

Glossaire des termes techniques

Version 1 – 20/07/2018

Guide réalisé par Richard Walter et Camille Desiles

Ce glossaire est basé sur la terminologie Omeka. Il a vocation à expliciter les termes techniques utilisés par Omeka ou dans EMAN en précisant à chaque fois la définition usuelle et, au besoin, celle utilisée dans le cadre de la plateforme. En particulier dans Omeka, de nombreuses réalités ont une terminologie fluctuante : nous avons alors sont à chaque fois des renvois vers le terme que nous préconisons.

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

Ce glossaire a pour vocation d'évoluer avec la mise en place de nouvelles fonctionnalités (visualisation, transcription, multilinguisme, etc.). N'hésitez pas à proposer de nouveaux termes à traiter. Il n'intègre pas pour l'instant le vocabulaire éditorial et scientifique (de l'édition critique à l'édition génétique).

Index

Ancre	2	Module.....	6
Annotation.....	2	Moissonnage.....	7
Balise	2	Mots-clés	7
Champ	2	Notice.....	7
CMS	2	Nuage de tags	7
Collection.....	2	Numérisation des images.....	7
Contenu	3	OAI-PMH	7
Creative Commons	3	Omeka.....	7
CSV.....	3	Onglet	8
Dublin Core.....	3	Open Source.....	8
Encodage	3	Page aveugle	8
Enregistrement	4	Page web.....	8
Entrée	4	PDF.....	8
Entrepôt OAI-PMH.....	4	Plateforme	9
Exposition virtuelle.....	4	Plugin	9
Facette	4	Relations	9
Fiche	4	Serveur.....	9
Fichier	4	Signet	9
Formulaire	5	Simple Pages.....	9
Framework.....	4	Site	9
Graphe.....	5	Tags	10
HTML	5	TEI	10
Huma-Num	5	Thème / Template	10
Image.....	5	Transcription	10
Item	5	Type de contenu.....	11
Item Relations.....	6	Unicode / UFT8	11
Item type metadata	6	Valeur	11
JPG/JPEG.....	6	Visualisation par graphe.....	11
Licence / licence libre	6	Wysiwyg	11
Lien hypertexte.....	6	XML.....	12
Métadonnées personnalisées.....	6		

A

Ancre : Une ancre est l'endroit de destination d'un lien interne dans une page web ; cela permet de se rendre directement sur une position précise de la page. Dans Omeka, on les utilise pour faire des notes de bas de page dans les « simples pages ».

Annotation : L'annotation est ce qui se rajoute au texte, elle peut être de nature explicative ou critique. Elle commente et ne décrit pas.

B

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

C

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

Sur Omeka, un **Contenu** est découpé en plusieurs champs qui ont une ou plusieurs **valeurs (entrées)** ; les champs désignent aussi les métadonnées **Dublin Core** comme les **métadonnées personnalisées** ou tout espace à remplir dans un **formulaire** du **site**.

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

Omeka se sert de la technologie CMS pour éditer du contenu élaboré mais c'est une bibliothèque virtuelle et non un CMS comme Wordpress, le plus connu actuellement des CMS.

Collection : Une collection rassemble un ensemble de documents dont la cohérence est fixée par le concepteur de la collection. Dans la terminologie Omeka, une collection rassemble des notices mais peut également rassembler des sous-collections (avec le **module CollectionTree**). La granularité et la composition d'une collection peuvent être très variées. Par exemple, il existe une collection sur l'auteur *Mouloud Feraoun* sur *Espace Afrique-Caraïbe*, une collection *Notes éparses* dans le projet Valery ou une collection par année de publication dans le projet *Aragon et la Presse*, etc.

Sur Omeka, on peut mettre sur une collection des métadonnées **Dublin Core** mais aussi, variante EMAN, des **métadonnées personnalisées**.

Contenu : C'est la structure de base d'Omeka ; par convention nous l'écrivons ici avec un C majuscule. Le Contenu permet de décrire de façon précise l'objet que l'on étudie (exemplaire, facsimilé, édition ...) en tant qu'objet physique ou en tant que ressource numérique. Un contenu est structuré en plusieurs **champs** (métadonnées **Dublin Core** ou **métadonnées personnalisées**) qui peuvent une ou plusieurs **valeurs**. Structure centrale d'un **site** Omeka, le Contenu peut être regroupé en **collection** et avoir des **fichiers**. Il est appelé dans la base de données ou sur certains écrans de **modules** « item ».

Creative Commons : Publiées dès 2002, les licences *Creative Commons* (CC) proposent une solution légale aux personnes souhaitant offrir une autorisation non exclusive de reproduire, distribuer et communiquer une œuvre au public à titre gratuit. Elles permettent de faire apparaître clairement au public les conditions de la licence de distribution et de réutilisation de cette création. Parmi les 6 niveaux de licences *Creative Commons*, la **plateforme** EMAN utilise celle exprimée en français et permettant un partage strictement dans les mêmes conditions : Partage à l'Identique 3.0 (CC BY-SA 3.0 FR).

CSV : Le « Comma-separated values » (CSV) est un format informatique stockant des données sous forme de **valeurs** séparées par un code ; le plus souvent le séparateur est la virgule ou le point-virgule. La représentation usuelle de ces données mémorisées en CSV est le tableau. Ce format ne permet pas d'enrichissement typographique (gras, italique, etc.), il conserve du texte brut. Omeka se sert de ce format pour faire des imports en masse : la tabulation des données en CSV permet de faire une correspondance avec la structuration des informations en **champs**. EMAN a rajouté l'export en CSV. Et il ne faut pas oublier que les contenus doivent être codés en Unicode.

D

Dublin Core : Le Dublin Core (dublincore.org) est une initiative de description de ressources numériques qui propose un socle de 15 éléments. Ces 15 **champs** sont les suivants : Titre, Créateur, Sujet, Description, Source, Éditeur, Date, Couverture, Relation, Format, Langue, Type, Identifiant, Contributeur, Droit. Son intérêt réside dans le fait que c'est devenu un standard international et qu'il permet donc l'échange de données entre différents projets ou outils. Pour plus de précisions sur l'application du Dublin Core, voir la présentation très précise sur le site de la BNF : http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html. Le Dublin Core n'impose pas de compléter tous les **champs** mais pour une meilleure diffusion et pérennité de vos données, il est important d'en remplir le plus possible.

Il peut y avoir un raffinement des 15 **champs** avec l'initiative du Dublin Core qualifié mais avec Omeka et EMAN nous utilisons plutôt en complément des **métadonnées personnalisées**. La **plateforme** EMAN propose une utilisation type de chaque **champ** mais cela est personnalisable en fonction de chaque projet. L'acceptation du **champ** Date par exemple peut varier d'un projet à un autre : date d'écriture, date d'édition, etc.

E

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

Dans le projet EMAN, nous utilisons les formats **HTML**, **XML** et **TEI**.

Enregistrement : Terme consacré dans les bases de données : c'est la colonne horizontale d'un tableau avec de multiples colonnes verticales (qui sont les **champs**). C'est équivalent à une **notice** (dans le seul usage du terme et non dans celui plus restrictif d'Omeka) à une **fiche**. Chaque objet du corpus est un « enregistrement » que cela soit un **Contenu**, une **collection** ou un **fichier**.

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

Entrée : Ce terme est utilisé dans Omeka pour désigner la ou les **valeur(s)** d'un **champ**. Chaque **champ** a une entrée minimum et en a autant que de besoin. Il n'y a donc pas besoin d'utiliser des séparateurs entre les différentes **valeurs** car dans Omeka chaque entrée est autonome dans le **formulaire**. Chaque **valeur** est donc égale à une entrée différente.

Exposition virtuelle : L'exposition virtuelle est un moyen de plus en plus utilisé par les musées, les détenteurs de fonds culturels, ainsi que les artistes pour diffuser leurs œuvres sur le web.

Sur Omeka, une exposition virtuelle permet de créer des parcours thématiques à partir des données publiées. Cependant, une exposition virtuelle ne remplace pas l'organisation du corpus en **collections** et sous-collections. Une **collection** permet de restituer intellectuellement la structure d'un corpus tandis qu'une exposition virtuelle donne la possibilité de mettre en scène les **Contenus** de ces **collections** dans une logique de valorisation. L'autre différence réside dans la pérennité des données : tandis que vous pouvez exporter dans d'autres formats les **collections** et les **Contenus**, il n'est pas possible de le faire pour une exposition virtuelle.

F

Facette : La recherche à facettes permet de filtrer des résultats en choisissant un ou plusieurs critères (les facettes). C'est souvent utilisé dans les **sites** d'e-commerce pour filtrer des résultats en fonction de catégorie. Ce n'est une réelle méthode de recherche d'information, mais plutôt un outil de support à la recherche. Il est à distinguer de la recherche par critères (qui est souvent appelé « recherche avancée »), sur laquelle la recherche par facette vient dans un second temps.

Framework : Un *framework* propose une infrastructure de développement pour un informaticien afin qu'il puisse créer une application (des lignes de code permettant de réaliser des actions). Pour Omeka, on utilise le *framework* ZENDEK et c'est à travers lui qu'on fabrique des **plugins**. Les **sites** internet ne sont pas conçus par des *framework* mais via des éditeurs **HTML**.

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

Fichier : Un fichier informatique est au sens commun, un ensemble de données numériques réunies sous un même nom, enregistrées en un seul contenant sur un support de stockage permanent.

Dans Omeka, le fichier est obligatoirement associé à un **Contenu**, il est l'étage le plus bas de l'architecture d'Omeka (**collection**, **Contenu**, Fichier) et il ne peut effectivement être consulté qu'en passant par un **Contenu**. Généralement ce sont des images mais cela peut être n'importe quel type de document, comme les traitements de texte, les vidéos ou les sons. Mais Omeka permet une visualisation aisée essentiellement pour les **images** et les **PDF**. On peut attribuer sur chaque fichier des métadonnées **Dublin Core** mais aussi, variante EMAN, des **métadonnées personnalisées**.

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

Sur Omeka, il existe trois types de formulaire pour « **Contenu** », « **Collection** » et « **Fichier** ». Mais pour les expositions virtuelles ou les « simples pages », on utilise aussi des formulaires. Cette méthode de saisie est très pratique car elle propose de faire la saisie à travers une interface graphique *a priori* aisée à comprendre.

G

Graphe : voir **Visualisation**.

H

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

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

La **plateforme** EMAN est hébergée sur une machine virtuelle de la grille de services d'Huma-Num.

I

Image : pour la numérisation et la qualité des images, voir **JPG**. Pour le nommage des fichiers, Omeka renomme de toute façon les images de façon aléatoire car il copie le fichier original en 4 formats :

- « Fullsize » : adaptation plein écran,
- « Square_thumbnails » : vignette en carré,
- « Thumbnails » : vignette,
- « Original » : format original.

Il garde toutefois le nom original dans la base de données.

Item : C'est le terme utilisé dans l'interface d'administration Omeka pour désigner parfois un **Contenu**, en particulier quand il s'agit des **métadonnées personnalisées**. La terminologie d'Omeka étant fluctuante sur cette notion, nous privilégions le terme **Contenu**.

C'est aussi le sigle du laboratoire hébergeant la **plateforme** : l'Institut des textes et manuscrits modernes (CNRS-ENS).

Item Relations : **module** d'Omeka qui met en relation deux **Contenus**. Le **module** n'est pas adapté aux **collections** et aux **fichiers**. Cette fonctionnalité est utilisée pour créer des **relations** entre n'importe quel **Contenu** du **site** mais il faut que ces contenus existent au préalable. La **relation** sur Omeka est basée sur la triade « sujet prédicat objet ». Elle se crée automatiquement dans les deux sens. Vous n'avez donc pas besoin de la saisir de nouveau pour le deuxième **contenu** relié mais cela demande une cohérence pour la dénomination du « prédicat » (il faut par exemple systématiser voix active ou passive). Omeka propose des listes de **relations** qu'EMAN a francisées et vous pouvez aussi personnaliser ces **relations**.

Item type metadata : voir **Métadonnées personnalisées**.

J

JPG/JPEG : Le « Joint Photographic Experts Group » est un format standardisé d'images numériques qui permet une visualisation dans la plupart des logiciels images ainsi que sur les navigateurs internet. C'est le format image par excellence pour internet. Car il est un format compressé et permet de diffuser des images au moindre coût taille/qualité. Mais si la compression est trop grande, l'image s'affichera moins bien dès qu'on est dans du logiciel ou de l'impression de haute qualité.

Les préconisations sont de numériser dans des formats non compressés et de haute qualité (TIFF généralement) puis de faire une copie JPEG pour la visualisation via Omeka. Celui-ci peut stocker des images haute définition dans d'autres formats mais leur affichage va être très loin et le gain de qualité pour l'affichage ne sera pas visible à l'œil. Par contre il est impératif que le JPG soit de bonne qualité, avec au minimum 300 DPI. Il faut mieux réduire la taille de l'image que le nombre de pixels (DPI). De toute façon, Omeka retaillera l'image en fonction de l'écran de consultation.

L

Licence / licence libre : voir **Creative Commons** et voir **Open source**.

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

M

Métadonnées personnalisées : Les métadonnées permettent la description et le traitement des ressources numériques, elles sont généralement standardisées et à l'extérieur ou en entête du texte ou du document qu'elles décrivent. Le **Dublin Core** (DC) est une initiative de métadonnées.

Omeka utilise des métadonnées DC mais il est possible d'en rajouter pour les **Contenus** : ce sont les métadonnées personnalisées qui complètent celles du DC. Mais il est possible d'utiliser que les métadonnées DC ou que les métadonnées personnalisées, en fonction des besoins du projet. Par contre pour le moissonnage **OAI-PMH**, seules sont exposées et récupérées les métadonnées DC. Sur la **plateforme** EMAN, il est possible d'ajouter des métadonnées personnalisées sur les **fichiers** et sur les **collections**.

Module : voir **Plugin**.

Moissonnage : voir **OAI-PMH**.

Mots-clés : voir **Tags**.

N

Notice : voir **Contenu**.

Nuage de tags : voir **Tags**.

Numérisation des images : voir **JPEG**.

O

OAI-PMH : L'Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) est un protocole informatique développé afin d'échanger des métadonnées – uniquement celles-ci et non les ressources elles-mêmes. L'OAI-PMH définit deux types d'acteurs : les fournisseurs de données, qui déposent leurs métadonnées sur un **serveur** web appelé « **entrepôt** », et les fournisseurs de service qui collectent (qui « moissonnent ») ces données. Le moissonnage s'effectue à partir de requêtes formalisées à l'adresse de l'entrepôt, les résultats sont alors intégrés dans l'index ou le répertoire du **site** moissonneur. La liste des requêtes peut être consultée par exemple sur http://www.bnf.fr/documents/intro_oaipmh.pdf. Ce protocole est utilisé notamment par les Archives Ouvertes et les **entrepôts** institutionnels, il s'est aujourd'hui largement répandu dans les institutions patrimoniales et notamment les bibliothèques. Il permet entre autre de construire des **sites** portails thématiques avec uniquement le résultat de requêtes sur les entrepôts repérés sur cette thématique.

Omeka : Omeka est un logiciel de gestion de bibliothèque numérique mis à disposition sous **licence libre** (GNU – General Public License). De conception modulaire, l'outil permet à chaque **site** d'adapter les fonctionnalités proposées à l'aide de **plugins** et de **thèmes**. L'outil est développé par le [Roy Rosenzweig Center for History and New Media](#) (CHNM) de l'Université George Mason qui est aussi à l'origine du logiciel de gestion bibliographique Zotero. La première version date en 2009 ; début 2018 on n'en était qu'à la version 2.6, ce qui est le signe d'une certaine stabilité. Le logiciel est sous système LAMP (Linux Apache MySQL PHP) et bénéficie d'une importante communauté d'utilisateurs quasi-exclusivement dans le domaine de l'enseignement supérieur, de la Recherche, de la culture et des Bibliothèques.

« Omeka est conçu avec des non spécialistes, permettant aux utilisateurs de se concentrer sur le contenu et l'interprétation plutôt que la programmation. Il apporte le Web 2.0 (technologies et approches) aux sites Web universitaires et culturels pour favoriser leur interaction la participation. Il propose un design haut de gamme facile avec un système de thèmes simple et flexible. Son développement open source robuste et des communautés d'utilisateur assurent la stabilité d'Omeka et la durabilité » (traduction de la page de présentation du site <http://omeka.org>, mai 2017).

En 2018, une autre version d'Omeka, Omeka S, est sortie ; de conception différente (plus centrée sur les **tags** sémantiques que sur les **collections** bibliothécaires), elle devait remplacer la version précédente. Mais, la page d'accueil du **site** <http://omeka.org> indique explicitement une cohabitation : à gauche, Omeka S « for institutions managing a sharable resource pool across multiple sites » (pour les institutions gérant un ensemble de ressources partageables sur plusieurs

sites) ; à droite, la version que nous utilisons et qui est maintenant dénommée Omeka Classic « for individual projects and educators » (pour les projets individuels et les enseignants, *sic*).

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

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

Omeka est un logiciel Open source. Pour les contenus, nous utilisons la licence **Creative Commons**.

P

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

Avec Omeka, sont utilisés des **thèmes** prédéfinis pour personnaliser l'affichage des différentes pages utilisées sur un **site**. Le système Omeka permet d'en gérer un petit nombre : page d'accueil, page affichant un **Contenu**, une **collection**, un **fichier**, page affichant la liste de ceux-ci, pages d'administration (voir **Formulaire**) et « simples pages » pour la présentation du projet.

Page aveugle : C'est une page qui incite à cliquer sur les pages auxquelles elles renvoient plutôt que sur elle-même. Dans un sous-menu, c'est la page qui, en passant dessus avec le curseur, donne accès à d'autres pages sur lesquels on va cliquer plutôt que sur elle. Il importe donc de ne pas y mettre de texte important.

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

Sur EMAN nous exportons les notices dans un PDF avec structure textuelle.

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

Omeka est souvent présenté comme une plateforme, nous préférons le considérer comme un logiciel et EMAN comme une plateforme rassemblant des projets différents (qui sont à chaque fois une instance Omeka) mais qui utilisent le même logiciel Omeka, les **modules** et les **thèmes** personnalisés par EMAN. D'où le fait aussi qu'EMAN ne peut être considéré comme une bibliothèque numérique : tous les projets ne partagent que la technologie et la gestion de projet.

Plugin : En informatique, un *plugin* ou *plug-in*, aussi nommé **module** greffon, plugiciel est un paquet structuré de codes informatiques qui complète un logiciel hôte pour lui apporter de nouvelles fonctionnalités. Omeka est un logiciel très léger avec de nombreux plugins plus ou moins bien développés, certains directement par les concepteurs du logiciel, d'autres dans la communauté de développeurs travaillant sur Omeka.

Sur EMAN, nous utilisons plus d'une trentaine de plugins : nous les installons tous sur chaque **site** nous les activons en fonction des besoins. EMAN développe des plugins spécifiques qui sont diffusés sur la **plateforme** de partage de code GITHUB (<https://github.com/ENS-ITEM>). Dorénavant nous essayons d'utiliser le terme français « **module** ».

R

Relations : voir **Item Relations**.

S

Serveur : Un serveur informatique est un ordinateur qui offre des services à un ou plusieurs clients (parfois des milliers). Généralement, on parle de serveur pour désigner la machine qui héberge et diffuse des **sites** internet. Le serveur qui héberge la **plateforme** EMAN est sur la grille de service d'**Huma-Num**.

Simple Pages : Module d'Omeka qui permet de créer de manière simple des pages **HTML** statiques sur un **site** Omeka pour présenter un contenu appelé à être plus ou moins stable. Elles sont appelées « statiques » pour les différencier des formulaires qui éditent l'information contenue dans la base de données et qui changent à chaque évolution de la base de données. On les dénomme parfois pages « génériques » car elles servent généralement à présenter le projet, les auteurs, les droits, etc.

Nous n'avons pour l'instant aucune traduction satisfaisant à cette expression technique, on utilise aussi bien le singulier que le pluriel mais il est important de le mettre toujours entre guillemets et avec majuscules (les pages ne sont pas aussi simples que cela !).

Signet : voir **Onglet**.

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

Chaque projet sur la **plateforme** EMAN est un site à part entière qui utilise une version autonome et personnalisée d'Omeka. EMAN a donc autant de projets que de sites.

T

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

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

Dans Omeka, on fait la distinction entre les mots-clés qui sont reportés dans la **balise** « Sujet » du **Dublin Core** et le **signet** « tags », uniquement présent pour les **Contenus** et qui permet de réaliser un nuage de tags uniquement spécifique à Omeka.

TEI : La TEI, *Text Encoding Initiative*, est un langage d'**encodage** très utilisé en Humanités numérique. Le langage **HTML** étant trop uniquement consacré à la mise en forme et le langage **XML** trop générique, la TEI propose des schémas d'**encodage** selon les types de documents. Sa flexibilité permet de s'adapter à différents types de genres littéraires (poésie, littérature classique, théâtre, etc.), ou de « formes » (manuscrits, imprimés, images, dialogues, etc.). Elle donne les outils pour réaliser la description et le découpage éditorial d'un document et non d'un ensemble (au contraire de l'EAD). C'est un langage **XML** qui utilise des **balises** et une terminologie propre ; elle est basée sur un dictionnaire de **balises** et des « guidelines », des règles d'utilisations.

Omeka permet un export en **XML** mais n'utilise pas la TEI. EMAN utilise la TEI pour la **transcription** (**module** Transcript).

Thème / Template : Un *template* (ou « thème », « layout », etc.) désigne l'enveloppe graphique d'un **site internet**, indépendamment de son contenu. Il s'agit par exemple de la disposition des colonnes, du choix des caractères ou des couleurs, de la structure des différents éléments, etc. Un *template* propose plusieurs **pages web** de base et des feuilles de style. Cela permet de séparer le contenu (les données) et le contenant (le thème), celui-ci pouvant être changé facilement sur les différents **CMS**. Omeka propose plusieurs *templates* différents, la **plateforme** EMAN propose un seul *template* pour tous les projets afin de garder une cohérence entre les **sites** et également pour éviter une gestion compliquée de l'interface publique et de ses évolutions. Nous privilégions l'usage français de « thème ».

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

Dans EMAN, on utilise le **module** Transcript v.2. Le texte transcrit est une unité qui est traitée à part des **formulaire**s Omeka et n'est pas encore intégré aux métadonnées produites par ailleurs.

Type de contenu : Terminologie propre à Omeka. Le type de contenu regroupe dans un seul **formulaire** des **métadonnées personnalisées** en complément des métadonnées **Dublin Core**. C'est un rassemblement virtuel de métadonnées, ce qui permet d'avoir des **formulaires** avec uniquement des **champs** pertinents par rapport à l'objet à décrire. Cela permet aussi de catégoriser le document sur lequel vous travaillez et d'opérer par le moteur de recherche des sélections par type.

EMAN propose par défaut six types de contenus (Article de presse, Correspondance, Document pédagogique...) mais chaque projet peut en créer autant que de besoin. Les types de contenu sur Omeka concernent les **Contenus** mais sur EMAN ils ont été élargis aux **collections** et aux **fichiers**.

U

Unicode / UTF8 : L'Unicode est un standard informatique international qui permet de décrire toutes les lettres des différentes langues. Il vise au codage du texte écrit en donnant à tout caractère de n'importe quel système d'écriture un identifiant numérique, et ce de manière unifiée, quelle que soit la plate-forme informatique ou le logiciel utilisé (à la différence d'ANSI). Il est plus complet que le code ASCII qui ne possède pas de signe diacritique. Le code informatique d'Unicode est standardisé par l'UTF (Universal Character Set Transformation Format) ; nous sommes maintenant en UTF8. Le fait que le caractère soit codé en UTF ne veut pas dire qu'il va s'afficher correctement : il faut ensuite disposer de la police de caractère adéquate mais le caractère sera bien interprété informatiquement.

V

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

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

Omeka ne propose pas de visualisation graphique des **relations**. Sur la **plateforme** EMAN, nous travaillons sur des visualisations par graphe des **relations** établis par le module **Item Relations**.

W

Wysiwyg : Wysiwyg est un acronyme anglais qui signifie « *what you see is what you get* » : « ce que vous voyez est ce que vous obtenez ». Cela désigne une interface graphique (le plus souvent par **formulaire** ou bouton) qui permet de composer visuellement le résultat attendu sans passer par l'écriture et donc l'apprentissage de codes informatique. L'exemple classique est de pouvoir réaliser une **page** web sans passer par l'écriture de **balises HTML**.

X

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

Omeka n'utilise pas XML mais permet d'exporter les listes ou les notices dans ce format avec une version DCMES (**Dublin Core** Metadata Element Set) et une version propre Omeka du XML (incluant en plus les liens vers les fichiers, les **métadonnées personnalisées**, les **tags**).