

Séance 1 - Introduction

1. Du document au système d'information

Pendant longtemps, la documentation patrimoniale s'est organisée autour de **documents** au sens classique du terme : fiches papier, inventaires, catalogues imprimés, rapports de fouille, dossiers photographiques. Le numérique n'a d'abord fait que **transposer** ces documents vers des formats informatiques (bases locales, PDF, images numérisées).

Or, cette transposition masque un changement bien plus profond : aujourd'hui, **ce ne sont plus les documents qui sont centraux, mais les données**. Ainsi, une base de données patrimoniale moderne ne se limite plus à stocker des textes ou des images. Elle organise :

- des **entités** (objets, lieux, personnes, événements)
- des **relations explicites** entre ces entités
- des **identifiants stables** permettant de les citer, de les relier et de les réutiliser.

On passe ainsi :

- d'une **logique documentaire**
- à une **logique de système d'information patrimonial**.

2. Donnée, information, connaissance : trois niveaux distincts

Il est fondamental de distinguer trois niveaux, souvent confondus :

2.1 La donnée

La donnée est une **valeur brute**, qui peut être :

- un nom,
- une date,
- une mesure,
- une coordonnée,
- un identifiant.
- etc.

Exemples :

Séance 1 - Introduction

- « Athènes »
- « -450 »
- « terre cuite architecturale »

Une donnée **n'a pas de sens par elle-même.**

2.2 L'information

L'information est une donnée **contextualisée** :

- reliée à une entité,
- placée dans un champ,
- interprétable.

Exemple :

« Le sanctuaire est localisé à Athènes »

Ici, la donnée « Athènes » devient une information parce qu'elle est liée à un objet précis et à une relation explicite (localisation).

2.3 La connaissance

La connaissance émerge lorsque :

- les informations sont structurées,
- comparables,
- reliées entre elles.

Exemples :

- comparer des ateliers de production,
- analyser des circulations,
- reconstruire des réseaux spatiaux ou chronologiques.

L'enjeu du numérique en patrimoine est précisément là : **permettre le passage de la donnée à la connaissance.**

3. Qu'est-ce que l'informatique documentaire ?

L'informatique documentaire ne consiste pas à « faire de l'informatique » au sens technique du terme. Elle consiste à **penser la structure de l'information.**

Séance 1 - Introduction

Elle s'intéresse notamment à :

- la **modélisation des données**,
- les **métadonnées**,
- les **formats d'échange**,
- les **standards de description**,
- les **identifiants pérennes**,
- l'**interopérabilité** entre systèmes.

Dans le domaine patrimonial, elle se situe à l'interface entre :

- sciences humaines,
- archivistique,
- bibliothéconomie,
- informatique,
- humanités numériques.

4. Pourquoi ces questions sont devenues centrales ?

4.1 Explosion des données

La recherche en SHS mais aussi sa médiation numérique contemporaine produit :

- des volumes massifs de données,
- hétérogènes (textes, images, SIG, 3D),
- issues de contextes multiples (fouille, prospection, archives, musées).

Sans structuration rigoureuse, ces données deviennent :

- difficilement exploitables,
- rapidement obsolètes,
- impossibles à partager.

4.2 Exigences de la science ouverte

Aujourd'hui, la recherche impose :

- la **traçabilité**,
- la **citabilité**,
- la **réutilisabilité** des données.

Un simple PDF ou une base fermée ne suffit plus.

4.3 Pérennité et responsabilité scientifique

Publier des données patrimoniales engage une **responsabilité à long terme** :

- vis-à-vis de la communauté scientifique,
- vis-à-vis des institutions,
- vis-à-vis du public.

Cela implique :

- des formats durables,
- des standards reconnus,
- des identifiants stables.

5. Le rôle des standards

Les standards ne sont pas des contraintes arbitraires. Ils constituent un **langage commun**. Ils permettent :

- de comprendre des données produites ailleurs,
- de relier des corpus distincts,
- d'éviter les systèmes fermés et propriétaires.

Dans ce cours, nous travaillerons notamment avec :

- des **standards de métadonnées** (comme le Dublin Core),
- des **ontologies** (comme le CIDOC CRM, BIBO, etc.),
- des **référentiels contrôlés** (thésaurus, autorités).

Ces outils seront abordés progressivement.

6. Vers la publication numérique des collections universitaires

Une collection universitaire publiée en ligne n'est pas :

- une simple vitrine,
- un catalogue figé.

C'est :

- un **objet scientifique**,
- un **système de données**,

Séance 1 - Introduction

- potentiellement **interopérable** avec d'autres plateformes.

C'est dans cette perspective que s'inscrit l'usage de **Omeka S** :

- non comme un CMS (système de gestion de contenu) ordinaire,
- mais comme un **outil de publication sémantique**.

7. Qu'est-ce que un CMS ?

7.1 Définition générale de CMS et de "contenu"

Un **CMS** (*Content Management System*, système de gestion de contenus) est un logiciel qui permet de :

- créer,
- organiser,
- gérer,
- et publier des contenus sur le Web sans avoir à programmer directement en HTML ou en bases de données.

Mais cette définition est **trompeusement simple**, car elle masque une réalité essentielle : **tous les CMS ne reposent pas sur la même conception du "contenu"**.

Selon les outils, un contenu peut être pensé comme :

- une **page**,
- un **article**,
- un **objet documentaire**,
- une **entité décrite par des métadonnées**.

C'est précisément sur ce point que se joue la différence entre des CMS généralistes (WordPress, Wix) et des CMS patrimoniaux comme Omeka.

7.2 CMS généralistes : WordPress et Wix

Des CMS comme **WordPress** ou **Wix** sont conçus avant tout pour :

- la communication,
- la publication de pages,
- le marketing de contenu,
- la visibilité institutionnelle ou commerciale.

Séance 1 - Introduction
Leur unité de base est :

- la **page** ou l'**article**, organisés dans une hiérarchie de menus.

Les contenus sont :

- pensés pour être **lus**,
- rarement pour être **réutilisés comme données**,
- faiblement structurés sémantiquement.

Dans WordPress ou Wix :

- les métadonnées existent (titre, auteur, date),
- mais elles servent surtout :
 - à l'affichage,
 - au référencement (SEO),
 - à la navigation.

Elles ne sont **ni normalisées**, ni conçues pour :

- l'interopérabilité scientifique,
- la description fine d'objets patrimoniaux,
- la mise en relation explicite entre ressources.

Ces CMS sont donc **excellents pour des sites vitrines**, mais **inadaptés à la construction d'un système d'information patrimonial**.

7.3 Omeka : un CMS documentaire et patrimonial

Omeka n'est pas un CMS éditorial au sens classique. C'est un **CMS documentaire**, conçu dès l'origine pour :

- les bibliothèques,
- les archives,
- les musées,
- la recherche et l'enseignement.

Son unité de base n'est pas la page, mais :

- l'**item** (ressource ou contenu),
- décrit par des **métadonnées explicites** (qui font recours à des vocabulaires sémantiques et/ou des thésaurus).

Une image, un objet, une archive, un modèle 3D ou un mémoire de master sont :

- des **entités documentaires**,

Séance 1 - Introduction

- dotées de propriétés,
- potentiellement reliées entre elles.

Omeka repose sur :

- des **standards de métadonnées** (Dublin Core),
- des vocabulaires explicites,
- une séparation claire entre :
 - données,
 - métadonnées,
 - médias,
 - exposition.

Cette logique est **fondamentale en contexte patrimonial**, car elle permet :

- la description raisonnée,
- la traçabilité scientifique,
- la réutilisation des données.

8. CMS libre et ouvert vs CMS propriétaire

Le choix d'un CMS n'est jamais neutre.

Il engage :

- des pratiques scientifiques,
- des modèles économiques,
- des politiques institutionnelles.

8.1 Omeka : un CMS libre et ouvert

Omeka est :

- **libre** (licence open source),
- **ouvert** (code accessible),
- **interopérable**,
- porté par une communauté académique internationale.

Avantages

- transparence du fonctionnement ;
- pérennité des données (formats ouverts) ;
- possibilité d'export et de migration ;
- indépendance vis-à-vis d'un prestataire unique ;

Séance 1 - Introduction

- alignement avec les principes de la science ouverte.

Limites

- nécessite des compétences techniques (installation, maintenance) ;
- interface parfois moins "clé en main" ;
- demande une réflexion conceptuelle forte en amont ;
- coût humain (temps, formation).

Omeka **déplace l'effort** : moins de facilité immédiate, mais plus de maîtrise scientifique.

8.2 FLORA : un CMS propriétaire spécialisé

FLORA est un CMS patrimonial propriétaire, payant, utilisé notamment :

- par **Sorbonne Université** (ex Paris 4),
- par **INRAP** pour son système documentaire **DOLIA**.

FLORA propose :

- une solution intégrée,
- un accompagnement professionnel,
- une stabilité fonctionnelle,
- des outils adaptés aux institutions patrimoniales.

Avantages

- solution immédiatement opérationnelle ;
- support technique dédié ;
- homogénéité des pratiques ;
- cadre rassurant pour les grandes institutions.

Limites

- dépendance à un éditeur ;
- coût financier élevé ;
- évolutions conditionnées par le prestataire ;
- interopérabilité partielle ;
- difficulté de migration à long terme.

8.3 Un choix scientifique autant que technique

Le choix entre Omeka (libre, ouvert) et un CMS propriétaire comme FLORA n'est pas un simple choix d'outil. C'est un choix :

- de **gouvernance des données**,
- de **rapport au temps long**,
de **positionnement vis-à-vis de la science ouverte**.

Dans le cadre de l'UFR 03 et du master *Valorisation et médiation du patrimoine archéologique*, le choix d'Omeka permet :

- de comprendre les mécanismes de structuration ;
- de former des étudiants **autonomes et critiques** ;
- de produire des données réutilisables ;
- d'inscrire les mémoires de master dans un écosystème scientifique durable.