

## Déploiement de la norme *Records in Contexts* pour la gestion des collections de la Fondation SAPA

Baptiste DE COULON  
[baptiste.decoulon@sapa.swiss](mailto:baptiste.decoulon@sapa.swiss)  
Responsable plateforme en ligne, Fondation SAPA

### Résumé

*La Fondation SAPA, Archives suisses des arts de la scène, a migré en 2021 les métadonnées sur ses collections en RDF en suivant pour cela les recommandations du nouveau standard *Records in Contexts*. Cette migration a coïncidé avec la mise en ligne d'un nouveau portail de ses collections basé sur les principes des données ouvertes liées (*Linked Open Data*). Après deux ans d'implémentation, la Fondation a souhaité tirer un premier bilan afin que son expérience puisse servir aux institutions qui souhaiteraient suivre le même chemin.*

### Mots-clés

*Records in Contexts (RiC), Gestion des collections, Fondation SAPA, RDF (Resource Description Framework), Données ouvertes liées (Linked Open Data)*



Cet article est disponible sous licence [Creative Commons Attribution - Partage dans les Mêmes Conditions 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/).

## 1. Introduction

*« Dear, dear ! How queer everything is to-day! And yesterday things went on just as usual. I wonder if I've been changed in the night? Let me think: was I the same when I got up this morning? I*

*almost think I can remember feeling a little different. But if I'm not the same, the next question is,*

*Who in the world am I? Ah, that's the great puzzle!"*

*Lewis Carroll, Alice's Adventures in Wonderland*

Dans quelles mesures les bouleversements technologiques actuels nous contraignent-ils à remettre en question certaines de nos pratiques professionnelles ? Nos métiers et identités professionnelles sont-ils bien toujours les mêmes qu'hier ? Effectivement, ne nous sentons-nous pas un peu différent ce matin ?

Cet article tâche de rendre compte de ces questionnements sous la forme d'un premier bilan de l'adoption de la nouvelle norme de description archivistique *Records in Contexts* (par la suite RiC) par la Fondation SAPA. Cet article a été écrit avec l'espoir de permettre à d'autres institutions de préparer au mieux cette évolution.

*Remarques préliminaires : Nous avons tâché de ne conserver que les informations qui pourraient servir à d'autres. Il ne s'agit pas d'un bilan complet de nos choix. Nos analyses resteront brèves. Nous restons à disposition pour approfondir un aspect ou l'autre sur demande.*

## 2. Fondation SAPA

La Fondation SAPA, Archives suisses des arts de la scène a été créée en 2017<sup>1</sup>. Sa mission est de veiller à la bonne conservation des informations et des archives relatives au domaine des arts de la scène en Suisse. Elle ne vise pas à réaliser l'entièreté de cette tâche seule, mais agit aussi comme centre de compétences pour soutenir les institutions tierces dans cette mission. La Fondation SAPA est héritière de collections de plusieurs institutions précédentes dont elle a dû harmoniser les schémas de métadonnées afin de les rendre compatibles.

En 2021, la Fondation SAPA a mis en place un nouveau système de gestion de données archivistiques, ou *Archival Information System* (par la suite AIS), et un nouveau portail pour ses collections<sup>2</sup>. L'outil choisi pour ces deux fonctions<sup>3</sup> exploite de façon native les métadonnées formalisées en *Resource Description Framework* (par la suite RDF). Cela a été donc l'occasion d'implémenter la nouvelle norme RiC au travers de sa formalisation RDF, mais aussi de définir des formalisations RDF pour bien représenter les activités liées aux arts de la scène<sup>4</sup>. La Fondation SAPA est une des toutes premières institutions d'archives au niveau international à l'avoir fait dans ses outils de production.

---

<sup>1</sup> <https://sapa.swiss>

<sup>2</sup> <https://www.performing-arts.ch>

<sup>3</sup> Metaphacts Semantic Platform for Cultural Heritage and Digital Humanities, voir <https://hub.docker.com/r/metaphacts/glam-community>

<sup>4</sup> Voir à ce sujet les travaux de Birk Weiberg (Weiberg, 2020).

Le choix du langage RDF a été conseillé à la Fondation SAPA pour gérer l'héritage complexe des métadonnées hétérogènes rassemblées par ses institutions prédécesseuses.

La Fondation SAPA, dans le cadre de cette migration, a pu distinguer deux grands types de métadonnées à gérer : celles relatives aux documents de ses collections et celles relatives à ce qu'elle a appelé le «Panorama». Ce dernier regroupe les informations sur les personnes, les groupes, les événements, les œuvres ainsi que sur les liens qui les relie. La spécificité de SAPA réside dans le fait que ces deux ensembles de métadonnées sont d'une taille informationnelle équivalente. Environ 100'000 unités principales décrites chacune. Nous verrons que l'attention portée à ce Panorama est une donnée importante qui entraînera des conséquences sur notre formalisation RDF.

### 3. Les arts de la scène en RDF

Le domaine des arts de la scène dispose déjà de travaux sur sa modélisation sous forme RDF. Il existe en particulier un groupe de travail du 3WC sur le sujet<sup>5</sup> et une communauté très active dans ce domaine<sup>6</sup>. Leurs travaux intègrent aussi l'élaboration de bonnes pratiques vis-à-vis de leur modélisation au sein de la base de connaissances Wikidata<sup>7</sup>. Plusieurs institutions emblématiques de conservation de l'héritage culturel dans ce domaine comme la Pina Bausch Foundation ou la Merce Cunningham Trust ont d'ailleurs déjà implémenté des modélisations RDF de leurs informations (Bardiot, 2022). Toutefois, aucune n'a pour l'instant déployé également la nouvelle norme *Records in Contexts* pour représenter ses collections.

### 4. Records in Contexts

Les travaux en vue de l'élaboration d'une nouvelle norme de description archivistique ont été lancés par le Conseil international des Archives (par la suite ICA) lors de son congrès à Brisbane en 2012. Cette norme doit permettre de réunir en un seul modèle conceptuel l'ensemble des anciennes normes élaborées à partir des années 1990. Il a été décidé de suivre pour cela les recommandations sur le web sémantique (Gueguen et al., 2013) et d'utiliser donc une formalisation RDF (Francart, 2020). Cette nouvelle norme en cours d'élaboration a pris le nom *Records in Contexts*. Elle se décline sous deux formes, d'une part un modèle conceptuel (RIC-CM<sup>8</sup>) et d'autre part une ontologie (RIC-O<sup>9</sup>).

---

<sup>5</sup> Performing Arts Information Representation Community Group (PAIR-CG) : <https://www.w3.org/community/pair-cg/>

<sup>6</sup> Linked Open Data Ecosystem for the Performing Arts (LODEPA) : <https://linkeddigitalfuture.ca/fr/communaute/>

<sup>7</sup> WG6 Wikidata/Wikipedia

<sup>8</sup> <https://www.ica.org/en/records-in-contexts-conceptual-model>

<sup>9</sup> <https://www.ica.org/en/records-in-contexts-ontology>

## 5. Besoins identifiés

Les besoins identifiés par la Fondation SAPA qui ont déterminé le choix de RDF et RiC sont les suivants:

Une **adaptation aux exigences numériques actuelles**. Cela a été jugé comme indispensable pour gérer le basculement inexorable vers des versements de masse dans les services d'archives de documents nés numériques ou numérisés. Les volumes et les caractéristiques des documents numériques, même issus d'un bon système de gestion intégré des documents, demandent de disposer de nouveaux outils, dont la nouvelle norme de description est présentée comme l'élément de base. Cette adaptation repose sur les aspects suivants :

- Une **unification**. La nouvelle norme devrait permettre de réunir dans une formalisation commune l'ensemble des descriptions des composants d'un lot de documents. On compte parmi ces composants les éléments suivants : leur contenu, leur caractéristiques techniques, leurs prescriptions légales qui en déterminent l'accès et la conservation, les acteurs qui en gèrent la vie, les fonctions dont ils sont issus, mais également les événements qui rythment leur vie. Cette unification était aussi attendue dans notre cas pour fusionner les anciens modèles de données<sup>10</sup>.
- Une **automatisation**. La nouvelle norme devrait proposer un encodage technique des métadonnées qui permet d'automatiser leur création et leur gestion. Il devrait être suffisamment souple pour permettre de couvrir un maximum de situations. Certaines passerelles devraient aussi être prévues pour permettre aux archivistes et aux publics d'interroger les données de façon facilitée.

RDF et RiC devraient permettre de mieux gérer les processus d'**archivage numérique**. En particulier en intégrant la présence des copies des documents dans les flux de gestion des archives numériques.

Le cycle de vie d'un lot de documents ne se termine pas avec l'acquisition de leur statut d'archives. Leur conservation et leur diffusion induisent des activités de gestion supplémentaires et non totalement distinctes de leur phase de création. Ces activités doivent aussi pouvoir être couvertes complètement au même titre que les autres **phases de vie dans un continuum** de métadonnées (Kern et al. 2015).

L'adaptation aux exigences techniques actuelles permet aussi d'espérer une meilleure interconnexion entre les descriptions archivistiques et les données descriptives externes déjà disponibles. C'est-à-dire une **sortie de la logique des silos de données**. Cela aurait pour bénéfice de rationaliser les travaux, d'enrichir les descriptions et d'améliorer les interrogations croisées. On obtiendrait alors des données liées ou *Linked Data*.

Si les données descriptives sont adaptées au numérique et interconnectées avec d'autres données, alors il devrait être facile de permettre l'automatisation de leur interrogation, ainsi que leur **ouverture à la libre réutilisation**. On obtiendrait ainsi des données ouvertes liées ou *Linked Open Data*.

---

<sup>10</sup> On retrouve le même type d'ambition mais autour du modèle Linked Art dans le portail LUX de l'Université de Yale : <https://lux.collections.yale.edu/>

## 6. Premiers bilans

Après avoir résumé les attentes qu'ont pu susciter chez SAPA l'adoption de cette nouvelle norme et de son encodage technique, nous allons maintenant présenter le bilan que nous pouvons en tirer après une première phase de mise en œuvre. L'exercice est compliqué par le fait que certaines questions soulevées n'ont pas encore trouvé leur résolution dans la pratique.

## 7. Bilan du point de vue technique

La norme RiC s'exprime sous la forme d'une ontologie (RiC-O), or les **ontologies n'ont pas vocation à être simples**. Leur but est de réussir à représenter la réalité qui est par nature complexe. Au contraire des dialogues entre humains qui sont remplis de sous-entendus et de raccourcis - car nous partageons entre personnes interlocutrices un passé, une culture ou des savoirs - les ontologies, elles, sont des langages entre humains et machines. Les machines n'ont aucun passé.

Tout doit leur être explicite. C'est particulièrement laborieux pour des humains de lire et d'interroger strictement des ontologies, mais c'est le seul moyen de permettre aux machines d'opérer des raisonnements sur ces données (appelés *machine reasoning*).

Est-ce que la formalisation RDF d'une ontologie est utilisée uniquement pour l'échange de métadonnées ou est-ce un format d'exploitation interne de ces dernières ? C'est une question importante qui devrait être tranchée dès le départ. En effet, les **ontologies ne sont pas des formats de bases de données**. Elles ne peuvent avoir l'efficacité immédiate d'une formalisation sur mesure au sein d'une base de données, remplie de notions implicites. Les ontologies n'ont pas vocation à élaborer des outils simples et directement utilisables. Elles ont pour but de décrire de façon explicite une réalité pour que dans 20 ou 50 ans d'autres humains et d'autres machines puissent la comprendre. Il existe une tension entre les besoins d'efficacité d'un outil de travail et la volonté de recourir à des ontologies. La conjugaison des deux n'est pas impossible mais exige un réel travail d'innovation qui reste en grande partie à mener.

Les ontologies sont des objets complexes qui demandent un **savoir spécialisé**. Les institutions qui souhaitent mettre en œuvre un projet de migration de leurs données en RDF devraient disposer à l'interne de personnes formées à cet effet, ou du moins intéressées et prêtes à se former sur les ontologies.

Selon notre expérience, nous proposons de distinguer **trois grands ensembles** de métadonnées couverts par la norme RiC :

- les Record Resources (à savoir les descriptions des contenus des lots de documents);
- les Instanciations (leurs matérialisations) ;
- ce que nous avons appelé le «Panorama», c'est-à-dire les réseaux des groupes, d'événements, des règles et des personnes qui concourent à la création et à la gestion de ces documents.

Selon nos observations, la norme RiC et sa formalisation RDF sont particulièrement développées pour décrire les Record Resource. Cependant, pour les deux autres ensembles, il faut utiliser principalement d'autres normes comme CIDOC-CRM ou PREMIS.

En ce sens, il est important de comprendre que l'implémentation RDF de RiC, c'est-à-dire l'ensemble des classes et des propriétés définies par RiC (RiC-O), ne peut suffire à la description RDF d'un lot de documents d'archives. Il est nécessaire de recourir à un **assemblage d'ontologies**. Cela est voulu et souhaitable, mais un peu déroutant au départ. Adopter RiC veut dire employer RiC, mais aussi tout un ensemble d'autres ontologies. En effet, il est plus simple (et efficace) dans la mise en œuvre de données décrites selon RDF de ne pas recréer des classes ou des propriétés qui existeraient déjà ailleurs, mais de réutiliser ces dernières. Les schémas de métadonnées RDF sont donc souvent un assemblage d'ontologies plus ou moins spécialisées. Cela peut sembler nuire à l'unification des descriptions mais finalement les ontologies utilisées sont souvent les mêmes. Il en existe des génériques que l'on pourrait dire de haut niveau (RDFS, OWL, etc.) qui permettent d'exprimer les fonctions RDF principales telles que : cette entité «appartient à» (rdf:type) la classe d'entités unetelle. Il en existe d'autres très précises adaptées à un domaine spécifique (Ebucore<sup>11</sup> pour les données audiovisuelles, PREMIS<sup>12</sup> pour les activités de préservation numérique, etc.).

Un assemblage d'ontologies ne peut être facilement appréhendé, géré et utilisé que s'il est publié dans une **ontologie spécifique, ou maison**, sous forme OWL<sup>13</sup>. En ce sens, une bonne pratique voudrait que les différentes ontologies de référence (CIDOC-CRM, Frobroo, RiC, etc.) ne soient pas utilisées en direct dans l'assemblage, mais plutôt de façon indirecte grâce à la propriété <rdfs:subClassOf>. Ainsi, une ontologie maison serait à créer. Chez SAPA, il s'agit de spao (pour *Swiss Performing Arts Ontology*). Toutes les classes et les propriétés seraient créées dans cette ontologie et seraient reliées aux ontologies de référence par le biais de <rdfs:subClassOf>. Cela a l'avantage de permettre à l'ontologie maison d'être indépendante des mises à jour des ontologies de référence mais aussi de pouvoir se relier à plusieurs d'entre elles. Par exemple, la classe <spao:Person> aurait un lien <rdfs:subClassOf> avec <rico:Person> et avec <crm:E21\_Person>.

D'un point de vue technique, les ontologies et leur formalisation RDF nécessitent de mettre en place un système d'identifiants uniques (URI) pour toutes les entités décrites. Toutefois si ces identifiants sont nécessaires au fonctionnement interne des logiciels de gestion des descriptions RDF, leur mise à disposition au sein des portails des collections à destination des publics (et parfois des archivistes) est malheureusement encore trop rare. Or, renoncer à publier ces derniers et à assurer leur pérennité contrevient au projet d'intégrer les métadonnées archivistiques au Web sémantique et empêche leur réelle réutilisation par des tiers. La Fondation SAPA a donc décidé dans ce but de publier l'ensemble de ces identifiants et s'engage à les maintenir sur le long terme<sup>14</sup>. Cela lui a permis, par exemple, de les référencer à titre d'identifiants externes dans Wikidata<sup>15</sup>.

En conséquence des difficultés évoquées préalablement le marché **manque actuellement d'outils logiciels clé en main** pour éditer des descriptions d'archives qui répondent aux normes RiC et à sa formalisation RDF. Le problème de l'absence de production de nouveaux

<sup>11</sup> <https://tech.ebu.ch/metadata/ebucore>

<sup>12</sup> <https://www.loc.gov/standards/premis/>

<sup>13</sup> <https://www.w3.org/2001/sw/wiki/OWL>

<sup>14</sup> Elle utilise pour cela le système Universally unique identifier (UUID). Ex : <http://data.performing-arts.ch/r/198c11e9-95d7-434c-b8cd-b5dc021d15a3>

<sup>15</sup> Voir «identifiant SAPA» (P8974) : <http://www.wikidata.org/entity/P8974>

outils par des entreprises spécialisées va bien sûr se résorber<sup>16</sup>, mais la question de la mutualisation des coûts de développement et de la conservation de l'expertise en main publique dans ce domaine devrait être envisagée comme prioritaire dès à présent<sup>17</sup>.

SAPA a choisi d'utiliser un framework logiciel RDF générique publié sous licence libre<sup>18</sup>. Ce logiciel n'a pas de spécifications déterminées pour l'usage sur des corpus d'archives. Or, malgré les développements réalisés, nous n'avons pu construire pour l'instant qu'un outil aux fonctionnalités archivistiques simples. L'élaboration d'un outil de travail métier élaboré est un projet qui dépasse une institution unique. Cet effort devrait se réaliser de façon mutualisée.

La question de la distinction entre le AIS (*archival information system*) en lui-même et la salle de lecture en ligne est un aspect essentiel pour les développements dans ce domaine, et ce du fait de la nature même du RDF. Il est nécessaire de disposer de deux entrepôts de données RDF (appelé communément triplestore) distincts entre la production et la diffusion. En effet, il n'est pas aisé de cacher des informations RDF sans en établir une sélection préalable au moment de l'export des métadonnées entre un entrepôt et le second.

Le groupe de travail de l'ICA sur RiC<sup>19</sup> a proposé une formalisation RDF des données. Il s'agit d'une formalisation particulièrement adaptée aux données exposées sur le Web, mais le langage RDF n'est que peu utilisé par les services informatiques en général, ce qui induit un **accompagnement technique compliqué**. Jusqu'ici, le RDF a connu un intérêt plutôt dans le domaine académique. Il est en ce sens difficile de trouver des partenaires techniques compétents dans ce domaine. L'industrie du Web a utilisé plutôt d'autres technologies pour gérer les graphes<sup>20</sup> qui s'apparentent au RDF tout en s'en distinguant. Un des reproches avancés face à la formalisation RDF concerne des problèmes de performance dans les requêtes. En conséquence, il est compliqué de recourir à des services externes commerciaux pour accompagner les institutions dans ce domaine<sup>21</sup>.

Concernant l'archivage numérique, le choix de la norme RiC et de sa formalisation RDF demande encore à prouver son avantage pratique. Théoriquement ces avantages ont été prouvés, mais dans la pratique, la question du **lieu de stockage des métadonnées descriptives complètes** reste encore à trancher. La synchronisation de ces métadonnées entre les dépôts numériques et les bases de données de production reste un défi. À noter que le standard *Oxford Common File Layout* (OCFL<sup>22</sup>) semble ouvrir de belles perspectives par la possibilité de mise à jour incrémentale.

---

<sup>16</sup> Par exemple, l'entreprise suisse docuteam a annoncé la sortie de son futur AIS RDF, docuteam context, pour 2024.

<sup>17</sup> On mentionnera le très bel effort déployé en ce sens par l'Association Memoriav qui a publié sous forme ouverte le code de son nouveau portail de diffusion : <https://gitlab.switch.ch/memoriav>.

<sup>18</sup> Metaphacts Semantic Platform for Cultural Heritage and Digital Humanities, voir <https://hub.docker.com/r/metaphacts/glam-community>

<sup>19</sup> Expert Group on Archival Description (EGAD) : <https://www.ica.org/en/about-egad>.

<sup>20</sup> comme Neo4j et GraphQL.

<sup>21</sup> Nous avons la chance d'avoir trouvé un partenaire académique d'accord de le faire, à savoir : Swiss Art Research Infrastructure (SARI) : <https://www.sari.uzh.ch/en.html>. Il existe toutefois aussi quelques rares services commerciaux comme Zazuko (<https://zazuko.com/>) en Suisse ou Sparna (<http://www.sparna.fr/>) en France.

<sup>22</sup> <https://ocfl.io/>

Dans un cas comme le nôtre, nous utilisons déjà une dizaine d'ontologies en plus de RiC<sup>23</sup>, y compris une ontologie maison dite spao. Si nous avons entrepris un effort conséquent de documentation sur nos outils internes, la solution pour donner accès à cette documentation pour des personnes externes n'a pas encore été résolue. En ce sens, le choix de solutions efficaces pour offrir une documentation complète de nos modèles de données à l'intention des publics est une tâche à ne pas sous-estimer.

La démultiplication des champs possibles et l'atomisation de l'information avec RDF rendent compliquée l'élaboration de **formulaires de saisie** complets. L'ergonomie est parfois perdue vis-à-vis de l'exhaustivité. Des solutions graphiques seront certainement trouvées à l'avenir pour résoudre ces problèmes, mais celles-ci n'enlèveront pas la complexité de la mise en œuvre et de la mise à jour.

Plusieurs **solutions fonctionnelles restent à trouver** dans les années à venir pour permettre une réelle utilisation de la norme RiC. Il existe par exemple une difficulté technique pour supprimer une notice RDF avec ses nombreux liens. Cela demande de mettre en place des processus de validation conséquents. Par ailleurs, la publication automatisée d'inventaires PDF à partir des entités RDF reste encore à réaliser. Le partage des scripts de réalisation dans ce domaine serait bienvenu. Enfin, la création d'interfaces de recherche avancées sur des métadonnées RDF qui seraient adaptées aux personnes non expertes est aussi un défi important qui reste à relever<sup>24</sup>.

L'entièreté des métadonnées RDF sur nos collections est à présent exposée sous licence CC0<sup>25</sup> et accessible via une interface de requêtes avancées<sup>26</sup> (qu'on appelle dans le langage technique une SPARQL *endpoint*). Il s'agit, à notre connaissance, d'une première sur l'ensemble des données d'un portail d'archives institutionnel<sup>27</sup>. Toutefois, cette offre n'a été que peu utilisée, outre une utilisation spécifique sur laquelle nous reviendrons. Nous avons identifié trois **freins à l'ouverture et la réutilisation de nos données**.

Le premier, qui est le plus petit, est qu'il faut apprendre le langage de requête SPARQL. Celui-ci peut être acquis avec plaisir et intérêt par les personnes qui disposent de temps pour le pratiquer régulièrement. Il se révèle toutefois difficile d'accès pour tous ceux et celles moins à l'aise avec l'informatique, ou ayant moins de temps à y consacrer.

Le second est plus complexe. En effet, pour interroger la base, il faut en connaître le modèle, c'est-à-dire les choix des propriétés et des classes utilisées ainsi que leur relation/enchaînement. Cela renvoie à la question de sa documentation.

---

<sup>23</sup> Functional Requirements for Bibliographic Records object-oriented extension to the the CIDOC Conceptual Reference Model (FRBRoo), CIDOC Conceptual Reference Model (CIDOC-CRM), RDA Registry (RDA), Preservation Metadata: Implementation Strategies (PREMIS), Resource Description Framework Schema

(RDFS), Simple Knowledge Organization System (SKOS), EBU Core Metadata Set (EBUCore), Schema.org, Ordered List Ontology (OLO), Web Ontology Language (OWL), SAPA Ontology.

<sup>24</sup> Voir à ce propos les pistes intéressantes offertes par des outils comme <http://sparnatural.eu/>

<sup>25</sup> <https://creativecommons.org/share-your-work/public-domain/cc0/>

<sup>26</sup> <https://www.performing-arts.ch/sparql>

<sup>27</sup> Parmi les exemples en Suisse de mise en ligne RDF mais partielle, voir Plüss & Padlina (2022).

Et enfin, le dernier réside dans le fait que les entrepôts de métadonnées (*triplestore*) exposés aux publics doivent pouvoir supporter des montées en charge importantes. Or, réussir à lever ce dernier frein n'est pas sans conséquences financières.

## 8. Bilan du point de vue de l'archiviste

Un **manque d'une communauté de pratique**. La formalisation récente de RiC sous forme RDF<sup>28</sup> induit que la communauté des archivistes qui teste sa mise en œuvre est encore réduite, mais il semble que les choses changent à présent rapidement dans ce domaine heureusement<sup>29</sup>.

Un **bouleversement de la pratique archivistique**. L'adoption de la norme RiC est intimement liée à la pensée en termes de triplets réunis dans des graphes. Le choix d'une modélisation sous forme d'un tel graphe induit un certain nombre de bouleversements dans la pratique de la description archivistique. Ces changements n'ont pas été, faute de retours d'expérience, bien documentés jusqu'ici. Par ailleurs, selon nos observations, les premières ébauches d'outils logiciels de description archivistiques qui mettent en œuvre RiC tentent à camoufler ces changements pour ne pas trop perturber les archivistes dans leurs travaux. Reste que ces changements existent et qu'ils ont aussi un potentiel bénéfique dont il s'agira de tirer parti. On peut relever en particulier les aspects suivants :

- **Nécessité d'efforts pédagogiques**. La compréhension des mécanismes des triplets et des graphes RDF est presque indispensable pour bien saisir la portée de la nouvelle norme RiC qui s'en inspire. L'ensemble reste toutefois technique et la marche est parfois haute pour nos collègues. Or, puisque c'est un passage presque obligé, il nous manque encore de vrais outils pédagogiques pour le faire.
- La dualité «**Record Resource**» (par la suite RR) vs «**Instanciation**». La norme introduit une nouveauté fondamentale, à savoir la distinction entre la description du contenu du document (RR) et sa matérialisation physique (Instanciation). C'est un pas qui n'est pas facile à franchir au départ pour les archivistes mais qui facilite grandement l'exploitation technique des données par la suite. Un document est décrit une première fois dans son contenu puis les descriptions de ses différentes matérialisations sont rattachées à cette première description. Par exemple la version analogique d'un document et sa version numérique issue du processus de numérisation constituent deux instanciations distinctes qu'un même document. Cette multiplication des matérialisations peut sembler une complication vis-à-vis de certains documents textuels analogiques, mais elle est absolument nécessaire avec les documents nés numériques ou dans les cas de copies numériques de documents sur support instable (par exemple, les bandes audiovisuelles). Reste que la définition spécifique du contenu de la RR est un enjeu pour chaque institution.
- À noter, que la norme introduit aussi des **distinctions parmi les RR** en établissant des sous-ensembles au niveau de leur nature : `rico:RecordSet`, `rico:Record`, `rico:RecordPart`. Toutefois, ces distinctions ne sont pas simples à gérer dans la

<sup>28</sup> <https://www.ica.org/en/records-in-contexts-ontology>

<sup>29</sup> À noter la journée professionnelle qui lui est consacrée, le 15 septembre 2023, par l'Association des archivistes suisses.

pratique même si elles ont certainement du sens au niveau ontologique. En effet, les relations entre ces types d'entité ne sont pas les mêmes. Les Record sont inclus (rico:isOrWasIncludedIn) dans les Record Set mais les Record Part sont constitutifs (rico:isOrWasConstituentOf) d'un Record.

Pour des archivistes habitués à créer et à lire les descriptions des lots de documents sur une feuille ou plus vraisemblablement dans une base de données relationnelle simple où les informations ajoutées existent sous forme de textes rattachés à l'entité décrite, l'image des confettis est bien adaptée (quoique déstabilisante peut-être) pour expliquer la mise en œuvre technique de RiC via RDF. Dans une description sous forme de graphe on assiste à une **atomisation de l'information**. Chacune des informations ajoutées à une description est une entité séparée reliée à la description du lot de documents via une propriété. Les entités ne sont structurellement pas hiérarchisées entre elles. Elles ne le sont que conceptuellement via les classes principales de l'ontologie utilisée. Finalement, c'est comme si on passait une ancienne description d'archives dans un de ces broyeurs de documents dont ne ressortent que des petits bouts de papier. Toutefois, grâce aux graphes, c'est-à-dire aux liens entre les petits bouts d'information, le sens n'est pas perdu. Les analyses portent même à penser que le sens est enrichi puisque les relations sont bien plus explicitées que dans les anciens types de description (Crupi, 2012).

L'adoption des nouvelles normes et langages d'encodage augmente la technicité des descriptions et pousse inexorablement vers une automatisation des versements de métadonnées et une **diminution de la saisie manuelle**. Notre expérience a démontré que la majorité des ajouts de nouvelles descriptions se faisaient finalement par des imports en masse à partir de données préparées sous forme de tableur ou directement extraites par l'analyse numérique des fichiers. Le recours à la saisie manuelle via les formulaires intervient plutôt pour corriger des détails et des petites erreurs sans devoir repasser par un import. C'est un changement de pratique qui peut exiger le recours à des compétences informatiques pointues qui devront éventuellement être cherchées à l'extérieur de l'institution.

Du fait du recours aux imports de masse et à cause de l'atomisation de l'information dans un modèle de graphes, les doublons se multiplient rapidement. Il faut alors établir une stratégie de **contrôle qualité** et se donner les moyens de sa mise en œuvre. En effet, on peut décider de conserver deux entités avec des informations identiques là où cela n'importe que peu (par exemple des dates). Mais dans d'autres cas, comme celui des personnes, il convient mieux de les réconcilier, c'est-à-dire de les fusionner d'une manière ou d'une autre. Or, la fusion n'est pas une chose tout à fait évidente dans un modèle de graphes et cette fonctionnalité devra être la plus simple possible pour les archivistes afin de contenir cette part du travail qui peut vite devenir chronophage.

Une **interconnexion limitée** a été relevée entre des entités internes à nos bases et celles existantes dans des bases ouvertes externes. Cette interconnexion n'a vraiment du sens que dans le domaine que nous avons appelé le Panorama, ou dans celui des thesaurus par exemple sur les matériaux. En effet, comme les fonds d'archives ont vocation à ne réunir que des documents originaux et uniques, l'opportunité de lier une RR avec d'autres RR identiques externes est presque nulle. On voit ici toute la différence avec les bibliothèques.

L'adoption des nouvelles normes pousse à un renversement de la perspective archivistique. Elle donne un **rôle central au contexte**. En effet, ce ne sont plus les fonds qui tendent à être

au cœur des efforts descriptifs mais bien le contexte (qu'on appelle chez SAPA le Panorama), géré de façon séparée des collections, qui obtient un statut d'équivalence avec ses dernières. Ce mouvement est renforcé par le fait qu'il s'agit des informations dont nous disposons en partage avec les autres institutions et où les synergies de travail et les échanges de données sont le plus fructueux. En ce sens, ce qui apparaissait au départ comme une spécificité de notre Fondation, à savoir l'importance des données sur le Panorama, pourrait bien devenir la norme à l'avenir pour les autres institutions. Il y aurait en tous les cas un réel avantage à prendre cet aspect en compte dès le départ. Pour les archives publiques, il s'agira alors de développer les informations sur l'histoire administrative (Coutaz, 2021).

## 9. Perspectives

Sur la base du bilan que nous avons pu tirer de ces premières années de déploiement de la norme RiC et du RDF, nous pouvons à présent relever quelques perspectives particulièrement prometteuses offertes par ces changements.

La **mutualisation des travaux de description**. Il existe du fait de la centralité regagnée du contexte un potentiel déplacement de nos efforts descriptifs dans ce domaine vers les outils externes à nos institutions mais partagés avec d'autres. La base de connaissance Wikidata est actuellement l'outil privilégié pour cela (de Coulon, 2018). La description des entités de contexte (personnes, groupes, événements, etc.) pourrait à l'avenir se réaliser directement dans ce type d'outil de façon mutualisée. Cette option est un des chantiers passionnants qui s'ouvrent à nous. Cela pose en tout cas la question de la complémentarité et redondance de nos bases entre elles. Il s'agirait alors pour nos institutions de s'engager dans ce qui pourrait être nommé les *Crowd-sourced Linked Open Data*.

Un **exemple de réutilisation**. Malgré les difficultés évoquées, il faut tout de même relever que l'ouverture de nos données a déjà trouvé une application pratique qui nous incite à regarder cette fonctionnalité comme extrêmement positive. En effet, l'association MemoriaV a décidé de diffuser au travers de son méta-portail de recherche sur le patrimoine audiovisuel suisse, appelé Memobase<sup>30</sup>, l'ensemble des métadonnées que les institutions mettraient à sa disposition, et pas seulement celles relatives aux projets de sauvegarde qu'elle a soutenus. Dans ce cadre, les premières expérimentations d'export du choix de métadonnées recherchées via notre SPARQL *endpoint* démontrent que nous pourrions automatiser ces tâches d'échanges à l'avenir et s'épargner de fastidieuses transformations manuelles de ces dernières. La disponibilité d'un SPARQL *endpoint* donne par ailleurs aux archivistes (et aux publics) de **nouveaux supers pouvoirs** sous la forme d'un accès avancé et détaillé aux données archivistiques. Cela était, à notre connaissance, très rare, voire inexistant jusqu'ici. En ce sens, le passage à la norme ne devrait pas se faire sans exiger la mise à disposition, du moins à l'interne, d'un tel outil d'exploration.

---

<sup>30</sup> <https://memobase.ch/fr/start>

## 10. Conclusion

La nouvelle norme *Records in Contexts* portée par le Conseil international des Archives n'est qu'une brique d'une métamorphose plus large de nos descriptions d'archives sous forme RDF. Il s'agit d'un bouleversement de nos pratiques professionnelles qui dépasse la nouvelle norme en elle-même. Il existe un grand potentiel de valorisation et d'automatisation dans le traitement des métadonnées par ce biais, mais l'accompagnement au changement des équipes (et des publics) ne doit pas être sous-estimé. Il ne serait d'ailleurs pas honnête de promettre un passage sans heurt d'une manière de faire à une autre. SAPA est toujours en train de travailler à ce changement de pratiques. L'enjeu serait de réussir à tirer le maximum des opportunités offertes par les nouvelles avancées technologiques tout en minimisant les contraintes.

Avec le recul, trois éléments nous semblent fondamentaux pour réussir un projet de passage à RDF pour une institution d'archives : définir un assemblage d'ontologies sous forme d'ontologie maison et la publier sous forme OWL, offrir un accès interne et public aux métadonnées via SPARQL *endpoint* et utiliser un système d'URI pour les notices descriptives qui soit visible pour les publics également. Ces trois éléments garantissent que le projet dispose de vraies bases saines d'un point de vue technique. Reste ensuite à établir les outils de gestion et d'interrogation facilités de ces métadonnées. Là les défis demeurent immenses.

Ces défis ne seront relevés que si l'ensemble de nos équipes y trouve aussi son compte. Or, seul l'engagement général de notre communauté professionnelle dans ce sens pourra nous permettre d'y arriver. Nous comptons donc sur vous !

## 11. Remerciements

Dans cet article, il n'a pas été possible de signaler l'apport précis de chacune et chacun. Nous tenions toutefois à remercier Birk Weiberg qui, avec un formidable talent et une grande passion, a créé la plus grande partie du RDF avec lequel nous travaillons aujourd'hui. Nous tenions aussi à mentionner l'apport de Beat Estermann qui dans ses fonctions de conseils a préconisé dès 2014 l'usage du RDF et également Julien A. Raemy qui a su par ses exemples donner foi dans les options prises. Ce texte a heureusement pu bénéficier d'une relecture critique très précieuse de la part de Roberta Padlina dont le savoir sur les ontologies dépasse largement le nôtre. Par ailleurs, la qualité de ce texte ne serait certainement pas la même sans les remarques judicieuses de Beate Schlichenmaier, Céline Bösch et Simona Generelli. Nous profitons pour les en remercier chaleureusement.

Finalement, nous aimerions remercier vivement l'ensemble de nos collègues qui ont bien dû nous accompagner dans ces implémentations mais qui ont su, malgré quelques passages difficiles, garder foi dans notre capacité de mener à bien ce projet. Merci également à notre Conseil de Fondation qui a su assumer cette prise de risque. Nous tâchons dans notre travail quotidien d'être à la hauteur de la confiance que toutes et tous nous ont témoignée.

## Bibliographie

- Bardiot, C. (2022). Comment transmettre l'héritage des arts de la scène ?. Dans B. Bohet et V. Pringuet (dir.), *Les devenirs numériques des patrimoines* (p. 303-315). Éditions de la Maison des sciences de l'homme.
- Coutaz, G. (2021, 26 février). L'histoire administrative, la grande oubliée de la gouvernance de l'information. *Archivalise(s) ; Archivistique et diplomatique des données et des documents*. <https://archivalises.hypotheses.org/651>
- Crupi, G. (2012). Universo bibliografico e semantic web. Dans F. Ciotti et G. Crupi (dir.), *Dall'Informatica umanistica alle culture digitali, Quaderni DigiLab* (vol. 2, p. 277–306). Università Sapienza di Roma.
- de Coulon, B. (2018). Archives et Wikidata. *Arbido*, 2018(2). <https://arbido.ch/fr/edition-article/2018/automatisation-versprechen-oder-drohung/archives-et-wikidata>
- Francart, T. (2020, 20 avril). RiC-CM, RiC-O : les Archives se dotent de leur modèle conceptuel (Records in Contexts). *Blog de SPARNA*. <http://blog.sparna.fr/2020/04/20/rico-records-in-contexts-archives-modele-conceptuel/>
- Gueguen, G., Marques da Fonseca, V., Pitti, D., et Sibille-de Grimoüard, C. (2013). Vers un modèle conceptuel international pour la description archivistique. Conseil international des archives. [https://www.ica.org/sites/default/files/EGAD\\_French.pdf](https://www.ica.org/sites/default/files/EGAD_French.pdf)
- Kern, G., Holgado, S., Cottin, M. (2015). Cinquante nuances de cycle de vie. Quelles évolutions possibles ?. *Les Cahiers du Numérique*, 11 (2), 37-76. <https://doi.org/10.3166/lcn.11.2.37-76>
- Plüss, R. & Padlina, R. (2022). Wissensnetz der Zürcher Ehedaten des 16.–18. Jahrhunderts: Eine Anwendung von Semantic-Web-Technologien im Archiv. *ABI Technik*, 42(4), 230-241. <https://doi.org/10.1515/abitech-2022-0043>
- Weiberg, B. (2020). Modeling Performing Arts: On the Representations of Agency. *Arti dello Spettacolo / Performing Arts*, 6, 50-56. <http://dx.doi.org/10.17613/6tsm-4787>