



RASAD 2017

Les fondamentaux de l'archivistique : un génome en évolution ?

Montpellier, 2-3 février 2017

Le projet de modèle conceptuel des archives, ou l'unification des normes de description

Hélène Zettel

Introduction

Je vais vous présenter le projet de modèle conceptuel RiC pour *Records in Contexts*, modèle qui a été développé par le groupe d'experts pour la description archivistique au sein du Conseil international des archives avec la collaboration en tant que membre pour la France le SIAF et les AN en la personne de Florence Clavaud.

1. poser des éléments de contexte
2. présenter le modèle conceptuel lui-même
3. apports et perspectives apportés par RiC

1. Contexte et objectifs

En 2012, le Conseil international des archives a constitué le groupe EGAD (dont l'acronyme développé signifie « Groupe d'experts sur la description archivistique »). Le mandat d'EGAD, qui courait de 2012 à 2016, était de travailler à l'élaboration d'une norme de description archivistique qui réconcilie et intègre les 4 normes de description actuelles de l'ICA, tout en s'appuyant sur elles.

Ces normes avaient pour partie fait l'objet de révision, afin de s'adapter aux évolutions permanentes des technologies de communication qui constituent le fondement des systèmes de description archivistique.

Ces normes souffrent cependant d'un manque d'articulation entre elles. Elles ont été élaborées à la suite les unes des autres : leur développement s'est donc étalé sur une période longue courant de 1994 pour la première version de ISAD(G) à 2008 pour ISDIAH, et sans qu'une vision d'ensemble prévale dès l'origine. Il y a bien une idée de lien entre elles mais qui s'articule mal : ainsi, il est prévu un lien entre ISAAR(CPF) et ISAD(G), mais uniquement dans un seul sens.

À ce manque de cohérence s'ajoute aussi une utilisation très inégale des quatre normes : ISAD(G) est de loin la norme la plus répandue, dû au fait qu'elle reflète la pratique descriptive ancienne et prédominante consistant à faire une description unique, centrée sur les archives et basée sur la provenance. Des trois autres normes, qui concernent plus le contexte de production et de conservation et qui doivent compléter ISAD(G), ISAAR-CPF connaît une certaine utilisation, tandis ISDF et ISDIAH qui en découlent sont peu répandues.

Par ailleurs, le développement de ces normes s'est fait en cohérence avec les technologies de diffusion qui prévalaient dans les années 1990-2000, à savoir les langages relationnels d'une part, les langages de balisage d'autre part. Ils sont le reflet d'une description conçue comme une transposition dans le monde numérique d'habitudes issues du papier, envisageant systématiquement la présentation des instruments de recherche de façon linéaire.

De plus, la pratique a fait ressortir la limite de ces modèles, dans le sens où le modèle hiérarchique ne permet de rendre qu'imparfaitement la complexité du monde, limitant par conséquent la richesse contextuelle.

Or, il existe depuis les années 1960 des technologies qui permettent de représenter les données en graphe, notamment via le langage RDF diffusé à partir des années 1990. Ce langage permet une représentation des données descriptives beaucoup plus souple et plus expressive, à condition de s'assurer de la qualité et de l'uniformité de ces données.

Le mandat donné à EGAD était donc d'établir une nouvelle norme de description qui concilie et intègre les quatre standards actuels, tout en se basant sur les possibilités offertes par le modèle de représentation en graphe afin que les données archivistiques puissent être mieux exploitées et plus visibles, notamment à l'heure du web sémantique.

EGAD a pu s'appuyer sur d'autres travaux de modélisation en cours dans d'autres communautés patrimoniales proches comme les bibliothèques (modèle FRBR) et les musées (modèle CIDOC-CRM). Il a pu également s'inspirer des travaux d'autres communautés archivistiques à l'échelle nationale comme l'Australie et la Finlande.

Une première version du modèle conceptuel de RiC a été publiée par EGAD en septembre 2016 pour appel à commentaires qui s'est clôturé mardi dernier. Des ateliers organisés par le SIAF et les AN ont permis de recueillir les commentaires pour la communauté française.

Dans un second temps, le modèle conceptuel sera complété par une ontologie qui sera sa traduction technique, et dont la première version devrait paraître pour commentaire en mars prochain.

Nous en sommes encore aux premières étapes de l'élaboration de cette nouvelle norme, qui nécessitera également des solutions d'implémentation qui permettront aussi la transition technique d'ISAD(G) vers RiC-CM.

2. Présentation du modèle conceptuel

Le modèle conceptuel de RiC repose sur l'agencement d'éléments descriptifs fondamentaux réutilisables à loisir. Le modèle a donc cherché à définir les éléments fondamentaux de la description archivistique à la plus petite unité possible, à donner les caractéristiques de ces éléments et à faire des liens entre eux.

Il distingue donc trois éléments :

- -les **entités**, qui sont les éléments fondamentaux de la description,
- -les **propriétés**, qui sont les attributs qui qualifient ces entités,
- -les entités sont agencées entre elles par un système de **relations**.

Si l'on file la métaphore de la molécule qui donne le titre à cette intervention, on pourrait dire que les entités sont des atomes que l'on relie entre elles par des liaisons chimiques que seraient les relations. L'ajout de nouvelles relations avec d'autres entités peut se faire dans de multiples directions : c'est pourquoi on parle désormais de description multi-dimensionnelle. La souplesse de RiC permet donc de rendre la richesse et la complexité du contexte d'évolution du document.

2.1. Les entités sont au nombre de quatorze. Les documents en eux-mêmes peuvent entrer en trois catégories :

- le *Record*, le document
- le *Record set*, le groupe de documents. C'est dans cette entité que les différents niveaux de description peuvent être compris, puisque les *Record sets* peuvent s'emboîter.
- le *Record component*, le composant du document, c'est-à-dire un élément distinct du document puisqu'il peut avoir ses propres caractéristiques, mais qui en constitue un élément indivisible. Par exemple, le sceau pour la charte.

On constate déjà une première distinction introduite par RiC entre le document (la pièce) et le fonds. Dans ISAD(G), le document est considéré comme le plus petit niveau de description, et est donc décrit de la même manière que les autres niveaux hiérarchiques du fonds. Dans RiC, le document ou *Record* est une entité distincte qui, certes, partage certaines propriétés avec le groupe de documents ou *Record set*, mais a certaines en propre. Il est par exemple possible qu'un *Record* ait un producteur différent que le *Record Set* dans lequel il est contenu.

L'autre distinction notable est le fait qu'un même *Record* puisse appartenir à plusieurs *Records Sets* à la suite ou simultanément. On passe ainsi d'une description hiérarchique, allant du général au particulier, du fonds à la pièce, à une description multi-dimensionnelle dans laquelle la description hiérarchique multi-niveaux d'un fonds n'est plus qu'une possibilité.

On entend par *Agent* une personne ou un groupe qui est responsable de ses actions et de leurs conséquences. Cela englobe les producteurs d'archives mais peut aller bien au-delà. La description des Agents est plus riche et plus qualifiée que ce que l'on peut faire actuellement, avec de nouvelles entités. On distingue ainsi la Position, le titre, de l'Occupation ou le métier.

RiC accorde une place importante au domaine des fonctions, qui regroupe :

- l'*Activity* ou action
- la *Function* (fonction), qui se distingue du domaine fonctionnel (*function abstract*). Le modèle considère qu'il y a d'une part les fonctions qui sont spécifiques au contexte et d'autre part des fonctions d'ordre général, indépendantes du contexte socio-culturel (par ex : promouvoir l'archivage numérique en France versus définir la politique archivistique).

Enfin, un certain nombre d'entités sont communes à celles définies par d'autres communautés et peuvent se traduire sous forme de vocabulaire normé : la typologie documentaire, la date, le lieu et le concept. RiC a souhaité en faire des entités, considérant que ce sont des concepts qui peuvent évoluer dynamiquement dans le temps et qu'ils nécessitent d'avoir leurs propres propriétés.

2.2. Les propriétés sont trop nombreuses pour être toutes citées. Elles peuvent être catégorisées selon trois groupes :

- les propriétés communes à toutes les entités,
- les propriétés communes à plusieurs entités,
- les propriétés propres à une entité.

Elles sont principalement issues des normes de description habituelles et peuvent faire l'objet de vocabulaires contrôlés ou non.

2.3. Les relations

Dans la version actuelle du modèle, la liste des relations se présente sous forme d'un long tableau qui liste toutes les relations envisageables entre deux entités avec chaque fois la relation inverse (est parent de / est enfant de).

C'est la partie la moins aboutie du modèle et qui n'a pas encore trouvé sa forme de présentation définitive. C'est aussi là toute la richesse du modèle.

3. Quels sont les apports de RiC ?

Les relations constituent l'apport majeur de RiC, bien que la zone des relations apparaît déjà dans les normes actuelles. L'enjeu est de caractériser plus finement les relations et non de les laisser « libres ».

L'analyse du modèle conceptuel montre que la plupart des champs que comptent les quatre normes de description trouvent leur équivalent au sein des entités et des propriétés de RiC. L'évolution tient plutôt dans l'éclatement des différents éléments de la description, qui sont gérés de façon indépendante et peuvent être mis en relation afin de construire la description. Cette idée de séparer les éléments de la description n'est pas nouvelle puisque les normes existantes en ont jeté les bases, mais le modèle conceptuel de RiC pousse la logique dans cette direction. Mieux structurées, les données de description peuvent être plus facilement exploitées par des machines et ouvrir à de nouvelles connaissances et de nouveaux modèles.

Avec RiC, les principes de base de l'archivistique sont respectés : le principe de provenance, le respect des fonds, la description contextualisée.

La souplesse introduite par RiC doit permettre d'aller plus loin et de mieux refléter la richesse des interactions entre les individus et les groupes, d'exprimer le mieux possible le contexte qui entoure la naissance et la vie du document, par exemple en documentant les lacunes d'un fonds ou les sélections faites volontairement ou au cours du temps. Ou, autre exemple, un document qui a été produit par une administration peut se trouver conservé dans le fonds d'un autre service. Le *Record* pouvant être lié à plusieurs *Records sets*, le document en question pourra être rattaché à la fois au service qui l'a produit et au service qui l'a versé. Cette souplesse dans la modélisation permet de respecter et de dépasser le principe de provenance.

L'objectif est aussi de mieux prendre en compte la description des archives numériques pour lesquelles la notion de dossier a tendance à éclater. Par exemple, au sein d'une même base de données peuvent se trouver réunies des données produites par des services différents.



RASAD 2017

Les fondamentaux de l'archivistique : un génome en évolution ?

Montpellier, 2-3 février 2017

Présentation du référentiel des producteurs de l'administration territoriale de 1800 à nos jours

Hélène Zettel

Origines du projet et objectifs

Lors des RASAD de 2009 a été évoqué le projet de création d'un référentiel national des notices de producteurs de l'administration territoriale.

L'idée était que, face à des services que l'on peut rencontrer à l'identique dans chaque collectivité, le réseau aurait tout intérêt à regrouper ses efforts, d'une part pour recenser les services producteurs et leur donner des formes autorisées, et d'autre part, pour rationaliser en mettant en commun la rédaction de notices d'autorité pour ces producteurs.

=> Une convention est signée en 2010 entre le SIAF et l'AAF.

Dans un premier temps, mise en ligne de formes autorisées du nom des collectivités à travers un référentiel.

Dans un second temps, commencer la rédaction de notices d'autorité génériques pour les producteurs identifiés.

Outils et méthodologie

Composition du groupe.

Réunions et méthodologie :

1 à 2 séances par an en présentiel afin d'acter les évolutions essentielles du référentiel.

Le reste des échanges se fait sous forme de mails.

Les travaux du groupe ont permis de publier la version 1.4. du Référentiel en janvier.

À ce jour, 200 notices sont en ligne.

Outils :

- guide pour la saisie des notices d'autorité sous AtoM : sera prochainement l'objet d'une révision (http://www.archivistes.org/IMG/pdf/manuel-atom_2012_octobre_2014.pdf)
- la plateforme AtoM : <https://aaf.ica-atom.org/>

Un formulaire permet la saisie et la publication en ligne de notices d'autorité conforme à la norme ISAAR(CPF). L'export en xml-eac est possible afin d'instancier et de réutiliser la notice générique dans son propre système d'information archivistique.