



HAL
open science

Mettre en pratique les recommandations sur les archives ouvertes de nouvelle génération de COAR pour l'archive ouverte institutionnelle Horizon Pleins textes de l'IRD

Pascal Aventurier

► To cite this version:

Pascal Aventurier. Mettre en pratique les recommandations sur les archives ouvertes de nouvelle génération de COAR pour l'archive ouverte institutionnelle Horizon Pleins textes de l'IRD. 3. Colloque International sur le Libre Accès, ESI. Ecole Supérieure de de l'Information, Rabat (Maroc)., Nov 2018, Rabat, Maroc. hal-01941353

HAL Id: hal-01941353

<https://hal.science/hal-01941353v1>

Submitted on 6 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mettre en pratique les recommandations sur les archives ouvertes de nouvelle génération de COAR pour l'archive ouverte institutionnelle, Horizon Pleins textes de l'IRD

*Putting into practice the recommendations of
COAR Next Generation Repositories report
into the context of the institutional repository,
Horizon Plein textes from IRD*

P. Aventurier, IRD- France

Résumé

Longtemps les archives ouvertes ont été cantonnées à des fonctions, déjà très importantes, de stockage et de diffusion de l'information. Elles ont connu des évolutions fonctionnelles progressives depuis les années 2000, cependant un récent rapport de COAR (Confederation of Open Access Repositories) sur les archives ouvertes de nouvelle génération explique que celles-ci n'ont pas développé pleinement leur potentiel et ne sont pas assez connectées. Ce travail analyse comment les principales recommandations du rapport COAR pourraient être appliquées à l'archive ouverte institutionnelle Horizon Plein textes de l'IRD.

Mots-clés. Archive ouverte ; Accès ouvert ; Standards ; Système d'information ; COAR ; Fonctionnalités.

Abstract

For a long time, open archives have been confined to important functions of storage and dissemination of information. They have experienced progressive functional evolutions since the 2000s, however a recent report from Confederation of Open Access Repositories (COAR) on open archives of new generation explains that they have not developed their full potential and are not enough connected. This work analyzes how the main recommendations of the COAR report could be applied to the IRD's Institutional repository "Horizon Plein textes".

Keywords. Repository; open access; standards; information system; COAR; Functionalities.

1 Introduction

L'Institut de Recherche pour le Développement (IRD) est un organisme de recherche public français qui propose un modèle original : le partenariat scientifique équitable avec les pays en développement, principalement ceux des régions intertropicales et de l'espace méditerranéen, pour lier réduction des inégalités de développement, changements globaux, atteinte des objectifs du développement durable et innovation fondée sur la recherche. L'IRD a inscrit dans ses textes le recensement et la diffusion de ses connaissances scientifiques et plus récemment, dans son contrat d'objectifs et de performance 2016-2020, la diffusion des documents en libre accès. Ainsi est née il y a plus de 30 ans la base documentaire Horizon qui est l'une des plus ancienne archive ouverte institutionnelle française (Couperin 2017).

Longtemps les archives ouvertes ont été cantonnées à des fonctions, déjà très importantes, de stockage et de diffusion de l'information. Elles ont connu des évolutions fonctionnelles progressives depuis les années 2000 comme l'ajout de fonctions bibliométriques, la gestion des listes de publications des chercheurs, la production de statistiques de consultation, la mise à disposition de flux RSS, ou encore une exposition des données avec le protocole OAI-PMH. Cependant un récent rapport du groupe de travail « Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group » de COAR (Confederation of Open Access Repositories) sur les archives ouvertes de nouvelle génération explique que celles-ci n'ont pas développé pleinement leur potentiel (Rodrigues et al. 2017). Elles devraient évoluer pour être mieux intégrées dans l'écosystème de la production scientifique mais aussi dans les systèmes d'information des institutions. Le rapport du MIT sur le futur des bibliothèques (MIT 2016) va aussi dans ce sens et recommande que les bibliothèques génèrent des plateformes de contenu ouvertes et interopérables qui explorent de nouvelles méthodologies innovantes pour la découverte, l'organisation, le partage ou la production d'informations.

L'objectif de ce travail est tout d'abord de présenter et d'analyser comment les recommandations du rapport COAR pourraient être appliquées à Horizon, puis de dégager les fonctionnalités qui semblent les plus importantes, enfin de préciser le niveau d'avancement et de standardisation des développements libres réutilisables.

2 L'archive ouverte institutionnelle Horizon

Horizon¹⁰ recense maintenant plus de 98000 documents dont 65000 PDF téléchargeables en libre accès en son volet Pleins textes. L'accroissement d'Horizon est d'environ 3000 documents par an (Cavet 2014). Les statistiques de consultation montrent qu'environ 7500 fichiers sont téléchargés chaque jour, dont 70% par des publics des pays du Sud.

Horizon utilise pour sa partie gestion documentaire le logiciel Cadic Intégrale de la société CADIC et le moteur SearchServer pour l'indexation des références et des PDF depuis 2006. Des efforts importants sont faits

¹⁰<http://www.documentation.ird.fr/>

pour numériser des documents qui constituent un patrimoine important pour l'Institut (ROSSI 2017).

Horizon est dupliquée quotidiennement en SQL sur un environnement qui permet de fournir des services à valeur ajoutée aux chercheurs tels que des identifiants pérennes pour les références bibliographiques, des *landing pages*¹¹ des notices, des listes dynamiques de publications (individuelles, UMR) exportables en différents formats, intégrables (fonction 'embedded') dans des pages web personnelles ou de laboratoires, et des flux RSS. C'est dans cet environnement que sont disponibles au format Dublin Core les 98000 *records* de l'entrepôt OAI. En effet, Horizon est moissonnée en OAI-PMH par, entre autres, le CLACSO (Consejo Latino Americano de Ciencias Sociales)¹², Base (Bielefeld Academic Search Engine)¹³, OpenAire¹⁴ pour l'Union Européenne ou Isidore¹⁵ pour les SHS, le portail de l'information publique environnementale Tout sur l'environnement¹⁶.

Horizon dispose d'un troisième environnement : Horizon "Bibliométrie". Il est alimenté par les publications de l'IRD repérées dans le Web of Science (WOS) de manière semi-automatique à l'aide d'une base de reconnaissance des segments d'adresse des auteurs. Cette base "Bibliométrie" gère l'attribution des publications à l'unité de recherche, pour les chercheurs de l'IRD, mais également pour ceux des UMR (Unité Mixte de Recherche¹⁷) ; les métadonnées sont en outre enrichies par la mention des pays du Sud signalés dans l'affiliation d'au moins un des co-auteurs. Les références bibliographiques des chercheurs de l'IRD collectées pour la base "Bibliométrie" sont également importées dans Horizon et reçoivent un identifiant interne. Les résultats de la veille sur le WOS sont diffusés via un bulletin mensuel interactif et des listes de publications dynamiques pour chaque UMR et chaque chercheur de l'IRD. La base "Bibliométrie" est également utilisée pour alimenter des systèmes d'information institutionnels sur les activités de recherche.

Enfin, Horizon interagit avec l'archive nationale française HAL¹⁸. Une fonctionnalité permet au chercheur de déposer aisément un PDF dans le sous-ensemble HAL-IRD via l'identifiant interne IRD qui déclenche l'autocomplétion des métadonnées bibliographiques. Le chercheur se contente alors d'ajouter sur HAL le texte intégral dans sa version autorisée. Dans l'autre sens, tout dépôt dans HAL d'une (co) publication d'un agent IRD est signalé automatiquement à l'équipe Information Scientifique et Technique et est ajouté dans Horizon, après vérification et demande éventuelle de complétion.

¹¹ Les *landing pages* sont les pages sur lesquelles pointent les moteurs de recherche, par exemple.

¹²<https://www.clacso.org.ar/>

¹³ <https://www.base-search.net/?l=fr>

¹⁴ <https://www.openaire.eu/>

¹⁵<https://www.rechercheisidore.fr/>

¹⁶ <https://www.toutsurlenvironnement.fr/>

¹⁷ Une Unité mixte de recherche (UMR) est une structure administrative constituée par la mise en commun par différents établissements d'enseignement supérieur et de recherche de moyens humains et matériels fondée sur une stratégie et un projet scientifique communs.

¹⁸<https://hal.archives-ouvertes.fr/>



Figure 1 : L'interface de la page d'accueil Horizon Plein textes (version en pré-prod au 02/09/2018)

La figure 1 présente l'interface et quelques fonctionnalités, dont les statistiques de consultation.

Les services actuellement offerts par Horizon

Le tableau 1 ci-dessous présente les services actuellement offerts par la base Horizon.

Tableau 1 : Services offerts par la base Horizon

Type de service	Description
Services aux chercheurs - Déposer	<p>Deux modes d'alimentation :</p> <p>Manuel : dépôt par mail du document, puis enregistrement par les documentalistes ou veille du service IST sur les revues, les sites d'éditeurs, les sites d'UMR puis enregistrement par les documentalistes</p> <p>Environ 1500 notices par an</p> <p>Semi-automatique : veille hebdomadaire sur le Web of Science via une équation de recherche sur</p>

	<p>les affiliations IRD, export dans EndNote : import mensuel des notices dans Horizon Environ 1400 notices par an Dans tous les cas, ajout et vérification des droits pour le texte intégral</p>
<p>Services aux chercheurs : Bibliométrie</p>	<p>Bulletin de veille mensuel, plus de 550 abonnés Suivi des publications : rapport bibliométrique Production de listes dynamiques individuelles ou d'UMR</p>
<p>Valoriser : dans les systèmes externes et les moteurs de recherches</p>	<p>Moissonnage OAI-PMH et compatibilité avec les directives européennes (OpenAire) et les infrastructures nationales (Isidore, ...) Indexation Google optimisée (déclaration de sitemap¹⁹ pour le référencement des PDF)</p>
<p>Valoriser: les fonctionnalités de la plateforme</p>	<p>Recherche dans les métadonnées Recherche plein texte dans le PDF Flux RSS Export Dernières saisies</p>
<p>Valoriser : les statistiques de consultation</p>	<p>Mesurer l'impact de la production de l'institution ou de sa production personnelle pour le chercheur via l'analyse des consultations de nos documents sur internet : indicateurs par document, ou sur des ensembles de documents (thématique, auteur, langue...) présentation dynamique des documents via des outils de mapping données librement accessibles par des URL libres</p>

Le fonds documentaire IRD (FDI), collection patrimoniale des publications des scientifiques de l'Institut, a été distingué fin 2017 par l'attribution du label CollEx, collection d'excellence pour la recherche, pour une période de cinq ans (2018-2022) reconductible. Ce fonds, conservé sur le site de la Délégation régionale Île-de-France à Bondy, est intégralement répertorié dans la base Horizon. Ce label permet de répondre à des appels à projets pour la constitution de fonds ou le développement de services pour les chercheurs.

Horizon tient également lieu de catalogue collectif de bibliothèque pour les dix bibliothèques du réseau international de l'IRD.

¹⁹ Un sitemap est un fichier XML ou texte standardisé qui indique au moteur de recherche les adresses URL des pages et leur relation pour pouvoir mieux les indexer.

2.1 Evolutions envisageables pour la base Horizon

Plusieurs fonctionnalités sont à l'étude à différents niveaux. Elles n'ont pas encore été planifiées mais apporteraient de la valeur ajoutée à Horizon. Il s'agit de :

- L'implémentation du module HAL, qui est une fonctionnalité du logiciel Cadic Intégrale, dans Horizon pour permettre d'envoyer directement des métadonnées dans HAL. Cela permettrait aux Unités Mixtes de Recherches pour lesquelles HAL (l'archive ouverte française) est le système d'évaluation de référence, d'avoir l'ensemble de leurs métadonnées dans une seule base ;
- L'intégration de l'identifiant numérique pour chercheurs, Orcid²⁰, afin de permettre une alimentation automatique des comptes chercheurs Orcid à travers cet identifiant unique ;
- L'ajout d'information pour lier la publication au jeu de données et réciproquement ;
- L'implémentation de l'API ISTE²¹ qui permet de lier les 20 millions d'articles scientifiques achetés par ce projet et mis à disposition des chercheurs des organisations françaises.
- La mise à disposition des données bibliographiques en données liées ouvertes (linked open data) et alignées avec des vocabulaires de référence permettant ainsi de s'affranchir encore davantage du silo de données que représente une base de données.

3 Les recommandations du rapport COAR

En 2016, COAR (Coalition of Open Access Repositories) a lancé un groupe de travail pour identifier les fonctionnalités centrales pour la nouvelle génération d'archives ouvertes ainsi que les architectures et technologies qui permettraient de les implémenter. L'orientation globale est de mieux connecter les archives ouvertes en réseau pour constituer un patrimoine mondial de connaissances en encourageant la création de services à valeur ajoutée entre les entrepôts de données de documents que sont les archives ouvertes²².

Le rapport COAR identifie de nouvelles fonctionnalités principales et présente des recommandations pour l'adoption de nouvelles technologies, standards et protocoles qui peuvent aider les entrepôts à mieux s'intégrer dans l'environnement du web et à jouer un rôle plus large dans l'écosystème de la communication scientifique. Le rapport présente onze nouveaux comportements, technologies, standards ou protocoles qui faciliteraient le développement de services novateurs comme le travail en réseau, la validation par les pairs, la notification ou l'usage d'évaluation.

²⁰Orcid : <https://orcid.org/>

²¹Istex : <https://www.istex.fr/>

²² En anglais, le terme "repository" convient pour les entrepôts de données et de documents. En français, le terme "document repository" a conservé la traduction "archive ouverte" alors que les anglo-saxons ont abandonné le terme "open archive" et utilisent "document repository".

3.1 Présentation des recommandations du rapport

Les onze comportements “centraux” cités dans le rapport Next Generation of repositories de COAR sont traduits et présentés ci-dessous avec les technologies associées à chacun d’entre eux et une description.

1- Exposer les identifiants.

Les identifiants doivent être disponibles pour la landing page mais aussi pour le document en PDF par exemple.

Technologie associée : Signposting

2- Déclarer les licences au niveau de la ressource.

La licence et les conditions d’utilisation doivent être appliquées au niveau de chaque objet.

Technologies associées : Signposting, Creative Commons Licence

3- Découvrir à travers la navigation.

Lien direct vers les objets sans passer par la landing page car les outils ne peuvent pas suivre les liens et pour pouvoir accéder aux métadonnées en temps réel via une API.

Technologies associées : Signposting

4- Interaction avec des ressources (Annotation, Commentaire, Review)

Possibilité d’interagir avec les ressources : commenter, mention “j’aime”,...

Technologies associées : Activity Streams 2.0, Atom syndication format, Web annotation Model and Web Annotation Protocol, International Image interoperability Network

5- Transfert de ressources

Les ressources (et/ou leurs métadonnées) ne devraient pas être copiées, principalement pour diminuer les temps d’accès.

Technologies associées : IPFS, Resourceync, SWORD

6- Batch Discovery.

Recherche de plusieurs ressources à travers différents entrepôts pour repérer et récupérer des documents pour le text mining ou la préservation par exemple.

Technologies associées : ResourceSync, SignPosting, Sitemaps

7- Collecter et exposer l’activité

L’activité de collecte et d’exposition dynamique et en temps réel des activités (par exemple, des informations sur les changements, ajouts, commentaires, annotations, peer-review, accès, téléchargements, etc.) concernant les objets scientifiques qu’ils hébergent.

Technologies associées : Activity Streams 2.0, Linked Data Notifications, ResourceSync (Change notifications), Signposting, WebMention, WebSub, Other messaging protocols (AMQP, Kafka)

8- Identification des utilisateurs

L'identification des utilisateurs permet d'être reconnu par le système, et de recevoir des recommandations, pour découvrir de nouvelles productions scientifiques potentiellement intéressantes pour lui.

Technologies associées : ORCID ;Social NetWork Identities, WebId

9- Authentification des utilisateurs

Demander aux utilisateurs de s'identifier eux mêmes avec des identifiants quand ils interagissent avec des ressources.

Technologies associées : http Signatures, OpenId Connect 1.0, WebID/TLS

10- Exposer des métriques standardisées

Partage des données d'interaction ou des statistiques de consultation des utilisateurs, soit en mode Push (en utilisant Sushi), soit en mode Pull avec les protocoles de récupérations de statistiques par les agrégateurs (par exemple, OpenAire utilise Piwik).

Technologies associées : Counter, SUSHI, ETAG, Usage Metrics Service Provider for repository IRUS-UK ou Piwik RAMP. Voir la partie 5 transfert de ressources.

11- Préservation des ressources

Permettre la préservation sur le long terme. Le groupe de travail a statué que la préservation est une activité extrêmement complexe, impliquant l'adoption de politiques, normes, pratiques et technologies appropriées déjà mises en oeuvre par des communautés existantes.

Par contre la mise à disposition suivant le comportement 5 "Batch transfert" relève bien de la responsabilité des archives ouvertes et entrepôts.

Technologies associées : les mêmes que le point 5.

3.2 Analyse des recommandations

Parmi les technologies exposées, certaines sont déjà bien connues mais peut-être pas suffisamment implémentées dans les archives ouvertes comme par exemple la Creative Commons Licence, SWORD ou OpenId Connect pour l'identification.

Les technologies Activity Streams, mais surtout RessourceSync et Signpost, semblent vraiment importantes et sont, d'après COAR en train d'être implémentées dans des plateformes.

Activity Streams : <https://www.w3.org/TR/activitystreams-core/>

Activity Streams 2.0 est une recommandation du W3C depuis le 23 mai 2017. Cette spécification décrit un modèle pour représenter les activités potentielles et terminées en utilisant le format JSON. Elle est destinée à être utilisée avec des vocabulaires qui détaillent la structure des activités et définissent des types spécifiques d'activités d'un nouvel objet pertinent ou d'une modification de métadonnées existantes. Il permettra aussi de proposer du contenu à des utilisateurs.

Signposting : <http://signposting.org>

Signposting est une approche pour rendre le web académique plus convivial pour les machines en exposant les relations en tant que liens typés dans les en-têtes de liens HTTP.

Par exemple dans une “landing page”, la page html décrivant la ressource pour les consultations humaines, l’en-tête HTTP contient des informations typant la relation auteur et le reliant avec son identifiant Orcid (figure 2). Cette information est directement interprétable par les machines²³.

```
Link: <http://orcid.org/0000-0003-4018-5153>  
; rel="author",  
<http://orcid.org/0000-0002-2787-7166>  
; rel="author",  
<http://orcid.org/0000-0003-3749-8116>  
; rel="author"
```

Figure 2 : un exemple d'en tête de signposting

Signposting sera implémenté dans Dspace ⁷²⁴ et est déjà implémenté dans OJS (Open Journal System²⁵), un logiciel libre pour la gestion complète de journaux scientifiques, dans des entrepôts de données et dans le schéma Datacite pour décrire les jeux de données scientifiques²⁶.

Resourcesync : <http://www.openarchives.org/rs/toc>

Resourcesync est une spécification s'appuyant sur les sitemaps qui peuvent être utilisés par les administrateurs pour fournir des informations permettant aux systèmes tiers de rester synchronisés avec les ressources de l'entrepôt-source à mesure qu'elles évoluent (création, mise à jour, etc..). ResourceSync a été créé par la SLOAN Foundation²⁷ et le JISC²⁸.

L'utilisation de vocabulaires contrôlés en format normalisé SKOS²⁹ pour décrire les objets, permettant ainsi une description avec des référentiels communs et de mieux relier ou agréger les données est très importante. Cette recommandation n'apparaît que très peu dans le rapport.

Enfin, il est nécessaire d'avoir des agrégateurs ou concentrateurs de données (“hub” en anglais) interopérables pour permettre les échanges d'information à travers les archives ouvertes et les entrepôts de données. On peut se demander s'il s'agira de nouveaux systèmes ou si des plateformes comme Openaire ou HAL peuvent jouer ce rôle.

²³<https://signposting.org/author/springer/>

²⁴ <https://wiki.duraspace.org/display/DSPACE/DSPACE+Release+7.0+Status>

²⁵ <https://pkp.sfu.ca/ojs/>

²⁶<https://signposting.org/adopters/>

²⁷<https://sloan.org/>

²⁸<https://www.jisc.ac.uk/>

²⁹<https://www.w3.org/2004/02/skos/>

4 L'application de ces recommandations dans le contexte d'Horizon Pleins textes

Le principe d'échanges de contenus est très intéressant pour une archive ouverte "modeste" et offre de nouvelles possibilités bien supérieures au protocole OAI-PMH. Les technologies mentionnées sont plutôt légères à mettre en oeuvre et les comportements listés dans le rapport n'impliquent pas de modification fondamentale de l'architecture d'une archive ouverte.

La COAR préconise de focaliser les services sur les ressources plutôt que sur l'archive. Avec des outils qui ne relèvent certes pas encore des préconisations techniques de type Signposting ou autres, la base Horizon propose déjà des services associés aux ressources, qu'il s'agisse des sitemaps permettant de rendre visibles les PDF dans des moteurs de recherche, ou encore via les urls pérennes pour les métadonnées, ou les listes de publications dynamiques : c'est bien la ressource en elle-même, et non pas l'archive Horizon Pleins textes, qui est "poussée" ici sur le web.

Le rapport conforte des développements que nous avons envisagés comme l'attribution d'un identifiant Orcid pour les chercheurs, les échanges de métadonnées ou les liens entre archive ouverte (dépôt de documents) et entrepôt de données.

L'implémentation des fonctionnalités peut à moyen terme remettre en cause l'outil sur lequel est développé notre archive ouverte.

Cependant la base Horizon et son architecture actuelle permettraient de mettre en place facilement des recommandations du rapport. La base Horizon possède des permaliens vers les landing pages qui pourraient être utiles pour implémenter Signposting et elle présente déjà des sitemaps qui pourraient être reliés à la spécification Resourcesync. Le fait de connaître et de maîtriser la structure de la base peut aussi être un atout, bien que les recommandations ne portent pas sur la structure de la base en elle-même.

Parmi les faiblesses, on peut noter que Horizon a deux fonctions : un catalogue pour gérer des fonds papier, le Fonds Documentaire de l'IRD et ceux des bibliothèques du réseau IRD, et une fonction d'archive ouverte institutionnelle. Il faudrait pouvoir continuer à différencier ces deux fonctions dans les échanges de métadonnées et de documents.

Le rapport COAR insiste sur l'importance d'un système d'archives ouvertes distribué, et non pas centralisé. Il faudrait donc que les archives ouvertes connectées suivent l'ensemble des recommandations pour réaliser pleinement le potentiel des technologies projetées dans le rapport.

Enfin, deux besoins ne sont pas abordés dans les recommandations :

le besoin de listes bibliométriques n'est pas évoqué alors que les archives ouvertes participent à l'identification des productions scientifiques des organismes de recherche et dans certains organismes représentent l'une des fonctionnalités principales.

Le web sémantique et les données liées ouvertes (linked open data). Pourtant, le linked open data est une composante de plus en plus importante du web, en particulier dans la recherche d'information scientifique. En exposant les données - données bibliographiques ou autres- converties dans le standard RDF, celles-ci peuvent être publiées

et liées avec d'autres données, ce qui permet de naviguer entre les données de la même manière qu'entre les pages web qui sont reliées avec des liens hypertextes. En exposant les données d'une archive ouverte en linked open data on expose les données en dehors des silos de données et on augmente le potentiel de réutilisation, permettant à d'autres projets, utilisateurs finaux, ou à des outils qui savent lire ces formats normalisés d'exploiter ces données. Certaines archives ouvertes exposent leurs données dans ce format via un sparql endpoint comme dans la plateforme Persée par exemple (<http://data.persee.fr/>). Pour augmenter le potentiel de réutilisation il est important de lier ses données avec des vocabulaires contrôlés, (Gottschling et Subirats 2014) qui sont communément utilisés et centraux dans le nuage du linked data³⁰ ou des sources de qualité plus spécifiques.

5 Conclusion

Cette réflexion doit être considérée comme un préalable à l'intégration progressive des fonctionnalités et des standards communs qui permettront de rendre l'archive ouverte plus attractive et de mieux répondre à de nouveaux besoins institutionnels et scientifiques. Elle pourra aussi permettre à d'autres archives ouvertes d'effectuer la même analyse.

Les technologies préconisées sont en cours d'implémentation dans différentes plateformes qui agrègent des données provenant des archives ouvertes et des entrepôts de données, et proposent des services à valeurs ajoutées. Il s'agit de :

- OpenAIRE³¹ – Europe
- National Institute of Informatics (NII)³² - Japan
- US Next Generation Repositories Implementers Group
- CARL Open Repositories Working Group³³ - Canada

Il faudrait que la grande majorité des entrepôts ou archives ouvertes adoptent ces technologies pour que les échanges puissent réellement s'opérer.

La plupart de ces technologies devraient être disponibles en standard sur les principaux logiciels libres permettant de construire des archives ouvertes ou des entrepôts de données (Dspace, Eprints, Dataverse...) et devraient sans doute aussi être développées dans les plateformes nationales (HAL), ou sur des outils moissonnant différentes sources comme BASE (Bielefeld Academic Search Engine).

Le 6 juin 2018, un workshop sur l'adoption des recommandations a mis en avant les priorités suivantes (SHEARER et al. 2018) :

³⁰<https://lod-cloud.net/>

³¹<https://www.openaire.eu/>

³²<https://www.nii.ac.jp/en/>

³³<http://www.carl-abrc.ca/advancing-research/institutional-repositories/open-repositories-working-group/>

- Travailler avec les développeurs des logiciels d'archives ouvertes pour faire connaître ces recommandations et développer les cas d'usages pour assurer une plus grande compréhension et un plus grand engagement de la communauté;
- Donner plus d'information à ces logiciels sur les technologies du rapport ;
- Aider à trouver des fonds de la part des organisations intéressées par ces technologies dans les différentes plateformes ;
- Maintenir un annuaire des implémentations dans les différents logiciels ;
- Assurer un suivi sur le développement des fonctionnalités de réseaux sociaux.

Ces actions, que le groupe de travail Next Generation of Repositories de COAR devra mettre en place, permettront de suivre les différentes implémentations pour réaliser une infrastructure distribuée et globale, associée à des services de qualité, qui bénéficiera à l'ensemble de la communauté académique.

Remerciements

L'auteur remercie l'équipe IST (Informatique Scientifique et Technique) de l'IRD pour leurs conseils et leurs contributions sur l'évolution technique de l'archive ouverte Horizon Plein textes

Références

- CAVET, Dominique (2014). Horizon l'archive institutionnelle de l'IRD. 45p. In *Archive nationale HAL, archives institutionnelles, quelle complémentarité ? Urfist de Bordeaux*. 2014/06/11, [En ligne]. Disponible à : http://weburfist.univ-bordeaux.fr/wp-content/uploads/2016/03/2014_URFIST-CAVET_IRD.pdf (Page consultée le 30 août 2018).
- COUPERIN. GTAO – Groupe “ Etude Enquête Archives Ouvertes”
COUPERIN :Résultats de l'enquête (2017). [En ligne]. Disponible à : http://www.couperin.org/images/stories/AO/Depouillement_enquete_AO2017.pdf (Page consultée le 30 août 2018).
- GOTTSCHLING, Maxie ; SUBIRATS, Imma (2014). 7 things you should know about Linked Data. [En ligne]. Disponible à : <https://www.coar-repositories.org/community/events/archive/repository-observatory-third-edition/coar-talks-ir-cris-interoperability/second-edition-linked-open-data/7-things-you-should-know-about-open-data/>. (Page consultée le 30 août 2018).
- MIT. MASSACHUSETTS INSTITUTE OF TECHNOLOGY (2016). Institute-wide Task Force on the Future of Libraries. Preliminary Report. [En ligne]. Disponible à : <https://future-of->

libraries.mit.edu/sites/default/files/FutureLibraries-PrelimReport-Final.pdf

(Page consultée le 30 août 2018).

RODRIGUES, Eloy; SHEARER, Kathleen; COAR Next Generation Repositories WG (2017). Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group », [En ligne]. Disponible à : <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf> (Page consultée le 30 août 2018).

ROSSI, Pier Luigi (2017). Numérisation et bibliothèques électroniques : bilan de 15 années de formations réalisées par l'IRD au bénéfice des professionnels de pays de l'Afrique francophone, 12 p. In *Colloque International du Cinquantenaire de l'Ecole des Bibliothécaires, Archivistes et Documentalistes (EBAD) : Les Sciences de l'Information Documentaire au Service de la Recherche, de la Formation, de l'Intégration et du Développement Durable*, Dakar (SEN), 2017/11/14-16.. [En ligne]. Disponible à : <http://www.documentation.ird.fr/hor/fdi:010071833> (Page consultée le 30 août 2018).

SHEARER Kathleen; RODRIGUES Eloy; BOLLINI , Andrea, CABEZAS, Alberto; CASTELI, Donatella; CARR, Les; CHAN, Leslie; HUMPHREY, Chuck; JOHNSON, Rick; KNOTH, Petr Knoth, MANGHI, Paolo MATIZIROFA, Lazarus; PERAKAKIS, Pandelis; SCHIRRWAGEN, Jochen; SMITH, Tim; VAN DE SOMPEL, Herbert; WALK, Paul; WILCOX, YAMAJI, Kazu. (2018). Next Generation Repositories. Scaling up repositories to a global knowledge commons ». Zenodo. [En ligne]. Disponible à : <https://doi.org/10.5281/zenodo.1306452>. (Page consultée le 30 août 2018).