
AtoM 3: Étude de faisabilité Proposition

Préparé par:

AIM 25 (Royaume-Uni)
Artefactual systems (Canada)
Docuteam (Suisse)
Imagiz (Royaume-Uni)
Nothing Interactive (Suisse)
Zazuko (Suisse)

20 Septembre 2018

Ceci est une version révisée de la proposition (28 novembre 2018). Le conseil d'administration de la Fondation Access to Memory (AtoM) a décidé d'omettre les sections consacrées aux phases de développement ultérieures et aux estimations de coûts afin que la proposition se concentre, lors de la première consultation auprès la communauté, sur les principes de conception d'AtoM 3.

Table des matières

Résumé analytique	3
Structure du document	4
1. Pourquoi remplacer AtoM 2?	5
2. Principes de conception pour AtoM 3	5
3. Qu'est-ce qu'une étude de faisabilité?	7
4. Construire une étude de faisabilité: description détaillée	7
4.1 Exemples d'utilisateurs cibles	7
4.2 Cas d'utilisation 1: convertir les descriptions archivistiques existantes en Web des données	8
4.3 Cas d'utilisation 2: permettre d'améliorer la description archivistique	9
4.4 Cas d'utilisation 3: fédérer et exposer les données	10
4.5 Tâches pour l'étude de faisabilité	11
4.6 Evaluation de l'étude de faisabilité	13
5. Architecture technique proposée (description de haut-niveau)	13
Annexe A: Argumentaire pour le web des données	15
Un meilleur partage des informations	15
Agrégation et mises à jour plus faciles dans les portails multi-référentiels	15
Plus de contexte, plus de sens, plus de nuances	16
Un meilleur soutien pour la normalisation	17
Répondre aux besoins avec AtoM 3	17
Annexe B: Architecture technique proposée (détails)	18

Résumé analytique

Ce document présente une proposition concernant l'établissement d'une étude de faisabilité pour une application logicielle qui succéderait à Access to Memory 2.x (AtoM 2). AtoM 2 est une application internet open source multilingue, conçue pour permettre la description et l'accès virtuel à des fonds d'archives¹. Articulé autour des normes de description du Conseil International des Archives (ISAD(G), ISDIAH, ISAAR (CPF) et ISDF²), AtoM 2 a été mis en oeuvre dans le monde entier, pour être utilisé soit par des institutions individuelles (un site par institution) soit par plusieurs établissements au niveau régional et national (portails multi-référentiels).

Du 23 au 25 août 2018, un groupe d'organisations vouées au développement et à la prise en charge de logiciels open source s'est réuni au King's College de Londres pour discuter des caractéristiques et des besoins techniques de ce que l'on appelle actuellement AtoM 3. Ce document est le résultat de cet atelier de trois jours et se présente sous la forme d'une demande de financement pour l'étude de faisabilité. Les représentants des organisations suivantes ont participé à l'atelier et ont écrit conjointement ce document:

AIM25 (Royaume-Uni)

AIM25 est un organisme de bienfaisance qui soutient l'accès à plus de 150 institutions archivistiques dans la région de Londres, dont des autorités locales, des universités et des organismes savants et scientifiques tels que la Royal Society et la Wellcome Library. Il regroupe et permet une recherche croisée des descriptions des fonds d'environ 17 000 archives conservées par ses membres, et gère également le thesaurus archivistique britannique. AIM25 a dirigé et collaboré à quatre projets de Web des données utilisant des données archivistiques. <https://aim25.com/>

Artefactual Systems (Canada)

Artefactual est spécialisé dans le développement et le soutien de logiciels open source dans le secteur du patrimoine culturel. La société est le principal développeur d'AtoM 2 et, forte de 10 archivistes, possède une vaste expérience des pratiques archivistiques et de la mise en oeuvre de normes de description archivistique. <https://www.artefactual.com/>

Docuteam (Suisse)

Docuteam soutient un grand nombre d'organisations dans leurs activités de gestion des informations et d'archivage, couvrant tout le cycle de vie de l'information. La société développe également des logiciels open source pour la conservation numérique des archives et des bibliothèques de recherche. <https://www.docuteam.ch/fr/>

Imagiz (Royaume-Uni)

Imagiz est une société de conception et de développement de sites Web possédant une vaste expérience dans l'aide apportée aux organisations culturelles et universitaires pour améliorer la gestion, la disponibilité et la présentation de leurs archives. Imagiz se spécialise dans les migrations de données et le

¹ <https://www.accesstomemory.org>

² <https://www.ica.org/en/public-resources/standards>

support technique d’AtoM, créant et gérant de grandes collections en ligne pour un certain nombre d’institutions britanniques de premier plan. <https://imagiz.com/archios/>

Nothing Interactive (Suisse)

Nothing Interactive est une société de conception de sites Web qui appuie ses clients dans le domaine de la conception centrée sur l’utilisateur, afin d’optimiser l’expérience de ces derniers avec le produit numérique. Nothing Interactive s’efforce d’adopter une approche holistique et offre une expertise en matière de concept, de conception et de développement. Nothing Interactive collabore notamment avec le Parlement suisse pour promouvoir l’éducation et aide des éditeurs suisses comme Klett dans leur transformation numérique. Nothing Interactive prend également en charge et contribue à différentes technologies open source permettant une meilleure collaboration et / ou facilitant la gestion de contenu. <https://www.nothing.ch/fr>

Zazuko (Suisse)

Zazuko, fondée en 2014, est une société de conseil pour le Web sémantique et les technologies associées. La société est en discussion avec diverses agences gouvernementales suisses telles que le Secrétariat d’Etat à l’économie, les Archives fédérales, l’Office fédéral de la statistique et l’Office fédéral de la topographie, ainsi que des entreprises clientes des secteurs de la communication, de la navigation et des sciences de la vie. Zazuko est co-créateur du paquet open source RDF.js pour JavaScript, en collaboration avec le MIT CSAIL (sous la direction de Sir Tim Berners-Lee) et IDLab de l’Université de Gand. <http://zazuko.com/>

Structure du document

Les auteurs de ce document proposent de concevoir AtoM 3 comme une plateforme capable de prendre en charge le Web des données tout en maintenant les fonctionnalités d’AtoM 2. Le “Web des données” permet de structurer les données de manière à ce que des sources d’informations disparates puissent être liées entre elles de manière compréhensible par les ordinateurs. Cela permet une recherche, une agrégation et une visualisation des données beaucoup plus robustes, ce qui facilite la mise à jour des portails multi-référentiels et permet aux archives d’utiliser les ressources en ligne existantes telles que les autorités et vedettes matières. Voir l’[Appendix A](#) pour plus d’informations sur le Web des données et la [section 2](#) pour les principes de conception proposés pour AtoM 3.

Les auteurs cherchent des fonds pour construire l’étude de faisabilité d’AtoM 3. Une étude de faisabilité, ou EF, est un projet conçu pour démontrer la faisabilité du développement d’une application logicielle afin de répondre aux besoins de parties prenantes définies. Pour ce faire, il vérifie qu’un ensemble de fonctions logicielles souhaitées fonctionnera comme prévu une fois implémenté avec l’architecture technique et l’ensemble des outils choisis. Voir la [section 3](#) pour la définition plus détaillée d’une EF.

L’EF proposée est décrite comme un ensemble de personnages d’utilisateurs, de scénarii d’utilisation, de scénarii utilisateurs et de tâches. La [Section 4](#) écrit en détail tous ces aspects de l’EF et définit également le rôle des testeurs dans l’évaluation du succès de l’EF. La [Section 5](#) propose une architecture technique de haut niveau conçue pour atteindre les capacités fonctionnelles envisagées pour AtoM 3, qui sera testée

lors de l'EF; l'[Appendix B](#) fournit un aperçu plus détaillé de l'architecture et de l'infrastructure technologique proposées.

1. Pourquoi remplacer AtoM 2?

AtoM 2 est une version réorganisée d'ICA-AtoM, dont le développement a commencé en 2006. ICA-AtoM a été introduit en tant que système de production en 2008 et est toujours utilisé dans certaines institutions à travers le monde. AtoM 2 a été conçu pour prendre en charge une plus grande évolutivité, de meilleures capacités de recherche et une interopérabilité des données plus robuste, mais il ne s'agissait pas d'une refonte complète du système. ICA-AtoM et AtoM 2 sont donc des applications logicielles conçues il y a plus de 12 ans, une longue période dans le monde du développement logiciel. Depuis lors, de nouveaux cadres de développement, bibliothèques logicielles et outils ont été introduits et affectent la conception des applications développées pour le Web; l'infrastructure technologique de base qui compose AtoM 2 est devenue obsolète et difficile à maintenir.

Outre les avancées technologiques, les modifications apportées aux concepts de base relatifs à la création, l'affichage et l'exposition des données via des interfaces de recherche ont entraîné des modifications fondamentales dans la manière dont une application comme AtoM pourrait répondre aux besoins des services d'archives. En 2016, le groupe d'experts sur les descriptions archivistiques du Conseil International des Archives (ICA) a publié *Records in Contexts (RiC)*³, un modèle conceptuel de données couplées pour la description archivistique, conçu pour intégrer et remplacer les normes de description archivistique existantes de l'ICA. Ce groupe d'experts travaille actuellement sur une ontologie pour le modèle. La publication de RiC et les développements d'autres normes et d'ontologies pour décrire les ressources indiquent des changements fondamentaux dans la manière dont les archivistes envisagent la manière dont les archives sont décrites et mises à la disposition des utilisateurs. Les applications logicielles conçues pour la description archivistique doivent s'adapter à ces évolutions.

2. Principes de conception pour AtoM 3

L'un des principes de conception fondamentaux d'AtoM 3 est l'intégration des fonctionnalités actuelles d'AtoM 2 mais offrant une meilleure expérience utilisateur et permettant aux organisations de partager des informations via l'utilisation du Web des données. AtoM 3 doit également s'assurer que les données d'AtoM 2 puissent être importées dans AtoM 3 avec une relative facilité. De plus, bien que le Web des données constitue un nouveau paradigme pour la description archivistique, les auteurs de cette proposition sont parfaitement conscients du fait que la plupart des archivistes ne veulent pas nécessairement apprendre de nouveaux modes de description ou avoir à abandonner des travaux antérieurs pour pouvoir utiliser un nouveau logiciel. Gardant cela à l'esprit, nous avons répertorié ci-dessous certaines des fonctionnalités clés d'AtoM 2 et considérons qu'il s'agit des fonctionnalités fondamentales que son successeur doit prendre en charge.

³ <https://www.ica.org/fr/records-in-contexts-ric-les-archives-dans-leur-contexte-presentation-congres-2016>

- **Standardisé:** possibilité de créer des descriptions archivistiques à l'aide de modèles basés sur les normes de description de l'ICA et sur des normes supplémentaires telles que les Règles canadiennes pour la description des documents d'archives, Dublin Core et MODS
- **Application web:** tous les utilisateurs finaux ont besoin d'un navigateur Web et d'une connexion Internet, ce qui signifie qu'ils peuvent accéder à leur site et l'éditer depuis n'importe quel ordinateur
- **Affichage de l'objet numérique:** dans le cadre des métadonnées descriptives - l'utilisateur peut naviguer de la miniature à la version pleine résolution (téléchargeable si autorisé)
- **Taxonomies hiérarchiques** modifiables
- **Multilingue:** capacité à coordonner les traductions de la communauté pour chaque version; capacité à traduire le contenu et les éléments de l'interface utilisateur
- **Interface configurable:** de nombreux éléments de l'interface utilisateur peuvent être facilement modifiés (libellés de l'interface utilisateur, menus, termes contrôlés)
- **Affichage hiérarchique:** possibilité de générer et de parcourir l'arborescence complète des descriptions d'archives
- Possibilité de créer des **thèmes personnalisés**
- Possibilité de rédiger des **notices d'acquisition** et d'y associer des descriptions archivistiques
- Capacité à séparer les **notices de description de celles relatives aux acteurs** et à intégrer les informations pertinentes relatives aux acteurs dans les descriptions archivistiques
- Capacité à prendre en charge **plusieurs dépôts d'archives**
- Possibilité d'**importer et d'exporter en bloc** les notices de descriptions, d'autorités et celles relatives aux dépôts en formats XML ou CSV via l'interface utilisateur

AtoM 3 devrait prendre en charge les fonctionnalités décrites ci-dessus, ainsi que les suivantes:

- **Prise en charge du Web des données:**
 - **Sérialisation des descriptions archivistiques en RDF⁴**
 - Capacité à établir **un lien** sémantique avec des **ressources externes** telles que les vedettes d'autorités, les vocabulaires contrôlés et les fonds d'archives d'autres institutions
 - Possibilité pour les utilisateurs d'**interroger et d'afficher de manière dynamique le contenu** de sites AtoM disparates et d'autres générateurs de Web des données
 - Possibilité d'**exposer les descriptions à des agrégateurs de données** tels que Digital Public Library of America, Europeana, ArchivesHub, ArchivesCanada, etc.
 - Possibilité d'intégration d'**outils de visualisation** RDF pour prendre en charge géocartographie, histogramme et autres représentations de données graphiques
- **Continuité:** modèles et interfaces utilisateur familiers aux archivistes, et possibilité d'importer des descriptions d'archives à partir d'AtoM 2 et d'autres outils de création EAD
- **Fédération des données:** séparation du front-end et du back-end, permettant d'afficher le contenu de plusieurs instances principales dans un seul portail front-end

⁴ RDF signifie Resource Description Framework, modèle type d'échange de données sur le Web utilisant les triplets du Web des données. Voir <https://www.w3.org/RDF/>.

- **Gestion améliorée des objets numériques:** meilleure prise en charge des visionneuses d'objets numériques, y compris les visionneuses IIF⁵; intégration avec des systèmes externes tels que DAMS et des systèmes de conservation numérique
- **Accessibilité Web:** prise en charge des normes internationales et des meilleures pratiques en matière d'accessibilité Web pour permettre aux personnes malvoyantes d'utiliser AtoM 3

3. Qu'est-ce qu'une étude de faisabilité?

Dans le cadre du développement logiciel, une validation de principe est un projet conçu pour démontrer la faisabilité du développement d'une application logicielle pour répondre aux besoins d'intervenants définis en testant qu'un ensemble de fonctions logicielles souhaitées fonctionnera comme prévu une fois implémenté avec l'architecture technique et les outils sélectionnés. Pour entreprendre une EF l'équipe de conception doit établir les besoins des parties prenantes via des scénarii d'utilisation et des scénarii utilisateurs. L'équipe de conception définit ensuite un ensemble de fonctions logicielles et une infrastructure technologique censées satisfaire les besoins des parties prenantes, mais qui doivent être testées par l'EF afin de prouver ces hypothèses. Une EF inclut une analyse et une conception considérables et juste assez de codage pour que la solution puisse être testée par les intervenants désignés. Une EF réussie permettra à l'équipe de conception de tester la faisabilité de l'infrastructure technologique proposée et de ses fonctionnalités. Cela permettra également aux parties prenantes d'évaluer si la conception logicielle proposée peut répondre à leurs besoins.





4. Construire une étude de faisabilité: description détaillée

Notre objectif est d'axer l'EF d'AtoM 3 sur les créateurs et les utilisateurs typiques de descriptions archivistiques. Nous proposons donc un ensemble de scénarii d'utilisation et de scénarii utilisateurs répondant aux besoins des personnages décrits ci-dessous.

4.1 Exemples d'utilisateurs cibles

Personnage	Motivation & frustration
Chercheur	<ul style="list-style-type: none"> ● Ma principale motivation est de pouvoir regrouper de manière intuitive différents documents issus de plusieurs sources et créer mes propres collections afin d'appuyer mes recherches. ● Ma principale frustration serait de découvrir une nouvelle interface plus jolie

⁵ IIF est un ensemble de spécifications conçues pour améliorer la capacité des organisations à afficher leur contenu numérique en ligne. Voir <https://iiif.io/>.

	<p>mais n'ayant pas toutes les fonctionnalités pouvant m'aider à regrouper et rechercher mes données.</p>
<p>Responsable de service d'archives</p> 	<ul style="list-style-type: none"> • Ma principale motivation est de voir que le travail de mon organisation dans AtoM 2 n'est pas perdu et que les données créées par mon organisation peuvent désormais être utilisées de nouvelles façons que nous n'aurions pas imaginées possibles. • Ma principale frustration serait d'avoir à allouer des ressources pour apprendre à utiliser une nouvelle interface ou de recréer des descriptions déjà existantes.
<p>Archiviste</p> 	<ul style="list-style-type: none"> • Ma principale motivation est de me voir offrir un outil me permettant de travailler avec de grands volumes de données, et de pouvoir effectuer des opérations plus rapides et plus efficaces pour la création de métadonnées. • Ma principale frustration serait d'avoir l'impression que l'outil me contraint plus qu'il ne me fait gagner en autonomie.
<p>Conservateur</p> 	<ul style="list-style-type: none"> • Ma principale motivation est d'avoir un moyen puissant pour trouver des données, en utilisant des critères précis et spécifiques, et d'être capable de <i>le montrer</i> grâce à une grande exposition virtuelle. • Ma principale frustration serait d'avoir un outil puissant centré sur les fonctionnalités qui n'attire pas les visiteurs sur notre site.

4.2 Cas d'utilisation 1: convertir les descriptions archivistiques existantes en Web des données

Assertions:

1. Les utilisateurs d'AtoM 2 pourront passer sur AtoM 3 sans avoir à entreprendre de migrations de données chronophages.
2. Il sera possible de convertir les descriptions hiérarchiques traditionnelles en RDF.
3. Il sera possible d'afficher les descriptions RDF sous forme de représentations visuelles hiérarchiques et non-hiérarchiques.

Utilisateur	Scénario
-------------	----------



En tant que responsable d'un service d'archives, lorsque je mettrai AtoM 3 en place pour gérer et diffuser nos données archivistiques, j'aimerais pouvoir migrer aisément mes descriptions archivistiques d'AtoM 2 sans avoir à effectuer de mise en correspondance ou de nettoyage des données, ou toute autre tâche coûteuse et chronophage, afin de pouvoir protéger l'investissement de mon organisation dans l'actuel AtoM 2 et de minimiser les coûts du passage à la version d'AtoM 3.

Comment répondre aux besoins exprimés dans le scénario

1. Démontrer qu'il est possible d'importer les descriptions d'AtoM 2 dans AtoM 3 en utilisant scriptage et automatisation.
2. Démontrer que les descriptions importées d'AtoM 2 peuvent être sérialisées en RDF.
3. Démontrer comment les données peuvent être représentées hiérarchiquement et non-hiérarchiquement sur Internet.

4.3 Cas d'utilisation 2: permettre d'améliorer la description archivistique

Assertions:

1. Les utilisateurs d'AtoM 3 pourront créer des descriptions archivistiques plus rapidement et de manière plus efficace en utilisant des liens automatiques vers des ressources externes
2. Les utilisateurs d'AtoM 3 pourront créer des descriptions archivistiques avec un contexte plus riche en établissant un lien avec des ressources externes
3. Les utilisateurs d'AtoM 3 pourront créer des descriptions archivistiques avec un contexte plus riche en incorporant des termes issus d'ontologies du Web des données

Utilisateur	Scénario
	En tant qu'archiviste, lorsque j'ai de gros volumes de documents à décrire, j'aimerais pouvoir créer des liens qualifiés vers des ressources externes existantes afin de pouvoir effectuer des tâches de création de métadonnées plus rapidement, avec moins d'erreurs et un contexte plus riche, pour produire des métadonnées de qualité supérieure et réduire les contraintes de temps du personnel.
	En tant qu'archiviste, lorsque je décris des documents d'archives ayant une provenance complexe ou nuancée, je souhaiterais créer un nombre conséquent de relations qualifiées avec des acteurs et des concepts pertinents, ou d'autres descriptions, afin de pouvoir refléter plus précisément la complexité de la provenance.



En tant que responsable des archives, je souhaiterais m'assurer que les descriptions créées par mon personnel soient conformes aux normes et utilisent autant que possible les vocabulaires et autorités contrôlés existants dans un souci d'exactitude des données tout en offrant une plus grande richesse de découverte virtuelle.

Comment répondre aux besoins exprimés dans le scénario




1. Démontrer liaisons et désambiguïsation grâce à l'utilisation d'autorités externes. Par exemple:
 - afficher la désambiguïsation (Kingston, Ontario vs Kingston, Jamaica vs Kingston, London) de manière à laisser l'utilisateur choisir la bonne dénomination
 - montrer que John Smith en Alberta et Smith, John au Québec sont la même personne lorsqu'ils ont tous deux été liés à des autorités extérieures
2. Démontrer des moyens d'améliorer la saisie des données, par exemple:
 - remplir automatiquement des champs à l'aide d'autorités et vocabulaires externes
 - glisser-déposer des entités descriptives pour établir des relations
3. Démontrer qu'il est possible d'enrichir les descriptions archivistiques existantes en incorporant des termes provenant d'autres ontologies et vocabulaires.

4.4 Cas d'utilisation 3: fédérer et exposer les données

Assertions:

1. Il sera possible d'utiliser un seul site AtoM 3 pour regrouper les données de plusieurs sites AtoM 3 dans un portail de données
2. Une fonctionnalité de recherche robuste permettra aux utilisateurs d'extraire les résultats d'un ou de plusieurs sites AtoM 3 et de les visualiser sur Internet
3. AtoM 3 permettra une meilleure gestion des objets numériques afin d'améliorer l'expérience de l'utilisateur final
4. AtoM 3 permettra une meilleure découverte des descriptions d'archives sur le Web

Profil de l'utilisateur	Scénario
	En tant que responsable des archives dans une grande institution, lorsque je dois combiner les données archivées de plusieurs unités fonctionnelles, je veux pouvoir créer un site portail unique afin que chercheurs, archivistes et administrateurs puissent rechercher facilement dans toutes les collections de mon institution et afficher la liste des résultats de recherche dans une liste cohérente.

	<p>En tant que conservateur numérique, lors de la création d'une exposition virtuelle pour l'anniversaire d'un événement national important, je souhaite créer une collection personnalisée de descriptions archivistiques et d'objets numériques afin de pouvoir exposer les documents adéquats conservés dans mes archives.</p>
	<p>En tant que chercheur, lorsque je recherche des documents liés à mon sujet de recherche, je souhaite pouvoir effectuer une recherche dans plusieurs sites AtoM de diverses archives afin de pouvoir rechercher les documents adéquats conservés par ces archives.</p>
	<p>En tant que responsable du service des archives, lorsque nous publions nos descriptions archivistiques, je souhaite que les membres du public les retrouvent facilement via Google ou sur le site Web de mon institution mère afin de pouvoir remplir mon mandat: mettre à la disposition du public les fonds d'archives conservés dans mes dépôts.</p>

Comment répondre aux besoins exprimés dans les scénarii

1. Démontrer l'agrégation de contenu de plusieurs sites AtoM 3 dans un seul front-end
2. Démontrer la possibilité d'interroger des données d'un ou de plusieurs sites AtoM à partir d'un seul emplacement
3. Démontrer comment AtoM 3 sera capable de présenter du contenu avec les visionneuses IIIF, qui peuvent être utilisées pour assembler des objets numériques de plusieurs pages, fournir des fonctions de panorama et de zoom, permettre l'extraction de texte grâce à l'OCR, prendre en charge l'annotation et le marquage et monter des expositions virtuelles
4. Démontrer comment un utilisateur peut facilement implémenter un RDF structuré dans des schémas normalisés afin de produire de meilleurs résultats de recherche dans les principaux moteurs de recherche

4.5 Tâches pour l'étude de faisabilité

L'EF consistera à implémenter AtoM 3 dans un environnement de développement, en utilisant l'architecture technique décrite dans la section 5 ci-dessous, afin de tester les cas d'utilisation. L'environnement se limitera à tester les cas d'utilisation de façon minimale afin de démontrer la faisabilité de l'architecture proposée. Par exemple, dans un environnement de production réel et à grande échelle, le back-end d'AtoM 3 disposera de nombreux modèles de saisie de données et de nombreuses fonctionnalités conçues pour améliorer l'expérience utilisateur. De même, l'interface d'AtoM 3 offrira une expérience riche en fonctionnalités et sera conviviale pour les utilisateurs publics lorsqu'ils navigueront, rechercheront et afficheront du contenu. L'environnement de l'EF, en revanche, aura très peu de modèles de saisie de données et de fonctionnalités de navigation pour les utilisateurs finaux, notamment des

maquettes html ou même des armatures au lieu des formulaires fonctionnels, afin de démontrer la fonctionnalité sans la mettre en œuvre.

Les tâches requises pour exécuter avec succès une EF sont les suivantes. Notez que chaque cas d'utilisation s'appuie sur la fonctionnalité développée pour le cas d'utilisation précédent.

Cas d'utilisation 1: convertir les descriptions archivistiques existantes en Web des données

- Installation des serveurs et machines virtuelles pour héberger l'infrastructure technologique
- Installer les composants logiciels
- Sélectionner une ontologie⁶ et mettre en correspondance les données d'AtoM 2 avec celle-ci
- Migrer les échantillons de données d'AtoM 2 vers la plateforme de l'EF à l'aide de deux ou trois jeux de données d'AtoM 2
- Créer une interface Web ou modifier une interface Web existante pour interroger le Web des données et produire des visualisations hiérarchiques et non hiérarchiques

Cas d'utilisation 2: permettre d'améliorer la description archivistique

- Créer une identité visuelle pour montrer à quoi pourrait ressembler une interface de gestion de contenu avec AtoM 3
- Préparer des armatures pour le back-end d'AtoM 3 afin de décrire les modèles de saisie de données exploitant le Web des données
- Créer un prototype cliquable pour permettre aux testeurs de comprendre l'expérience de saisie de données dans AtoM 3 en se basant sur les armatures

Cas d'utilisation 3: fédérer et exposer les données

- Créer une identité visuelle pour montrer à quoi l'interface utilisateur publique d'AtoM 3 pourrait ressembler, et notamment les visualisations de description d'archivage hiérarchiques et non hiérarchiques
- Préparer des armatures pour l'interface d'AtoM 3
- Créer un prototype cliquable pour l'interface d'AtoM 3
- Configurer deux ou trois sites back-end d'AtoM 3
- Configurer le prototype front-end d'AtoM 3 pour effectuer une recherche fédérée sur les trois sites principaux d'AtoM 3 et afficher les résultats
- Démontrer que les recherches effectuées dans les principaux moteurs de recherche, tels que Google, extraient les résultats du prototype front-end d'AtoM 3

⁶ Le Matterhorn RDF, développé par Docuteam et les Archives de l'Etat du Valais (Suisse), est un modèle de données utile pour les besoins de l'EF. Matterhorn RDF est basé sur des ontologies existantes et a pour but de fournir des métadonnées de conservation et de description pour les fonds d'archives numériques. <https://wiki.docuteam.ch/doku.php?id=docuteam:matterhornrdf>

4.6 Evaluation de l'étude de faisabilité

Conformément aux bonnes pratiques en matière de développement de logiciels libres, ce document ainsi que toutes les analyses, la documentation, le code logiciel et les résultats des tests liés à l'EF devront être mis gratuitement à disposition en ligne et être largement annoncés afin de sensibiliser et d'informer la communauté archivistique sur toutes les étapes du processus.

L'analyse itérative, la conception, le prototypage et les tests sont essentiels au développement d'une application logicielle qui atteint son objectif et répond aux besoins de ses utilisateurs désignés. Les membres de la communauté, en particulier les utilisateurs actuels d'AtoM 2, devront être invités à commenter cette proposition et tous les autres aspects de l'EF et devront être recrutés comme testeurs. L'équipe de conception mettra les sites hébergés d'AtoM 3 à la disposition des testeurs, leur donnera des conseils pour tester le logiciel, sollicitera et documentera les commentaires reçus des testeurs et apportera des ajustements à la conception et à la mise en œuvre de l'EF dans les cycles de développement itératifs afin de prendre en compte les commentaires. Cette approche est très similaire à celle du développement logiciel de l'application originelle ICA-AtoM, qui a grandement bénéficié d'un pool de testeurs dédié, qui ont fourni un retour constant sur les nouveaux développements dès les tout premiers jours du projet.

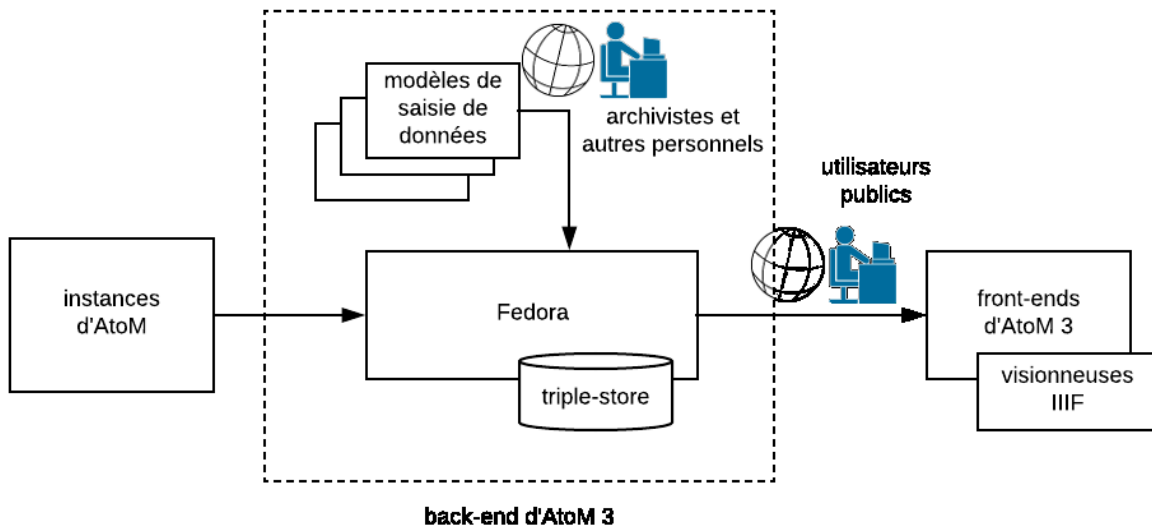
Les commentaires des testeurs doivent également être mis à la disposition des financeurs de l'EF, indépendamment de leurs interactions avec l'équipe de conception, afin de leur permettre d'évaluer dans quelle mesure l'EF répond aux attentes des utilisateurs en termes de fonctionnalité et de convivialité. Si les réactions des testeurs sont négatives ou incertaines, différentes approches de la conception du logiciel pourront s'imposer. C'est le but de l'EF - établir et tester certaines hypothèses sur le comportement d'une description archivistique et d'une plateforme de découverte en ligne, et vérifier qu'une conception et une infrastructure technologique données sont capables de prendre en charge les fonctionnalités souhaitées. Un développement complet sans une EF réfléchie et sans une réponse positive de l'utilisateur à la conception et aux fonctionnalités présenterait un risque élevé d'échec.

5. Architecture technique proposée (description de haut-niveau)

Comme il existe de nombreux outils open source capables de prendre en charge certaines des fonctionnalités proposées d'AtoM 3, l'EF et finalement AtoM 3 lui-même, ne doivent pas être entièrement refaits à partir de zéro. Au cours des mois qui ont précédé les réunions au King's College, les auteurs de ce document ont passé beaucoup de temps à examiner les projets open source existants afin de déterminer ceux qui pourraient être utilisés pour AtoM 3, faisant de Fedora⁷ un bon candidat pour la construction de la plateforme. Le diagramme ci-dessous illustre une architecture système de haut niveau utilisant Fedora comme plateforme de référentiel back-end pour gérer les objets numériques et les lier à leurs métadonnées descriptives, qui sont conservées dans un triple-store. Les métadonnées descriptives (ainsi

⁷ Fedora est un système ouvert de stockage de données numériques avec prise en charge native du Web des données. La communauté Fedora est soutenue par l'organisation à but non lucratif DuraSpace. <https://duraspace.org/fedora/>.

que les objets numériques associés) sont soit importés dans Fedora à partir des bases de données d'AtoM 2 et convertis au format RDF, soit entrés directement par les archivistes et autres personnels via des formulaires Web de saisie de données. Pour les utilisateurs publics, les données sont consultées et interrogées via des interfaces Web frontales à l'aide de la fonctionnalité de requête SPARQL⁸, qui permet d'afficher le contenu de manière hiérarchique ou non hiérarchique, l'affichage des objets numériques étant amélioré par les visionneuses basées sur la spécification IIIF. Un diagramme d'architecture plus détaillé et une description de l'infrastructure technologique sont présentés dans l'annexe B.



⁸ SPARQL est un langage de requête conçu pour récupérer et manipuler des données RDF.

Annexe A: Argumentaire pour le web des données

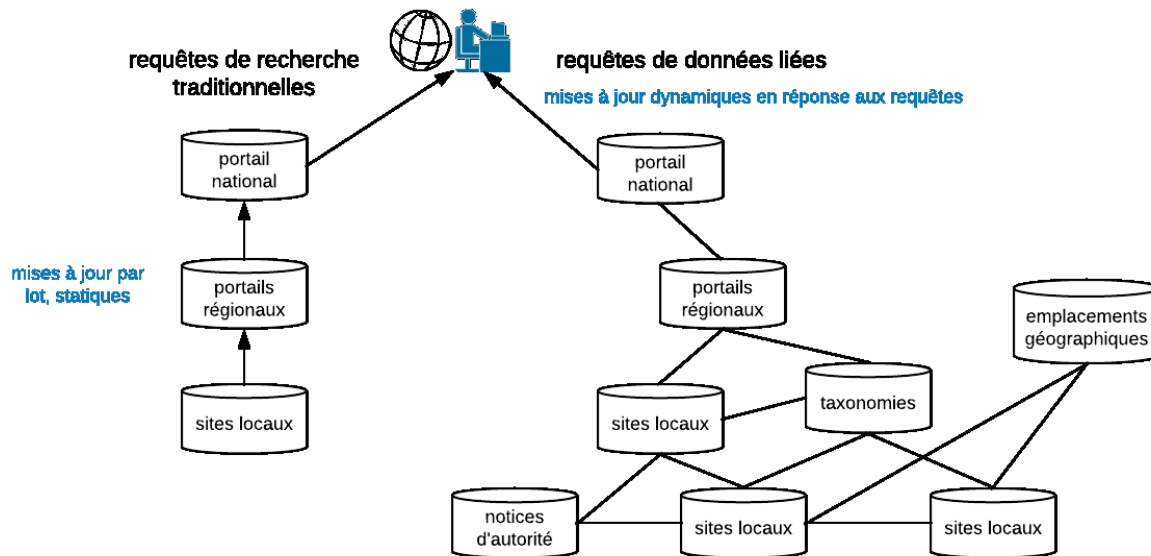
Un meilleur partage des informations

Le "Web des données" permet de structurer les données de manière à ce que des sources d'informations disparates puissent être connectées de manière compréhensible par les ordinateurs. C'est bien différent de l'incorporation de liens hypertextes dans des pages Web qui oblige l'utilisateur à ouvrir manuellement les liens afin de comprendre la relation entre les deux ressources. Par exemple, dans le monde du Web des données, un ensemble de documents et une notice d'autorité peuvent être liés de manière significative, de sorte qu'un moteur de recherche puisse comprendre que la personne ou l'organisation citée dans la notice d'autorité entretient une relation spécifique avec les documents, en tant que créateur, accumulateur, auteur, sujet, gardien, etc. Le Web des données facilite également le partage d'informations, de sorte que, par exemple, plusieurs ensembles de documents détenus par plusieurs organisations puissent être liés à une notice d'autorité centrale. Le Web des données permet également de faire des références croisées entre archives, bibliographies, citations, objets de musée, images et données cartographiques (parmi de nombreux types de ressources), rassemblant ainsi les communautés de gestion des informations patrimoniales. Des ressources utiles pour comprendre le Web des données dans le secteur du patrimoine culturel sont disponibles à l'adresse <http://publish.illinois.edu/linkedspcollections/outcomes/> et <http://lodlam.net/>. Voir aussi <http://www.canadiana.ca/pcdhn-lod> et <https://linkedjazz.org/>, deux exemples de projets utilisant le Web des données.

Agrégation et mises à jour plus faciles dans les portails multi-référentiels

L'agrégation de données dans les sites portail est mieux prise en charge par l'utilisation du Web des données que par une description hiérarchique traditionnelle. Bien qu'AtoM 2 soit utilisé par de nombreuses organisations pour décrire les fonds d'archives et les rendre disponibles en ligne, le logiciel n'a jamais pleinement répondu au besoin des archives d'exposer leur contenu sur le Web. L'agrégation de données provenant de plusieurs instances d'AtoM vers des sites portails tels que Archeon et Archives Canada, par exemple, est un processus relativement fastidieux d'importation de données mises à jour via CSV et EAD, et de remplacement du contenu existant. Les données résidant dans plusieurs emplacements sont rapidement désynchronisés avec les sites de type portail, et le problème est encore aggravé lorsque les données sont agrégées sur plusieurs niveaux (par exemple, des sites locaux contribuant à un portail régional; des portails régionaux contribuant à un portail national; des portails nationaux contribuant aux portails internationaux). Ce problème est dû en partie aux modèles de données impliqués: les descriptions mises à jour doivent être exportées et importées manuellement ou via des moissonnages OAI planifiés, ce qui revient essentiellement à des migrations récurrentes de données d'instances locales vers des agrégateurs de données.

Le Web des données, en revanche, est conçu pour l'interrogation dynamique de plusieurs sources d'informations et pour l'assemblage des résultats dans des interfaces Web. Les différences entre l'utilisation du Web des données et de la description archivistique traditionnelle pour l'agrégation de données dans des sites portail sont illustrées dans le diagramme ci-dessous.



Plus de contexte, plus de sens, plus de nuances

Le Web des données est également plus efficace pour représenter les relations complexes et les informations contextuelles dans la description archivistique. La description archivistique traditionnelle met l'accent sur la propriété et la détention des documents et sur la primauté d'un seul créateur pour un agrégat de documents. Cependant, le monde réel s'accorde rarement si clairement avec un tel modèle, en particulier à l'ère numérique. Les théoriciens archivistiques contemporains défendent de plus en plus le fait que les documents résultent d'interactions dynamiques entre divers individus et communautés et qu'attribuer la provenance d'un fonds, d'une série, d'un fichier ou d'une autre unité de description à une source unique est souvent trop simpliste. Les normes descriptives telles qu'ISAD (G), ISAAR et RAD (Règles canadiennes pour la description archivistique) prévoient le classement hiérarchique des agrégats de documents, les créateurs étant attachés au niveau de description le plus élevé possible et hérités des niveaux inférieurs. Les schémas de Web des données, d'autre part, permettent de capturer les relations riches entre individus et communautés qui interagissent avec les documents en tant que créateurs, gardiens et sujets. Plutôt qu'un récit dominant unique, une approche avec le Web des données permet à une multiplicité de récits d'être enregistrés sur une ressource, y compris une approche multi-provenance pour la description et le classement.

Dans AtoM 2, l'organisation hiérarchique et les relations limitées entre agents et documents sont appliquées par le modèle de base de données relationnelle sous-jacent sur lequel le logiciel est construit. Dans une plateforme utilisant le Web des données, des relations beaucoup plus expressives sont rendues possibles par l'utilisation d'ontologies conçues spécifiquement à cette fin. Records in Context (RiC), par exemple, offre de nombreuses possibilités pour exprimer les interactions complexes entre individus, organisations, fonctions et ensembles de documents.

Un meilleur soutien pour la normalisation

Une plus grande normalisation des pratiques descriptives est également plus facile avec le Web des données, conçu pour permettre des liaisons entre notices d'autorités, vocabulaires contrôlés et autres sources d'informations externes. Au sein d'un jeu de données, si une valeur est un lien au lieu d'un texte dans une plateforme utilisant le Web des données, la relation entre la description et la ressource externe liée peut être rendue explicite et la ressource liée peut être affichée avec les données auxquelles elle a été liée. Par exemple, une notice d'autorité pour une école peut inclure un lien vers une notice d'autorité pour un individu et le lien peut spécifier si l'individu a fréquenté l'école ou était un enseignant ou un administrateur, etc. D'autres liens entre l'école et les individus peuvent permettre aux moteurs de recherche de déduire les relations entre les individus, et les outils de visualisation utilisant le Web des données peuvent ensuite présenter ces relations d'une manière qui n'est pas possible si l'archiviste ou le chercheur ne font que suivre les liens incorporés dans des descriptions en texte libre. De nombreuses organisations peuvent utiliser les mêmes notices d'autorité plutôt que de rédiger laborieusement leurs propres récits administratifs et biographiques et les utiliser uniquement pour leurs propres descriptions archivistiques.

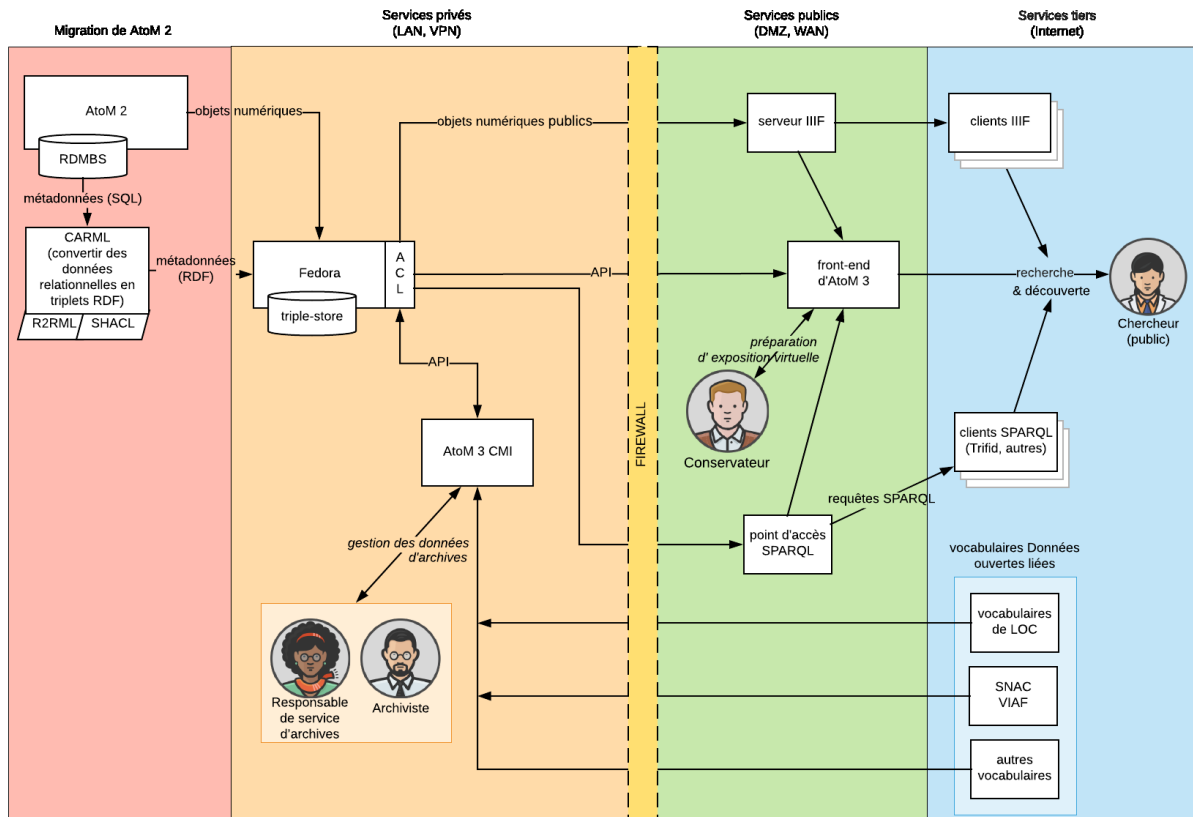
Répondre aux besoins avec AtoM 3

La disponibilité croissante des services liés au Web des données venant d'institutions du patrimoine culturel permet d'établir de tels liens mais cela nécessite que des applications utilisant ces données existent. Les notices d'autorité, les vocabulaires contrôlés et même les ontologies et les modèles de données archivistiques sont de plus en plus répandus et plus facilement disponibles, mais peu, voire aucune, application logicielle spécifiquement orientée vers les archives ne peut ingérer, réutiliser ces données ou contribuer à l'écosystème du Web des données. Les applications basées sur le Web des données comme Islandora et Samvera, et les agrégateurs de données tels que Digital Public Library of America et Europeana sont conçus pour les bibliothèques, qui concentrent leurs efforts sur les publications et les collections plutôt que sur l'accumulation naturelle d'ensembles de documents dans des contextes complexes de création et d'utilisation.

Avec le passage du Conseil International des Archives à une norme archivistique unique et unifiée basée sur le Web des données et suite aux autres changements survenus dans le domaine de la description archivistique, il est devenu urgent de disposer d'une application de description archivistique libre et gratuite qui puisse créer et exposer le Web des données et répondre aux exigences des petites comme des grandes institutions. AtoM 3 sera conçu pour répondre à ce besoin.

Annexe B: Architecture technique proposée (détails)

Le diagramme d'architecture et la description associée ci-dessous décrivent les composants de l'EF et la manière dont ils seraient utilisés pour accomplir les tâches décrites dans les cas d'utilisation et les scénarii.



Dans ce diagramme:

- Un composant de conversion vers le modèle RDF (R2RML) tel que CARML, permet de faire correspondre des données MySQL d'AtoM 2 en triplets RDF, et charger les triplets RDF en arrière-plan sur un serveur AtoM 3 fonctionnant sous Fedora. Le Shape Constraint Language (SHACL) est utilisé pour valider la structure des triplets RDF.
- Un script importe les objets numériques d'AtoM 2 à partir du système de fichiers d'AtoM 2 dans l'entrepôt de données Fedora et associe les objets numériques importés aux métadonnées RDF appropriées.
- Fedora assure le stockage et la gestion des métadonnées, le stockage des objets numériques, le contrôle d'accès (via WebAC), les API HTTP REST, la gestion des versions des ressources (via Momento), la vérification de la fixité et d'autres fonctionnalités du dépôt.
- L'interface de gestion de contenu (CMI) d'AtoM 3 fournit des modèles de saisie de données utilisés pour saisir directement des métadonnées et les écrit sur Fedora via l'API HTTP REST.

Bien que les données soient capturées dans Fedora sous forme de triplets RDF, les modèles de saisie de données ressemblent à ceux utilisés dans AtoM 2.

- Lors de la création et de la modification de descriptions, le CMI d'AtoM 3 permet de créer des relations certifiées à l'aide de vocabulaires établis issus du Web des données (LCSH, SNAC, VIAF, par exemple) afin d'enrichir les métadonnées archivistiques et de fournir davantage de contexte.
- Les listes de contrôle d'accès (ACL)⁹ permettent de gérer les autorisations des groupes et des utilisateurs pour les données et les objets numériques affichés par les API REST de Fedora.
- L'utilisateur peut télécharger des objets numériques sur AtoM 3 via le CMI, comme dans AtoM 2.
- Un serveur IIIF¹⁰ s'intègre à l'API REST de Fedora pour fournir des métadonnées et du contenu d'objets numériques aux clients IIIF. Les objets numériques disponibles via le serveur IIIF sont limités aux documents publics de l'ACL de Fedora.
- Un point d'accès SPARQL permet aux clients tiers SPARQL tels que Trifid¹¹ d'exécuter des recherches puissantes et ad hoc sur les données RDF d'AtoM 3 afin de faciliter les nouvelles requêtes et les agrégations de données. Le contenu disponible via le point d'accès SPARQL est limité aux documents publics de l'ACL de Fedora.
- L'interface d'AtoM offre une possibilité de recherche et de découverte publiques des données archivistiques. Une interface de recherche conviviale et familière interrogera le point d'accès SPARQL pour renvoyer et afficher les résultats de recherche standard.
- Des données archivistiques détaillées seront fournies au client d'AtoM 3 via l'API REST de Fedora.
- En utilisant les requêtes SPARQL une interface front-end d'AtoM permet d'interroger et de regrouper les résultats de plusieurs instances d'AtoM 3.
- Les collections virtuelles peuvent être organisées dans le front-end d'AtoM en combinant les ressources d'une ou de plusieurs instances d'AtoM.

⁹ ACL est largement utilisé pour la gestion des groupes d'utilisateurs et des autorisations, et est utilisé dans AtoM 2. https://fr.wikipedia.org/wiki/Access_Control_List.

¹⁰ IIIF, ou International Image Interoperability Framework, est une norme internationale bien acceptée, conçue pour prendre en charge le rendu des objets numériques dans les visionneuses. <https://iiif.io/>.

¹¹ Trifid est un outil open-source développé par Zazuko <https://github.com/zazuko/trifid>.