



Pourquoi et comment gérer ses fonds d'archives sous forme de données liées au moyen de logiciels libres ?

Exemple de la Fondation SAPA (Suisse)

Baptiste de Coulon, Fondation SAPA, Archives suisses des arts de la scène
LibreABC 2025, MEG (Genève), 9 septembre 2025

Plan

Introduction

Partie 1 : Création

Partie 2 : Maintenance

Partie 3 : Réutilisation

Focus : Outils

Conclusions

Fondation SAPA

Fondation SAPA, Archives suisses des arts de la scène

- Centre de compétences sur
 - la préservation des arts de la scène
 - la scène suisse des arts de la scène
- Centre de conservation des arts de la scène suisses

Fondée en 2017, héritière de plusieurs institutions précédentes

- Collection suisse de la danse (2011), Mediatektanz (2005), Archives suisse de la danse (1993)
- Schweizerische Theatersammlung (fondée en 1927)

Plus d'info sous <https://www.sapa.swiss/fr/uber-uns>

Baptiste de Coulon

Archiviste au sein de la Fondation SAPA depuis 2021

En charge de

- la maintenance des infrastructures et des données, relatives aux collections
- l'ouverture de l'accès numérique à ces données (open data et content)

baptiste.decoulon@sapa.swiss

<https://orcid.org/0009-0003-7448-7835>



Introduction

Sujet : données patrimoniales

Les institutions patrimoniales doivent se saisir des problèmes/questions posées par leurs données numériques.

La première étape est de **comprendre les questions posées**.

Il faut donc trouver un **langage** compréhensible permettant ce dialogue.

Les données liées et leurs modèles sont un langage qui permet aux non-informaticiens de comprendre les questions que leur posent les données.

Ce langage permet en retour de poser des questions aux données.

Introduction

Pourquoi utiliser des logiciels libres?

Car le passage aux données liées changent la perspective. Les logiciels ne sont plus au centre de l'attention, mais les données oui

Or, les logiciels libres sont construits sur la dissociation logiciel/données

Pour répondre à l'exigence "Public Money, Public Code" !

Pour mutualiser les frais de développement

Pour soutenir les institutions pauvres en ressources financières en supprimant les frais de licence

Introduction

Comment utiliser des logiciels libres?

En investissant de leur développement

En rejoignant/animant une communauté d'utilisateurs/trices

Remarques introductives (I)

1. Cette présentation expose les bouleversements induits au sein des institutions patrimoniales par la transformation des métadonnées sur leurs collections sous forme de données liées
2. Elle se focalise sur les données liées au sein d'un outil de gestion (type AIS/SIA)

Remarques introductives (II)

1. Cette présentation ne vise pas à présenter un retour d'expérience vis-à-vis de l'utilisation d'un logiciel en particulier
2. Il n'y sera pas présenté dans le détail ce que sont les données liées
3. N'y sera pas envisagé les données liées
 - a. au sein des paquets d'archivage numérique (SIP, AIP, DIP)
 - b. au sein des salles de consultation virtuelles
4. Les références données entre parenthèses renvoient aux listes présentes en fin de présentation



Partie 1 : création

Données liées

≠ principalement une innovation technique

Car la “nouvelle” technique ne va pas résoudre en soi les “anciens” et “nouveaux” problèmes mais elle va permettre surtout de mieux les comprendre et y répondre.

On va pas faire les mêmes choses en mieux. On va faire des choses différentes.

Il s’agit surtout d’un changement d’attention (sensibilité)

Les (méta)données numériques comme des “choses” fragiles et précieuses

Les données (et leur maintenance) sont des éléments essentiels et incontournables à la bonne conduite des travaux patrimoniaux

Données liées : définition

Les données liées sont une technologie de structuration des métadonnées descriptives.

Elle offre l'opportunité de traiter tous éléments descriptifs (objets et relations) sur pied d'égalité en permettant de les identifier distinctement de façon pérenne.

Données liées : déploiements

On peut envisager deux déploiements des données liées

- Étape 1 : à l'interne de l'institution

première étape, agrégation des données, pas forcément la plus simple (de Coulon, 2024a) mais certainement la plus importante

- Étape 2 : de façon ouverte, librement accessibles sur Internet (Open Data)

seconde étape, la plus classique et la plus exemplifiée, dont l'utilisation reste embryonnaire de la part des institutions

Données liées : fondamentaux

Les éléments fondamentaux d'un bon projet de données liées sont :

- La mise à disposition d'une **documentation du modèle de données** utilisé
- Valorisation des travaux de **maintenance**
- Une **interface générique de requêtes** permettant des interrogations précises des données

Données liées : bouleversements

Changement de perspective complet pour les institutions

- L'institution prend ses responsabilités vis-à-vis de ses données. Cela ne peut plus être la faute des "autres".
- Les données qui sont au coeur de l'attention
- Les données sont partagées entre les équipes. Fin de la logique des silos intra-institutionnels
- Le modèle de données doit être établi et discuté conjointement
- La maintenance des données a une place reconnue
- La complémentarité entre les bases internes et externes est reconnue
- La contribution sur des bases externes assumée

Données liées : avantages

- L'explicitation du modèle des données utilisé offre un gage de pérennité technique et institutionnel vis-à-vis du renouvellement des équipes
- L'indépendance vis-à-vis des logiciels (et de leurs vendeurs)
- Le déploiement d'une standardisation des descriptions (via les ontologies)
- La garantie de pouvoir répondre aux spécificités des besoins de chaque institution (via un modèle de données personnalisé)
- Une réutilisation démultipliée et simultanée des données

Modèle de données : définition

Un modèle de données liées est un schéma documenté de structuration des données.

Il est composé de

- classes d'objet
- propriétés

Un modèle est représenté au travers de sa documentation.

Modèle de données : définition

Il existe des modèles de données liées prêts à l'emploi qu'on appelle des ontologies.

Un modèle de données liées établi par une institution est habituellement composé à partir de différentes ontologies.

Les bonnes pratiques voudraient que pour forger son modèle de données l'on utilise au maximum des ontologies existantes avant de créer des nouvelles classes ou propriétés.

Ontologie : définition

On appelle ontologie, une représentation codifiée d'une partie de la réalité, publiée sous forme d'un modèle de données liées prêt à l'emploi rattaché à un espace de nom.

Elle est caractérisé par le fait que :

- elle est publiée
- elle dispose d'un espace de nom (préfixe)
- elle est référençable
- la responsabilité de sa maintenance est établie

Les ontologies disponibles sont souvent spécifiques à une réalité déterminée, un domaine d'activité par exemple.

Ontologie : exemples

EBUcore est une ontologie qui définit des classes d'objets et des relations spécifiques au domaine audiovisuels (images animées et sons). Elle est maintenue par European Broadcasting Union. SAPA l'utilise beaucoup pour décrire ses collections de vidéos.

Autre exemple : l'ontologie RiC-O basé sur la norme *Records in Contexts*.

Voir à ce sujet la présentation de Jan Krause-Bilvin qui suivra celle-ci, et notre article (de Coulon, 2004a)

SAPA RDF Data Model

En 2017, SAPA a décidé stratégiquement d'établir un modèle données de données liées pour décrire ses collections

En 2021, la Fondation a publié une première version en ligne de ces données sous forme liées suivant ce modèle.

En 2024, Elle a commencé à publier une documentation complète de son modèle de données sous forme SHACL : <https://shapes.performing-arts.ch>

SAPA RDF Data Model

Notre modèle de données comprend des parties des ontologies suivantes :

- rico : Records in Contexts (RiC)
- crm: CIDOC Conceptual Reference Model (CRM)
- FRBRoo: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism
- PREMIS : Preservation Metadata: Implementation Strategies
- EBUcore : European Broadcasting Union metadata specifications

et dans une moindre mesure:

- prov : Provenance
- rda : Resource Description & Access
- RDFS
- Schema.org
- olo : The Ordered List Ontology
- owl : Web Ontology Language
- ore : Object Reuse and Exchange. Open Archives Initiative.

SAPA RDF Data Model : SPAO

Au sein du SAPA RDF Data Model nous avons dû aussi développer une extension à l'ontologie RiC-O pour les besoins non couverts encore par cette dernière :

- spao : swiss performing arts ontology
<https://ontology.performing-arts.ch/spao>

Cette ontologie maison vise à compléter les besoins qu'elle n'a trouvé traités dans aucune autre ontologie.

Données liées : exemples de question

Voici les types de questions posées par les données :

- Quand tu écris URI, tu veux parler d'un Canton d'Uri en Suisse ? Ou d'un Uniform Resource Identifier?
- Es-tu certain que les objets dont cette personne est censé être l'auteur ont bien été créés avant sa naissance?
- Certains fonds sont insérés dans des dossiers, cela a-t-il un sens ?

Données liées : questions

Les trois questions évoquées posent des questions d'ordre divers :

- la première est une question identification/désambiguïsation
- le second est une question de logique général
- la dernière est une question de logique interne au modèle de données choisies

Dans tous ces cas, les données liées demandent aux institutions des réponses qui ne sont pas techniques mais stratégiques. Les réponses qui seront apportées auront une dimension technique mais là n'est pas l'essentiel.

Données liées : réponses

Les données liées permettent aux institutions de répondre aux questions posées plus haut, au moyen de :

- identification précise et pérenne des descripteurs et des autorités.
- interrogation croisée de plusieurs bases pour utiliser des données complémentaires non existantes dans les données de l'institution.
- établissant de modèle de données précis qui définissent des règles de logiques internes aux données.



Partie 2 : maintenance

Données liées et maintenance

Le travail des données liées comprend deux phases

- la création du modèle et la transformation des données selon ce dernier
- la maintenance du modèle et des données

Chez SAPA, on peut dire que depuis 2021, nous effectuons principalement des travaux de maintenance

Maintenance

La maintenance est une activité dévaluée et invisibilisée parmi les activités humaines mais elle reste fondamentale et omniprésente.

Elle a commencé à émerger comme récent objet d'étude des sciences humaines (Denis et Pontille, 2022) ce qui entraîne sa réévaluation.

Elle aurait par ailleurs une importance renouvelée vis-à-vis des défis du partage et de la préservation des ressources.

La maintenance engage des compétences spécifiques.

La maintenance qui vise le soin des choses demande une connaissance spécifiques de ces dernières.

La maintenance des données

Les données numériques sont des “choses” dont il faut prendre en soin.

Elles sont fragiles et précieuses.

Elles demandent une maintenance continue.

Les données et le patrimoine

Les données numériques sont souvent un impensé des travaux patrimoniaux.

Si elles sont considérées,

- elles le sont comme un aspect secondaire et technique.
- leur gestion est externalisée

Les données et les bugs

Les données numériques ne cessent pourtant d'intervenir dans tous les travaux patrimoniaux, souvent en posant des soi-disant problèmes ou des bugs.

On cherche alors l'outil logiciel qui permettra de résoudre ces problèmes.

Mon point de vue est que si les données interfèrent dans nos travaux c'est plutôt en posant des questions que des problèmes.

L'importance des bugs

Tout comme la maintenance, le bug a fait récemment l'objet d'une réévaluation dans la recherche scientifique (Vitali-Rosati, 2024).

Le bug doit être envisagé à la fois comme une question et comme une chance de mieux saisir la réalité numérique. Il est indissociable du travail numérique, ou du moins du travail autonome avec le numérique.

Savoir accueillir les bugs comme opportunité, permettraient de reconstruire un rapport riche et apaisé au numérique, et d'éviter de chercher à tous les prix (surtout les plus élevés) des machines qui nous les cachent.

Partie 3 : réutilisation

Données liées : réutilisation

On a souvent pensé que les réutilisations des données liées seraient externes (étape 2 du déploiement) mais l'expérience démontre que la première bénéficiaire est l'institution elle-même dès l'étape 1 du déploiement.

Exemples pour SAPA :

- formulaires de recherche avancée
- <https://tls.performing-arts.ch>
- <https://vocab.performing-arts.ch>

Focus : outils

Données liées : logiciels libres

Le choix des données liées par une institution patrimoniale permet que la structuration des données soient totalement indépendantes des logiciels pour les gérer.

L'attention (et les moyens financiers) des institutions devraient donc porter en premier lieu sur cette structuration (et la maintenance des données) et non sur les outils pour les gérer.

Cela induit une liberté gagnée pour le choix des logiciels.

L'interopérabilité étant une des forces des logiciels libres, ce sont ces derniers qui s'imposent dès lors rapidement.

Données liées : interfaces (niveau 1)

Les solutions logiciels permettant de maintenir des données liées doivent proposer de base les aspects suivants :

- un système d'identification pérenne (URI) des éléments
- un entrepôt de données (triplestore)
- un point d'accès permettant une interrogation riche des données selon un langage choisi (SPARQL)
- une documentation du modèle de données utilisées (SHACL)

Comme souvent, il s'agira ici d'un ensemble de logiciels nécessaires à ces tâches.

La plupart des logiciels disponibles à ce niveau sont libres.

Données liées : interfaces (niveau 2)

Au-delà du niveau 1, on retrouve avec les données liées aussi les attentes classiques vis-à-vis d'une solution logicielle de gestion d'archives comme :

- des formulaires de saisie manuelle
- l'affichage des données
- des formulaires de recherche avancées
- des fonctionnalités d'import en masse
- des fonctionnalités de changement en masse
- des fonctionnalités d'export des résultats

C'est à ce niveau-là que les logiciels sont les plus souvent développés de façon propriétaire.

SAPA : logiciels libres

Chez SAPA, nous privilégions depuis 2021 l'utilisation des logiciels libres pour la gestion de nos données liées.

Le fait qu'ils soient libres ne signifie pas que nous les gérons tout seul.

L'avantage c'est que nous pouvons choisir les prestataires pour nous épauler dans leur gestion. Et d'en changer. Ce que nous avons déjà fait.

Cela demande un effort de documentation de la part du prestataires externes.
Nous utilisons pour cela un dépôt Git.

Données liées : solutions logicielles libres existantes

Dans le domaine des données liées, on peut citer les solutions logicielles libres suivantes :

- ResearchSpace
- glam-community Metaphactory (plus maintenu)
- Arches
- OmekaS

Toutes vont proposer (grosso modo) les interfaces de niveau 1, par contre il est possible que certaines interfaces de niveau 2 fassent encore défaut

→ mutualisation du financement de leur développement

SAPA : logiciels de gestion d'archives

La Fondation SAPA utilise :

- avant 2021 : ScopeArchives, FileMaker, MS XLS
 - 2021 à 2022 : metaphactory (metaphacts)
 - 2022 à 2025 : glam-community (metaphacts) - open source
 - 2026 : passage à ResearchSpace (kartography) - open source

ResearchSpace

ResearchSpace est un logiciel de gestion de données patrimoniales sous forme de données liées

Le projet est géré actuellement par l'entreprise à but non lucratif [Kartography](#) (Londres) dont les équipes sont liées au British Museum.

Le projet est né en 2009 sous l'impulsion de la Fondation Andrew W. Mellon qui souhaitait éviter que les projets institutionnels soutenus recommencent toujours à développer les mêmes outils.

La société Metaphacts a été mandaté pour élaborer la première version open source. Elle en a tiré ensuite sa suite commerciale metaphactory.

La Fondation SAPA a l'objectif de migrer ses données liées dans ResearchSpace en 2026.

pour plus détails : <https://researchspace.org/>

SAPA : coûts

L'utilisation des logiciels libres, ne veut pas dire automatiquement que les coûts sont moindres mais cela garantit que l'argent est proportionnel aux travaux (et non pas une rente de monopole).

Chez SAPA, nous dépensons grosso modo 75'000.- CHF par année auprès de nos différents prestataires externes pour gérer des données.

- 25'000.- CHF pour le maintenance des logiciels (corrections de bugs, changements de l'interface, etc.)
- 10'000.- CHF de frais d'hébergement des données
- 20'000.- CHF de maintenance des données (nettoyage, ajout)
- 20'000.- CHF de développement dans la réutilisation des données

Références

Voir aussi

- Baptiste de Coulon (2024a), "Déploiement de la norme Records in Contexts pour la gestion des collections de la Fondation SAPA". Revue électronique Suisse de Science de l'Information (RESSI), (24).
<https://doi.org/10.55790/journals/ressi.2024.e1511>
- Baptiste de Coulon (2024b) , [L'impact de l'utilisation des technologies du Web Sémantique pour des institutions. L'exemple de la Fondation SAPA](#), présentation au WebSem.pro 2024 (Paris).

Bibliographie

- Marcello Vitali-Rosati (2024) *Éloge du bug. Être libre à l'époque du numérique*, éditions La Découverte/Zones.
- Jérôme Denis, David Pontille (2022), *Le soin des choses. Politiques de la maintenance*, éditions La Découverte/SH Terrains philosophiques.



Merci de votre attention !