinaison avec le MODS (Metadata Object Description), une spécification XML qui est également gérée par la bibliothèque du Congrès. Alors que METS enregistre les métadonnées techniques et les voies d'accès aux données, MODS sert à coder les métadonnées descriptives du contenu des objets.</p>	
<p>MARC21</p>	
<p>Les formats MARC sont des standards pour la restitution et l'échange de données lisibles par machines. Le <i>Network Development and MARC Standards Office</i> (NDMSO), soutenu par le MARC Advisory Committee, est responsable de la gestion et du développement de MARC 21. Une variante XML du format est également disponible. MARC21 est en usage dans toutes les grandes bibliothèques de Suisse.</p> <p>L'utilisation et la définition des champs individuels est consultable sous: http://www.loc.gov/marc/bibliographic/</p> <p>Le successeur de MARC21 est BIBFRAME, mais ce format n'est pas encore en utilisation³⁶. Une alternative à MARC est le → MODS, utilisable de manière plus globale.</p>	<p><i>Alors que la norme MARC définit quelle information est inscrite dans chaque champ, les règles de saisie formelles définissent la manière selon laquelle les informations contenues dans les champs seront retransmises.</i></p>
<p>DC</p>	
<p><i>Dublin Core</i> (DC) est une collection de conventions simples et standardisées pour la description de documents et d'autres objets.</p> <p>Au sein de <i>Dublin Core</i>, les Dublin Core Metadata Element Set et les DCMI Metadata Terms sont différenciés. Les premiers désignent 15 termes pour la description de métadonnées. Les deuxièmes visent en</p>	<p><i>Les données structurées en DC dans les en-têtes de pages HTML aident les moteurs de recherche à indexer les pages web, ainsi qu'à les hiérarchiser et les rendre visibles.</i></p>

³⁶ voir <<https://www.loc.gov/bibframe/>>, état : 22.12.2021.

<p>autre à relier ces termes avec des classes et des propriétés, des types de données et des vocabulaires.</p>	
<h2>Langages de balisage XML</h2>	<p><i>Les langages de balisage des textes sont lisibles par les machines et interopérable.</i></p>
<h3>ALTO XML</h3>	
<p>ALTO XML (<i>Analyzed Layout and Text Object</i>) est un standard servant à la description de données textuelles automatiquement reconnues, qui sont souvent utilisées dans le cadre de la reconnaissance optique de caractères.³⁷</p>	
<h3>PageXML</h3>	
<p>Page XML est un standard pour la description de données textuelles reconnues automatiquement, qui sont souvent utilisées dans le cadre de la reconnaissance optique de caractères. Page XML est souvent utilisé pour la création de vérité de terrain (Matériel d'entraînement pour les systèmes de reconnaissance textuelles)³⁸. Au contraire de →ALTO XML, PageXML permet d'annoter une information sémantique en <i>stand-off</i>³⁹.</p>	
<h3>HTML</h3>	
<p>HTML (<i>Hypertext Markup Language</i>) est un Markup (un langage de balisage) pour les pages web servant à définir leur structure (par ex. paragraphes, titres etc.). Le HTML est représenté par les navigateurs web. HTML5 est la version la plus récente de HTML et propose plusieurs éléments interactifs prédéfinis. Le format XHTML (<i>Extensible HyperText Markup Language</i>) représente une variante XML d'hypertexte compatible avec le HTML 4.01, qui a entre-temps perdu de son importance avec l'introduction du HTML5.</p>	

³⁷ <<https://github.com/altotxml/documentation/wiki/Versions>>, état : 21.03.2022.

³⁸ <<https://ocr-d.de/de/qt-guidelines/trans/trPages>>, état : 21.03.2022.

³⁹ <<https://github.com/PRIMA-Research-Lab/PAGE-XMLs>>, état : 21.03.2022.

SIARD	
<p>SIARD (<i>Software Independent Archival of Relational Databases</i>) est un format servant à l'archivage de bases de données relationnelles indépendamment de leur système de gestion. SIARD peut représenter des schémas, des relations, des métadonnées ainsi que le contenu de banques de données au format XML. Étant donné que SIARD est conforme à la norme SQL:2008, il peut archiver la plupart des systèmes de bases de données récemment développés. La conformité aux normes XML et SQL:2008 assure un stockage à long terme, durable et opérable, des bases de données de même que, en option, d'une large palette d'informations supplémentaires sur leur contexte d'utilisation. Logiciel gratuit, facile d'utilisation, SIARD Suite permet un archivage de bases de données et un transfert des enregistrements déjà archivés dans de nouvelles applications de bases de données.</p>	
TEI	
<p>TEI (<i>Text Encoding Initiative</i>) est un consortium qui a développé une norme pour l'encodage électronique de textes. TEI désigne également un langage de balisage ou un ensemble de grammaires (schémas XML) qui peuvent être combinés pour en définir un (désigné ici par TEI/XML). Les éditions numériques qui travaillent avec TEI utilisent en général un schéma TEI pour la saisie structurée des métadonnées archivistiques des sources de même que pour l'annotation critique et sémantique des données primaires. Les directives de transcription et d'édition doivent être documentées de manière exhaustive.</p> <p>Dans le cadre de la Text Encoding Initiative, un grand nombre d'initiatives apparentées sont développées afin d'élargir ou de préciser ce standard. Parmi eux, on citera tout particulièrement correspsearch (<https://correspsearch.net/de/start.html>, état : 21.03.2022), EpiDoc pour l'épigraphe (<https://epidoc.stoa.org/gl/latest/intro-intro-fr.html>, état : 21.03.2022), TEI Lex-0 pour les dictionnaires (<https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>, état : 21.03.2022), Vanilla TEI.⁴⁰</p> <p>CEI (<i>Charters Encoding Initiative</i>), qui s'appuie sur le TEI et l'élargit avec l'annotation de chartes réalisées avec les outils de la diplomatique (par ex. indications du préambule à un document ou de son eschatocole, etc.).</p> <p>MEI (<i>Music Encoding Initiative</i>) s'appuie sur les mêmes fondements et sur la logique du TEI mais représente un langage de balisage spécifique pour l'encodage de musique.</p>	

⁴⁰ TEI XML Vanilla est un nouveau «format minimal», projet porté par l'association [e-editiones](#) et également intégré au [TEI-Publisher](#), voir <<https://e-editiones.org/news/tei-vanilla-a-scoop-of-tei-for-everyone/>>, état 21.03.2022.

<p>Le TEI-Publisher est conçu pour la publication de documents encodés en TEI XML ainsi que pour l'indexation dans un existDB. Le Publisher suit les prescriptions du Single-Source Publishing, qui permet de produire diverses variantes d'un même document.</p>	
<p>Une rature dans un texte imprimé ou manuscrit est signalée par le tag . Pour une machine ou un être humain, cela permet d'indiquer clairement que les signes inclus dans les balises ont été barrés.</p>	
<p>EAD</p>	
<p>EAD (<i>Encoded Archival Description</i>) est un standard de documentation XML édité par la bibliothèque du Congrès pour la description de fonds d'archives et de répertoires dans les archives, musées et bibliothèques⁴¹.</p>	
<p>Ontologies</p>	
<p>OWL</p>	
<p>OWL (<i>Web Ontology Language</i>) est une extension du langage de représentation RDF permettant de constituer des vocabulaires et ontologies plus complexes (inclut le →RDFS). La notation est basée sur le RDF. Grâce à des ontologies, un domaine (ses composantes et leurs relations) peut être formellement décrit. Les ontologies existantes peuvent être réutilisées.</p>	
<p>SKOS</p>	
<p>SKOS (<i>Simple Knowledge Organization Systems</i>) est un langage formel standardisé pour organiser des connaissances, par exemple sous forme de thesaurus.</p>	
<p>CIDOC-CRM</p>	
<p>CIDOC-CRM (<i>Conceptual Reference Model de CIDOC</i>) est une ontologie qui modélise les concepts implicites et explicites ainsi que les relations entre institutions qui conservent et documentent un patrimoine (par ex. des musées). Le Comité International pour la Documentation (CIDOC) est un comité international pour la documentation muséale. Des sections</p>	

⁴¹ Introduction en français : <https://www.ead-bibliotheque.fr/>

spécifiques permettent l'élargissement de CIDOC-CRM à certains domaines, comme FRBR . Dans le cadre du consortium Data for History (http://dataforhistory.org/), une version de CIDOC-CRM adaptée aux besoins des historien·ne·s est en cours de développement.	
DOLCE	
DOLCE (<i>Descriptive Ontology for Linguistic and Cognitive Engineering</i>) est une ontologie permettant de modéliser un discours humain.	
RiC-O	
RiC-O (<i>Records in Context-Ontology</i>) est une Ontologie → OWL pour la description de fonds d'archives et des unités qui en dépendent.	
Schema.org	
Initiative pour une structuration uniforme de données sur des sites web à partir d'une ontologie développée par schema.org. Une ontologie hébergée sur Schema.org est friend-of-a-friend, afin d'exprimer les relations entre personnes	
Ontologies RDF/S-OWL simples : → Dublin Core pour les métadonnées ou → Friend-of-a-friend pour les personnes. Ontologies de domaines : → FRBR ou SPAR-Serie (FRBR réutilisé) Application de ces ontologies : DBPedia , une version de Wikipedia adaptée en RDF, ou Europeana , avec les systèmes d'interrogations idoines.	

Formats d'échange/protocoles d'échange

Web ARChive (WARC)	
Le format d'archive WARC permet de stocker des ressources numériques dans un fichier d'archive agrégé avec les métadonnées correspondantes. Il s'agit d'une évolution du format de fichier ARC de l' Internet Archive et représente le format le plus utilisé et le plus soutenu. Il existe un grand écosystème d'outils d'analyse, par exemple sous https://github.com/iipc/awesome-web-archiving , état : 21.03.2022.	

IIIF	
<p>L'<i>Image Interoperability Format</i> (IIIF) est géré par un consortium et établit un standard pour le transfert de données iconographiques, métadonnées incluses⁴². Le chemin d'accès et les informations sur les métadonnées sont fournis dans des « manifestes ».</p> <p>L'objectif de IIIF est de permettre l'intégration et la réutilisation d'images, de parties d'images, d'audio, de vidéo et de métadonnées sur des pages de tiers, sans qu'il soit nécessaire de dupliquer les fichiers d'images.</p>	

⁴² Voir le site *International Image Interoperability Framework*, o.D., <<https://iiif.io/>>, état : 23.12.2021.

[e-codices](#) reproduit des manuscrit médiévaux et modernes par le biais d'un visualiseur compatible avec un IIIF Viewer et publie tous ses contenus avec des manifestes IIIF (métadonnées et données iconographiques).

Distributed Text Services (DTS)

[DTS](#) est un standard, en cours de développement, qui définit une interface/API permettant de récupérer des textes et des parties de textes. Par analogie avec → [IIIF](#), DTS doit mettre à disposition des textes adressés de manière schématique. Plusieurs points d'accès permettent de s'adresser aux [collections](#), à la [navigation](#) et aux [documents](#).

Canonical Text Services (CTS)

La spécification des services de textes canoniques définit un service de réseau qui permet d'identifier des textes et de citer des fragments de textes en utilisant des notions telles que « œuvre » et « citation » couramment en usage dans les sciences de l'Antiquité et d'autres disciplines.

La page principale de CTS (<http://cite-architecture.org/cts/>) est actuellement hors-ligne. La page sourceforge a été actualisée pour la dernière fois en 2015: <http://cts3.sourceforge.net/>.

La [Perseus Digital Library](#) utilise une architecture CTS. L'*endpoint* Perseus CTS API est disponible sous : <http://cts.perseids.org/api/cts/>. Une interface compatible avec le navigateur peut être acquise ici <http://cts.perseids.org/>.

IPIF/ProsopogrAPhI International Prosopographical Interchange Framework)

Avec [IPIF/ProsopogrAPhI](#), l'Académie autrichienne des sciences propose un standard pour l'échange de données prosopographiques.

Stockage des données	
Git	
<p>Bien connu dans le développement de logiciels, le système de contrôle de version Git offre de la sécurité et de la flexibilité dans l'archivage de codes source et de données.</p> <p>GitLab est une plateforme git <i>Open Source</i> gérée en Suisse par SWITCH et par plusieurs universités (par ex. Genève et Berne). GitHub est un service web privé souvent utilisé.</p> <p>Lors du développement de Git, l'intégrité des codes sources administrés était la principale priorité. Le contenu des fichiers de même que les relations réelles entre données et répertoires, versions, tags et <i>commits</i> sont sécurisés par un algorithme de hachage cryptographiquement sûr (SHA1). Cela protège le code et l'historique des modifications contre toute modification malveillante ou accidentelle et garantit une traçabilité complète de l'historique.</p> <p>La flexibilité représente l'un des principaux objectifs dans le développement de Git. Git est flexible à plusieurs égards: dans la prise en charge de différents types de flux de développement non linéaires, dans son efficacité pour les petits ou les grands projets ou dans sa compatibilité avec plusieurs systèmes et protocoles déjà existants.</p>	
DLCM/OLOS	
<p>À Genève, le <i>Data Life Cycle Management</i> (DLCM) a assuré le développement de OLOS sur mandat de la Conférence suisse des Hautes Écoles. OLOS met à disposition une solution nationale pour le management de données de recherche (dans un sens large) et tout au long de leur cycle de vie.</p>	
DaSCH	
<p>Le <i>Swiss National Data & Service Center for the Humanities</i> (DaSCH) est une infrastructure de recherche nationale qui assure un accès à long terme à des données de recherche complexes dans le domaine des sciences humaines et qui peut également servir d'environnement de travail pour les projets. DaSCH favorise la mise en réseau des données avec d'autres bases de données (<i>Linked Open Data</i>) et crée ainsi une plus-value pour</p>	

<p>la recherche et pour le public intéressé. Déjà en ligne, la DaSCH Service Platform (DSP) permet le stockage, l'échange et le travail sur des sources primaires et des données dans les sciences humaines. Les données peuvent être interrogées et actualisées via des applications web génériques. Chaque objet, dans DSP, possède un identificateur spécifique permanent (ARK) qui permet de référencer des entrées de données individuelles. Dans les prochaines années, la plateforme sera progressivement développée pour élargir ses fonctionnalités.</p>	
<p>ZENODO</p>	
<p>Zenodo est un Repository du CERN cofinancé par l'Union européenne et utile pour le stockage de données scientifiques, de publications et de présentations.</p> <p>Les projets hébergés sur GitHub peuvent être sauvegardés à long terme sur ZENODO. Actuellement, il existe une restriction, car un ensemble de données ZENODO ne peut être alimenté que par un seul compte GitHub, ce qui n'est pas idéal du point de vue de la pérennité des données. Une connexion à Zenodo a également été suggérée il y a longtemps pour le Gitlab, mais rien n'a encore été fait (il existe néanmoins des scripts utiles en début de phase bêta).</p>	