

Année académique 2024-2025



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Informatique et documentation patrimoniale
TD 2 (mardi 08-10-2024)

Vincenzo Capozzoli

PARTIE I

Organisation des informations et objets

Importance pour :

- **Préservation**
- **Identification**
- **Utilisation**

Évolution des méthodes :

- **Inventaire manuel → Bases de données modernes → Web sémantique**

David Weinberger et ses "trois ordres d'organisation" (tiré de *Everything Is Miscellaneous*, 2007)

Premier ordre : Organisation physique

Définition : Organisation directe des objets

Exemples :

- Bibliothèque : Livres classés par auteur, thème, ordre alphabétique
- Musée : Artefacts classés selon provenance, époque

Limite :

- Un objet ne peut être classé que selon un seul critère
- Exemple : Un livre ne peut être à la fois classé par auteur et par thème sur une même étagère

Deuxième ordre : Le catalogue

Définition : Séparation entre objets et description.

- Permet de multiples entrées pour un objet (auteur, titre, sujet, etc.).
- Exemple : Catalogue de livres ou photos avec critères multiples (date, sujet).

Progrès avec les catalogues numériques :

- Moins de limites physiques.
- Catalogue plus flexible avec plus de critères possibles.

Limites des catalogues physiques

Problème : Catalogues traditionnels dépendent de la forme physique, limitant l'accès rapide à l'information.

Solution numérique :

- Catalogues numériques permettent plus de flexibilité dans l'organisation des données.
- L'**imagination** et la **patience du catalogueur** deviennent les seules limites.

Troisième ordre : Bases de données relationnelles

Définition : Catalogue devient une base de données enrichie de métadonnées (informations détaillées sur les objets).

Limite des premières bases de données :

- Rigidité des critères traditionnels, comme nom de l'auteur ou date de publication, ne permettant pas toujours de capturer la complexité des objets.
- Problème avec les objets sans nom ou date précise, forçant une catégorisation artificielle.

Complexité des objets

Réflexion : La réalité est souvent plus complexe que les schémas rigides.

Exceptions : Certaines informations ne peuvent être correctement décrites dans les bases de données rigides.

Everything is miscellaneous (David Weinberger)

Concept clé : Le monde est complexe et diversifié.

- Essayer de le décrire avec des systèmes trop rigides simplifie à l'excès et **dénature la richesse** des informations.

Besoin d'outils plus flexibles :

- Descriptions plus riches et précises pour mieux refléter la réalité des objets et informations.

Web sémantique : une organisation flexible

Définition : Extension du Web traditionnel (Tim Berners-Lee).

- Informations interconnectées via des **réseaux complexes de données**.
- Utilisation des **métadonnées** pour relier objets, concepts et idées.

Avantages :

- Bases de connaissances dynamiques.
- Recherche selon divers critères (auteur, sujet, époque historique, etc.).

Transition vers le Web sémantique

Évolution : Catalogue physique → Bases de données → Web sémantique.

Bases de données :

- Modèle plus sophistiqué pour organiser et rechercher des informations.
- **Systèmes de gestion de bases de données (SGBD) :** efficacité accrue.

L'essor des métadonnées

Schémas de description :

- **Dublin Core (DC), MARC.**
- Facilitation de l'accès et gestion des objets en environnement numérique.

Impact :

- Descriptions plus riches associées aux objets, améliorant les systèmes de gestion de bibliothèques numériques.

Web sémantique : Au-delà des bases de données

Avantage principal : Dépassement des limites des catalogues et bases de données traditionnels.

- **RDF (Resource Description Framework)** : Structure en triplés pour relier les données.
- **Requêtes Sparql** : Recherche sophistiquée et visualisations inédites.

Transition vers les bibliothèques digitales

Bibliothèque digitale (Digital Library, DL) :

- Collection d'informations numériques.
- **Objectifs :**
 - Conservation à long terme.
 - Gestion organisée.

Caractéristiques :

- Contenus riches et variés.
- Accessibles à diverses communautés.
- Gestion et consultation avec des **niveaux de qualité mesurables.**

Les débuts de l'informatique

Contexte : Emergence dans les années 1950-1960.

Origine du terme :

- Combinaison de "informat(ion)" et "(automat)ique".

Objectif initial : Concevoir des machines pour des calculs rapides et automatiques.

Avancée clé : Stockage permanent des données.

L'émergence de l'informatique

Définition : Informatique = représentation et gestion automatisée des informations.

Utilisation : Domaines divers comme la gestion des données et la recherche scientifique.

Impact : Large adoption des ordinateurs.

La naissance des systèmes de gestion de données

Problème initial : Besoin d'un **système de catalogage numérique** pour gérer la croissance des informations.

Solution : Création de standards pour structurer les données numériques.

Technologies clés :

- **Codage numérique.**
- **Métadonnées** pour décrire des informations (textes, images, sons).

Les premières bases de données et la gestion des métadonnées

Années 1970 : Introduction des **bases de données** pour un stockage organisé et flexible.

Systèmes de gestion de bases de données (SGBD) :

- Logiciels gérant de grandes quantités d'informations.

Métadonnées : Définition fine des caractéristiques des documents, permettant des liens complexes entre informations.

L'avènement du Web

Contexte : Le Web a transformé l'accès aux informations à une échelle mondiale.

Différentes étapes :

- La naissance du World Wide Web
- L'essor des navigateurs Web
- Le W3C et la standardisation du Web



L'inventeur du World Wide Web (WWW), Tim Berners-Lee

La naissance du World Wide Web (WWW)

Création : Mars 1989, CERN à Genève.

Projet : Un système **hypertexte distribué** reliant des documents sur plusieurs ordinateurs via Internet.

Technologies clés :

- **HTTP** : Hypertext Transfer Protocol.
- **HTML** : Hypertext Markup Language.

Résultat : Accès facile aux informations via des navigateurs, grâce aux liens hypertextes.

L'essor des navigateurs Web

Premiers navigateurs :

- 1990 : Prototype affichant uniquement du texte.
- 1993 : **Mosaic** introduit une interface graphique.
- 1995 : **Netscape** améliore la convivialité de la navigation.

Impact :

- Diffusion rapide du Web à l'échelle mondiale.
- Apparition des sites Web éducatifs, scientifiques, et industriels.
- Utilisation de **URL** (Uniform Resource Locator) pour naviguer entre les documents (iu, en français, « localisateur uniforme de ressource »). Une URL est simplement l'adresse d'une ressource donnée, unique sur le Web. En théorie, chaque URL valide pointe vers une ressource unique.

Le W3C et la standardisation du Web

Création du W3C (World Wide Web Consortium) : 1994,
collaboration entre CERN et MIT.

Mission du W3C :

- Standardiser les protocoles du Web.
- Promouvoir l'**interopérabilité** mondiale.

Résultat : Outils et applications compatibles entre différents systèmes, favorisant l'utilisation dans les domaines académiques, industriels et culturels.

Vers les bibliothèques numériques

Contexte : Numérisation des collections des bibliothèques et institutions dès les années 1990.

Objectif : Rendre accessibles des documents variés via des **bibliothèques numériques**.

L'évolution des bibliothèques numériques

Initiatives de numérisation :

- **Collections variées** : Livres, images, vidéos, manuscrits, artefacts culturels.
- **Accessibilité à distance** : Recherche facilitée pour des publics divers.

Exemples d'initiatives :

- **Digital Library Initiative (DLI)** (États-Unis) : Financé par NASA et NSF.
- **Projet européen Delos** (1996-2007) : Préservation du patrimoine culturel et historique.

L'importance des métadonnées

Rôle des métadonnées :

- **Description précise** des documents numériques (livres, images, artefacts).
- **Facilitation de l'accès** et de la recherche.

Standards utilisés :

- **Dublin Core (DC)** et **MARC** pour décrire les collections numériques.

Avantages :

- **Recherche facilitée** par des critères variés (auteur, date, sujet).
- **Interopérabilité** entre différents systèmes numériques.

Services avancés dans les bibliothèques numériques

Services clés : Annotations numériques et narratives interactives.

Objectif : Favoriser l'engagement et la collaboration dans les bibliothèques numériques.

Les annotations numériques

Définition : Les utilisateurs peuvent ajouter des **observations** ou **commentaires** sur des documents numériques.

Exemple : Université de Padoue – système d’annotations pour enrichir les documents numérisés.

Avantage : Favorise la collaboration entre utilisateurs et enrichit les collections.

Les narratives : un outil d'engagement

Narratives : Récits interactifs qui créent des parcours thématiques dans les collections numériques.

Inspiration : Techniques de **gamification** pour rendre l'expérience plus immersive.

Avantage : **Engagement accru** des utilisateurs à travers des explorations guidées par des experts.

Conclusion : Vers les bibliothèques numériques du futur

Environnements de recherche virtuelle (VRE) :

- **Consultation** et **participation** active à l'enrichissement des collections.
- Interaction et collaboration entre utilisateurs et contenus.

Futur des bibliothèques numériques :

- Intégration de technologies avancées.
- **Utilisateur** devient acteur clé dans la gestion de l'information.

PARTIE II

Du thésaurus à l'ontologie : Introduction

Thème : De l'usage des mots à la visualisation des données.

Problème : Fragmentation des savoirs patrimoniaux avec la numérisation.

Solution : Utilisation de **thésaurus**, **taxonomies**, et **ontologies** pour structurer les données.

Des données aux ontologies

Patrimoine numérique : Fragmenté, nécessite une meilleure gestion.

Rôle des institutions : Recentrer l'**accès aux données** et encourager l'interaction utilisateur.

Thésaurus et vocabulaire contrôlé : Améliorent l'interopérabilité et la structuration des données.

Architecture de l'information et cycle de vie des données

Architecture de l'information : Fondée sur le Web et les bibliothèques numériques.

Web 2.0 : Données participatives et réutilisables.

Web sémantique : Structuration des données pour exploitation automatique.

Le cycle de vie des données

Étapes : Création → Analyse → Conservation → Réutilisation.

Données patrimoniales : Numérisation des objets (textes, photos) et leur gestion.

Réutilisation des données : Paradigme du Web 2.0 et Web de données.

Données vs Métadonnées

Données : Informations brutes (photo, description).

Métadonnées : Informations contextuelles (nom, date, lieu).

Types de métadonnées :

- **Descriptive** : Nature et caractéristiques.
- **Gestion** : Localisation, droits d'accès.
- **Structure** : Organisation des documents.

Interopérabilité et principes FAIR

Interopérabilité : Capacité des systèmes à partager des données.

Principes FAIR :

- **Findable** : Données faciles à localiser.
- **Accessible** : Données consultables facilement.
- **Interoperable** : Capacité à interagir avec d'autres systèmes.
- **Reusable** : Données réutilisables sous certaines conditions.

Langages de description des patrimoines numérisés

Thésaurus : Vocabulaire contrôlé pour indexer et classifier les contenus.

Taxonomies : Classification hiérarchique des ressources.

Ontologies : Représentation des concepts et relations entre eux, facilitant la compréhension par humains et machines.

Le CIDOC-CRM : un exemple d'ontologie

CIDOC-CRM : Ontologie pour structurer les informations patrimoniales.

Exemple : Monument à Balzac de Rodin :

- Description des étapes de création et des relations entre l'œuvre, les acteurs, et les événements.
- Visualisation des multiples états et interprétations de l'œuvre.

Conclusion

Importance des thésaurus, taxonomies, et ontologies :

- Structuration des données culturelles et patrimoniales.
- Facilite la **recherche scientifique** et l'accès du **grand public**.
- Incontournable à l'ère du Web sémantique pour garantir la **pérennité** et la **diffusion des informations**.