

# Sciences humaines, sciences sociales à l'ère numérique : la mise en données et méthodes de modélisation des connaissances

Stéphane Pouyllau  
*Ingénieur de recherche hors classe CNRS*  
*Professeur attaché à l'université d'Evry-Paris-Saclay*  
*Orcid : 0000-0002-9619-1002*

Séance du 18 novembre 2024

## **1 Introduction (10 minutes)**

### **1.1 Présentation du cours**

- Objectifs du cours
- Importance de la structuration des données[1]
- Aperçu des formats CSV, SGBDR, XML (TEI), et RDF

### **1.2 Contexte et applications**

- Utilisations courantes de chaque format
- Exemples d'industries et de cas d'utilisation

## **2 Partie 1 : CSV (20 minutes)**

### **2.1 Introduction à CSV**

- Définition et historique
- Structure de base d'un fichier CSV

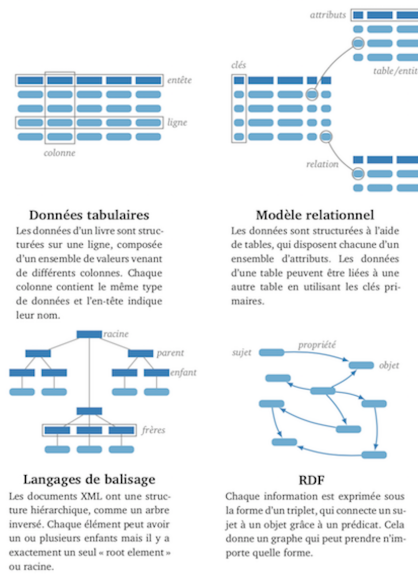


Figure 1: Seth van Hooland & al., 2016.

## 2.2 Syntaxe CSV

- Séparateurs de colonnes et de lignes
- Gestion des valeurs manquantes
- Exemple de fichier CSV

Listing 1: Exemple de fichier CSV

```
name,age,city
Alice,30,New York
Bob,25,Los Angeles
Charlie,35,Chicago
```

## 2.3 Avantages et inconvénients

- Simplicité et lisibilité
- Limitations en termes de structure et de types de données

## 2.4 Exercice pratique

- Créer un fichier CSV
- Importer et manipuler des données CSV avec un outil de tableur (Excel, Google Sheets)

## 3 Partie 2 : Modèle Relationnel (SGBDR) (20 minutes)

### 3.1 Introduction au modèle relationnel

- Définition et historique
- Concepts de base : tables, colonnes, lignes, clés primaires et étrangères

### 3.2 Syntaxe SQL

- Création de tables
- Insertion de données
- Requêtes de sélection
- Exemple de schéma relationnel

Listing 2: Exemple de schéma relationnel en SQL

```
CREATE TABLE Employees (  
    EmployeeID INT PRIMARY KEY,  
    FirstName VARCHAR(50),  
    LastName VARCHAR(50),  
    DepartmentID INT,  
    FOREIGN KEY (DepartmentID) REFERENCES Departments(DepartmentID)  
);  
  
CREATE TABLE Departments (  
    DepartmentID INT PRIMARY KEY,  
    DepartmentName VARCHAR(50)  
);
```

### 3.3 Avantages et inconvénients

- Intégrité des données et transactions
- Complexité de gestion des relations

### 3.4 Exercice pratique

- Créer un petit schéma relationnel
- Exécuter des requêtes SQL pour manipuler les données

## 4 Partie 3 : XML (TEI) (20 minutes)

### 4.1 Introduction à XML

- Définition et historique
- Structure de base d'un document XML

### 4.2 Syntaxe XML (TEI)

- Éléments et attributs
- Balises et nœuds
- Exemple de document XML TEI

Listing 3: Exemple de document XML TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Example Document</title>
        <author>John Doe</author>
      </titleStmt>
      <publicationStmt>
        <publisher>Example Publisher</publisher>
        <date>2023</date>
      </publicationStmt>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>This is an example of a TEI XML document.</p>
    </body>
  </text>
</TEI>
```

### 4.3 Avantages et inconvénients

- Flexibilité et extensibilité
- Complexité et verbosité

### 4.4 Exercice pratique

- Créer un petit document XML TEI
- Valider le document XML avec un outil en ligne

## 5 Partie 4 : RDF (20 minutes)

### 5.1 Introduction à RDF

- Définition et historique
- Concepts de base : triples, sujets, prédicats, objets

### 5.2 Syntaxe RDF

- Représentation en XML, Turtle, JSON-LD
- Exemple de document RDF

Listing 4: Exemple de document RDF en Turtle

```
@prefix ex: <http://example.org/> .
```

```
ex:Alice ex:knows ex:Bob .  
ex:Bob ex:age 25 .  
ex:Bob ex:livesIn ex:LosAngeles .
```

### 5.3 Avantages et inconvénients

- Flexibilité et interopérabilité
- Complexité de gestion des données

### 5.4 Exercice pratique

- Créer un petit document RDF
- Valider le document RDF avec un outil en ligne

## 6 Conclusion et Q&A (10 minutes)

### 6.1 Résumé des points clés

- Récapitulatif des formats CSV, SGBDR, XML (TEI), et RDF
- Comparaison des avantages et inconvénients

### 6.2 Questions et réponses

- Répondre aux questions des participants
- Discussion ouverte sur les applications pratiques

## 7 Ressources supplémentaires

- Liens vers des outils en ligne pour valider XML, manipuler CSV et visualiser des graphes.
- Lectures recommandées pour approfondir chaque sujet.

## 8 Matériel nécessaire

- Ordinateur avec accès à Internet pour les exercices pratiques.
- Outils de tableur (Excel, Google Sheets) pour manipuler les fichiers CSV.
- Outils de gestion de bases de données relationnelles (MySQL, PostgreSQL) pour les exercices sur le modèle relationnel.
- Outils de validation XML pour les exercices sur XML (TEI).
- Outils de validation RDF pour les exercices sur RDF.

## 9 Bibliographie

### References

- [1] Simon Hengchen et Max De Wilde Seth van Hooland Florence Gillet. *Introduction aux humanités numériques : méthodes et pratiques Sciences humaines et sociales*. De Boeck Supérieur, 2016. DOI: 10.3917/dbu.dewi.2016.01.