

Vers une architecture de réseaux d'archives en libre accès (archives ouvertes) dans le domaine des sciences et de la technologie agricoles

Par Imma Subirats, Irene Onyancha, Gauri Salokhe et Johannes Keizer

Organisation des Nations unies pour l'alimentation et l'agriculture (FAO)

Rome, Italie

{Imma.Subirats, Irene.Onyancha, Gauri.Salokhe,
Johannes.Keizer}@fao.org

Résumé. Le Réseau AGRIS (Système international d'information pour les sciences et la technologie) est le fruit d'une initiative internationale née de la collaboration entre différentes institutions, dont l'objectif est de promouvoir le libre accès à l'information sur la science et la technologie dans l'agriculture et ses domaines connexes. Ce document montre comment, au sein du Réseau AGRIS, l'utilisation des modèles d'archives ouvertes (*Open Archives – OA*) et les initiatives pour des archives ouvertes (*Open Archives Initiative – OAI*) permettent de résoudre les problèmes de diffusion et d'échange des résultats de la recherche agricole. Le manque de possibilités d'échange d'informations appropriées entre les chercheurs en sciences et technologie agricoles est un inconvénient majeur qui limite leur capacité à prendre efficacement en charge les questions du développement agricole. Les modèles OA promeuvent la diffusion des résultats de la recherche aux niveaux international, national et régional, supprimant ainsi les barrières imposées par les protocoles traditionnels qui régissent la publication scientifique. Le présent document explique comment relever le défi de l'accessibilité, de la disponibilité et de l'interopérabilité de l'information par l'échange d'expériences et des résultats de la recherche agricole.

Mots-clés : AGRIS, Recherche agricole et technologie, Archives ouvertes, Initiative pour des archives ouvertes, Fournisseurs de données, Fournisseurs de services, Profil d'application AGRIS, Systèmes d'organisation des connaissances, Interopérabilité

1 Introduction

Opérationnel depuis 1975, le réseau AGRIS est le fruit d'une initiative internationale lancée très tôt par différentes institutions oeuvrant, en étroite collaboration les unes avec les autres, pour la création d'un

système commun d'information en libre accès, au service de la science et de la technologie dans l'agriculture et ses domaines connexes. Jusqu'à la fin des années 1990, les travaux d'AGRIS consistaient essentiellement en une base de données bibliographiques centralisée et des produits associés. Mais depuis 2000, le réseau AGRIS oriente de plus en plus ses efforts vers le renforcement et la décentralisation de ses capacités auprès de ses centres de ressources participatifs, en les dotant des moyens nécessaires pour améliorer la gestion de l'information agricole au sein de leurs propres institutions. AGRIS a notamment pour principale mission de s'attaquer aux problèmes que pose l'accès aux publications scientifiques et technologiques, aux travaux universitaires mais aussi et surtout à la « littérature grise » (souterraine, non conventionnelle). En effet, si la littérature grise contribue substantiellement à la recherche agricole, en particulier dans les pays en développement, elle reste toujours insuffisamment accessible à la communauté agricole.

L'accélération des mutations technologiques offre de nouvelles opportunités en matière de partage de données, d'informations et de connaissances. Au cours de ces dernières années, le monde de la communication a connu plusieurs évolutions, dont les plus marquantes restent le libre accès et les initiatives pour des archives ouvertes. On entend par libre accès « la mise à disposition gratuite de documents sur le réseau public Internet, permettant aux utilisateurs de lire, télécharger, copier, transmettre, imprimer, rechercher – ou faire un lien vers – le texte intégral de ces documents (notamment des articles d'origine universitaire et scientifique), de les disséquer pour les indexer, de s'en servir pour l'élaboration de logiciels ou à toute autre fin légale, sans barrières financières, juridiques ou techniques autres que celles indissociables de l'accès et de l'utilisation d'Internet » (Initiative de Budapest pour l'accès ouvert – BOAI -, 2002). Le modèle d'accès libre permet à la fois une diffusion plus large des résultats de la recherche et une visibilité maximale, tout en contribuant, dans le même temps, à la suppression des contraintes qui pèsent communément sur les protocoles classiques de diffusion de la littérature scientifique. Il offre aux chercheurs la possibilité d'accéder à des documents en texte intégral via : les revues et magazines d'édition, qui rendent immédiatement accessibles leurs articles dès leur publication ; et l'auto-archivage, qui permet aux auteurs de « mettre eux-mêmes en ligne des documents numériques sur un site Internet accessible au public, de préférence dans

des archives ouvertes, conformément aux principes de l'OAI » (FAQ sur l'auto-archivage, 2006). L'Initiative pour des archives ouvertes (OAI) met au point et promeut des solutions d'interopérabilité visant à faciliter la diffusion efficiente de contenus. Elle se fonde sur le principe selon lequel l'interopérabilité permet de fédérer des archives ouvertes décentralisées, encourageant ainsi le développement de services à valeur ajoutée tels que les portails, les passerelles et les moteurs de recherche spécialisés, avec à la clé une plus grande visibilité des archives ouvertes. L'objectif ultime est de surmonter les obstacles à l'interopérabilité grâce aux archives ouvertes et au libre accès pour tous les documents numériques (Van de Sompel et Lagoze, 2000).

Il existe un lien très étroit entre les principes du réseau AGRIS et le modèle de publication en libre accès, d'une part, et l'Initiative pour des archives ouvertes, d'autre part. Avec la publication en libre accès, la communauté AGRIS entend non seulement renforcer la capacité de ses centres ressources à diffuser et partager les résultats de leurs recherches, mais aussi favoriser l'accessibilité et la visibilité des publications des chercheurs sur l'agriculture. Il va sans dire que le réseau AGRIS devra également répondre aux exigences et conditions spécifiques régissant les archives ouvertes.

Le présent document donne une définition et une description de ce qu'est une architecture de haut niveau, fondée sur le principe d'archives ouvertes accessibles à la communauté AGRIS. L'architecture comprend trois composantes : (i) la création de contenus avec des normes de description convenues, (ii) la « moisson » de contenus à l'aide de formats d'échange normalisés et (iii) la fourniture de services à valeur ajoutée aux utilisateurs grâce aux contenus échangés. Ce document est ainsi divisé en trois sections : la Section 3 donne un aperçu de l'architecture de la gestion des processus opérationnels (*workflow*) où sont décrites les différentes composantes et les relations qui les régissent. La Section 4 explique le principal objectif de cette nouvelle architecture, fondée sur le modèle du libre accès, et évoque les deux acteurs que sont le fournisseur de données et le fournisseur de services, leurs rôles respectifs et leurs formats d'échange normalisés, le protocole d'intégration via la « moisson » de métadonnées (données descriptives). La Section 5 explique en détail la justification et l'utilisation des normes d'échange, y compris les métadonnées, les Systèmes d'organisation des connaissances (*Knowledge Organization*

Systems – KOS) tels que les thésaurus et les formats communs d'intégration qui constituent l'élément structural du Web sémantique. Enfin, la conclusion, dans la Section 6, rappelle les prochaines étapes concernant la mise en oeuvre de cette architecture.

2 Objectifs

Le réseau AGRIS veut promouvoir le modèle de publication en libre accès et appliquer les normes et méthodologies OAI, puisque celles-ci contribuent sensiblement à la réalisation de l'objectif général que s'est fixé depuis longtemps AGRIS, à savoir, améliorer l'accessibilité et l'échange des informations et connaissances en science et en technologie agricoles entre les pays en développement. Parmi ses objectifs spécifiques figurent l'amélioration de la visibilité et de l'accessibilité des contenus, y compris ceux qui étaient autrefois inaccessibles, et la promotion de normes communes spécifiques ; deux aspects essentiels qui fondent les principes majeurs du réseau AGRIS. L'exploitation d'une nouvelle architecture au moyen d'outils permettant le libre accès constitue un avantage inestimable pour la communauté des chercheurs dans le domaine agricole, en ce sens qu'elle va bénéficier à la fois d'une meilleure visibilité et d'un lectorat plus large. De plus, les archives ouvertes facilitent l'accès de potentiels lecteurs qui, autrement, n'ont qu'un accès limité aux documents, en raison du coût élevé des principales revues. L'initiative permet non seulement de relier la recherche locale à celle internationale, mais également de mieux appréhender la production de la recherche agricole d'un pays. Le nouveau modèle offre une architecture dans laquelle la communication peut être définie entre deux niveaux de partenaires : les fournisseurs de données – ceux qui ont des métadonnées à exposer – et les fournisseurs de services – qui récoltent/moissonnent ces métadonnées. La communication entre les deux est facilitée par une solution d'interopérabilité appelée « moisson de métadonnées », qui permet aux fournisseurs de données d'exposer leurs métadonnées, via une interface ouverte, auprès des fournisseurs de services qui utilisent les métadonnées pour améliorer les services qu'ils offrent à leurs utilisateurs. La bonne qualité descriptive de l'information agricole dépendra de l'utilisation de normes communes spécifiques au domaine et au réseau. Parmi ces normes figurent le Profil d'application AGRIS (*AGRIS Application Profile – AGRIS AP*) et des Systèmes

d'organisation des connaissances tels que le Thésaurus AGROVOC, expliqué plus en détail à la Section 5. Elles offrent aux fournisseurs de services spécialisés en sciences agricoles plus de possibilités d'améliorer sensiblement leurs services à partir des métadonnées récoltées. Les normes et les outils seront promus au sein du réseau AGRIS afin de garantir le succès de la nouvelle architecture.

3 Éléments de l'architecture réseau : un nouveau processus opérationnel (*workflow*)

Comme le montre la Figure 1, l'architecture réseau AGRIS proposée ici décrit les principaux éléments et la dynamique du flux d'informations attendue au sein du réseau. De par sa conception, le processus opérationnel comprend trois activités principales : la gestion de contenus, l'exposition de métadonnées et la fourniture de services à valeur ajoutée.

Par activité de gestion de contenus on entend la capture, la description et le stockage de l'information au moyen d'outils et de méthodologies conçus à cet effet. Une liste de normes, de méthodes et de logiciels appropriés est mise en ligne sur le site Internet d'AGRIS et sur celui de la FAO dédié aux Normes de gestion de l'information agricole (*AIMS*¹). On peut citer comme nouvelle étape du processus de capture l'auto-archivage par les chercheurs de leurs documents en texte intégral. Les centres AGRIS vont promouvoir l'archivage de documents en texte intégral, soit par les auteurs eux-mêmes, soit par leurs représentants au sein de leurs institutions respectives.

¹ Site de la FAO dédié aux Normes de gestion de l'information agricole : <http://www.fao.org/aims/>. Dernière mise à jour en janvier 2007.

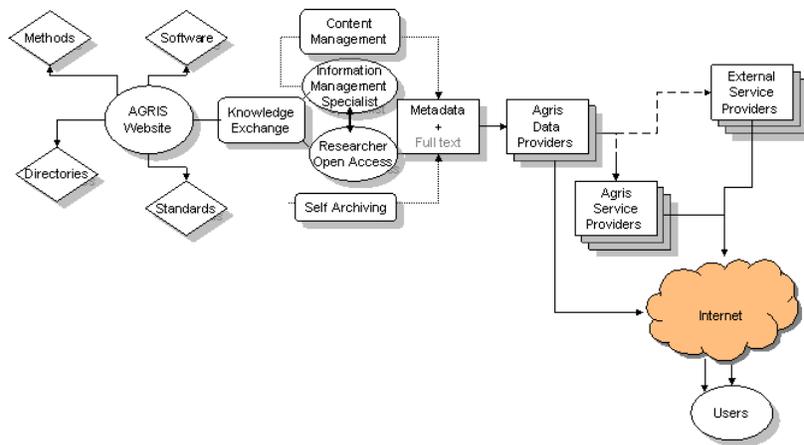


Figure 1. Le nouveau processus opérationnel (*workflow*) du réseau AGRIS

La deuxième partie du processus opérationnel consiste à exposer les métadonnées. Cela implique que les partenaires du réseau AGRIS deviennent des fournisseurs de données et interopèrent avec les fournisseurs de services en utilisant les normes et procédures OAI. Pour en garantir la visibilité, les fournisseurs de données devront s'inscrire et décrire leurs « collections » dans un registre en ligne sur le site Internet d'AGRIS. Le réseau AGRIS veut aller au-delà de l'échange de métadonnées en format Dublin Core simplifié (condition préalable pour faire partie de la communauté OAI), en recommandant l'utilisation du Profil d'application (PA) AGRIS, un format d'échange enrichi qui permet de récolter et d'échanger du contenu agricole. Le profil d'application et les systèmes d'organisation des connaissances AGRIS sont les principaux facteurs pris en compte dans la description de contenu. Ils permettent aux fournisseurs de services spécialisés de sélectionner et de récolter des données de haute qualité qui, en dernier ressort, améliorent la qualité de leurs propres services. La dernière étape de la gestion du processus opérationnel (*workflow*) concerne les fournisseurs de services spécialisés, qui apportent de la valeur ajoutée aux métadonnées récoltées en mettant à disposition des plates-formes

via lesquelles les chercheurs auront la possibilité d'interagir et de partager l'information. Le réseau AGRIS a créé des communautés disciplinaires qui partagent des activités et des services communs ; ces groupes constituent déjà une base à partir de laquelle les tâches et responsabilités des différents fournisseurs de services pourront être définies.

De même, les fournisseurs de services auront la possibilité de créer des passerelles spécialisées dans les différentes disciplines des sciences agricoles. En outre, les documents entreposés par les fournisseurs de données seront également moissonnés par d'autres fournisseurs de services qui sont en dehors du réseau AGRIS. L'image que présente, dans l'ensemble, le nouveau réseau AGRIS est celle d'un espace dans lequel tous les éléments décrits ci-dessus jouent un rôle actif dans le partage des métadonnées et des connaissances. La mise en oeuvre des réseaux d'archives ouvertes reste l'élément principal de la nouvelle architecture ; elle permettra aux chercheurs en agriculture d'accéder aux résultats de leurs recherches et de les partager.

4 Les fournisseurs de données : adopter le modèle de publication en libre accès

L'adoption du modèle de publication en libre accès exigera, au niveau institutionnel et de la part des fournisseurs de données, qu'ils mettent en place des mécanismes capables de gérer l'utilisation des entrepôts d'archives numériques. Ce sont des archives ouvertes pouvant héberger du contenu, par exemple, des articles déjà publiés (validés par un comité d'experts scientifiques ou postpublications), des articles non encore publiés (non validés ou prépublications), des thèses, des manuels, du matériel didactique, des supports de cours ou d'autres documents que les auteurs ou leurs institutions d'origine souhaitent mettre à la disposition du public sans barrières financières, ni entraves à leur accessibilité. Les infrastructures techniques nécessaires à la création d'entrepôts institutionnels sont déjà disponibles, comme le montrent à la fois leur coût d'installation limité et la grande diversité des logiciels en accès libre (*free open source*) qui respectent les normes requises. La mise en place de stratégies et politiques institutionnelles en matière de Technologies/Gestion de l'information et de la communication (TIC)/GIC, intégrant une gestion efficace des processus opérationnels (*workflow*) d'édition et de documentation, contribuera

également à favoriser davantage l'institutionnalisation du modèle de publication en libre accès. Parmi les autres questions à prendre en compte figure celle des droits associés à l'auto-archivage par les auteurs d'articles déjà publiés. La base de données ROMEO (*Rights Métadonnées for Open Archiving*), accessible à partir du site Web [SHERPA](#) (*Securing a Hybrid Environment for Research Preservation and Access*), donne des informations sur l'auto-archivage et la politique des droits d'auteur (*copyright*) des éditeurs. Pour faciliter ce processus, il importe que les avantages de la publication en accès libre (OA) soient clairement mis en avant, défendus et démontrés auprès des chercheurs et décideurs, qui sont à même de mettre en place des politiques capables de favoriser et/ou d'imposer le modèle OA. Les bibliothèques au sein des instituts et organisations devraient être à l'avant-garde de ce mouvement, en jouant le rôle important de facilitateur pour l'adoption de ce modèle, à travers des campagnes de sensibilisation des chercheurs et décideurs aux avantages et au rôle de la publication en libre accès dans la promotion de la visibilité de ce modèle et de son impact sur la production de la recherche institutionnelle.

4.1 L'exposition de métadonnées

Dans la nouvelle architecture, les fournisseurs de données ne se résument pas aux seuls centres de ressources AGRIS : toute institution de recherche et de technologie agricole désireuse de publier sa production scientifique sur le réseau AGRIS peut s'y inscrire, à titre de partenaire, et utiliser les normes et technologies communes du réseau. Il faut sensibiliser le fournisseur de données au fait que les entrepôts ou les archives ouvertes ne sont pas seulement des bases de données contenant des documents en texte intégral et leurs métadonnées. La Figure 2 décrit les différentes actions que nécessite la fourniture de données : le processus de capture de contenus au cours duquel les chercheurs peuvent archiver eux-mêmes leurs documents ; le catalogage-indexation des données par les gestionnaires de l'information, et l'exposition des métadonnées sous différents formats. Trois différents types d'entrepôts sont définis dans le réseau AGRIS : l'entrepôt dynamique – une « couche » logicielle qui permet la compatibilité OAI et qui agit comme interface entre les archives ouvertes et les fournisseurs de services OAI. La « couche » reçoit les requêtes OAI et génère des réponses dynamiques ; l'entrepôt statique – qui est utilisé pour de petites collections de métadonnées. Un entrepôt

statique peut être généré périodiquement par un script (suite de programmes) qui extrait l'information à partir d'une base de données existante (Hochstenbach, Jerez et Van de Sompel, 2003) ; et enfin l'entrepôt hébergé - la solution des institutions qui ne peuvent exposer sur Internet leurs propres métadonnées : d'autres institutions le font à leur place. Un fournisseur de données peut exposer des métadonnées en différents formats ; ce qui permet à différents fournisseurs de services de récolter ces métadonnées, de disposer de plus de canaux de collecte et de jouir de plus de visibilité. Le format de base des métadonnées nécessaire pour faire partie d'une communauté OAI s'appelle Dublin Core non qualifié ; toutefois, ce n'est pas une condition suffisante pour faire partie de la communauté scientifique agricole. Le PA AGRIS a été créé spécialement pour améliorer la qualité et l'interopérabilité des métadonnées dans ce domaine. Les fournisseurs de données peuvent proposer aux utilisateurs le format Dublin Core non qualifié et le PA AGRIS pour qu'ils soient visibles par un plus grand nombre de moteurs de recherche et de fournisseurs de services spécialisés dans le réseau AGRIS et au-delà. L'initiative d'AGRIS va encourager l'échange de métadonnées via la norme PA AGRIS, afin de garantir la bonne qualité descriptive des ressources d'information agricole.

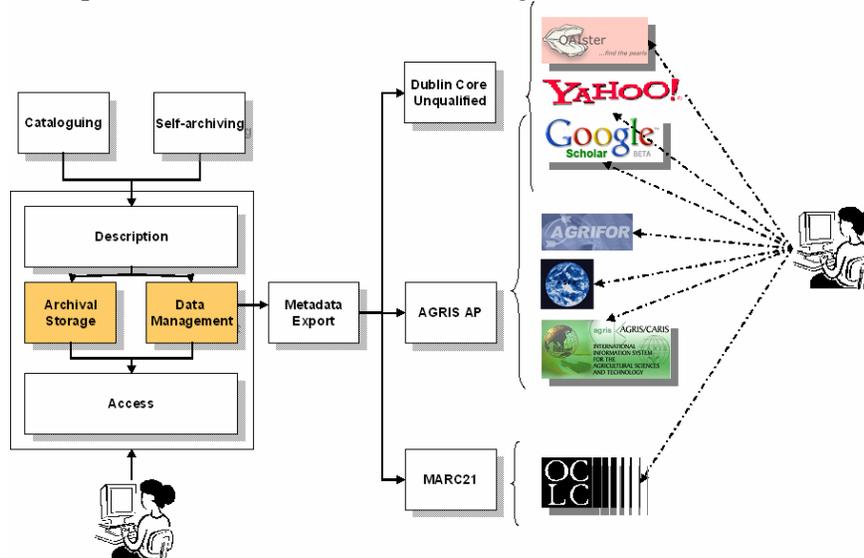


Figure 2. Les fournisseurs de données

5 La récolte de métadonnées : créer un canal de communication

La récolte de métadonnées est une solution d'interopérabilité permettant aux fournisseurs de services de communiquer avec les fournisseurs de données. Cette solution repose sur une appellation uniforme (métadonnées normalisées), une syntaxe commune et un protocole régissant la récolte de métadonnées. Le protocole OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) a été mis au point pour permettre de collecter les métadonnées des enregistrements stockés dans les archives ouvertes. Ce protocole est basé sur une architecture client-serveur (les fournisseurs de services demandent à accéder aux métadonnées qui se trouvent dans les archives ouvertes des fournisseurs de données). Ainsi, les fournisseurs de services peuvent utiliser des métadonnées issues de différentes archives pour mieux confectionner leurs services. Au cours des cinq dernières années, le protocole OAI-PMH s'est fait largement connaître comme moyen de connexion avec des archives ouvertes décentralisées. La principale raison réside dans la simplicité de sa mise en oeuvre. De même, plusieurs outils ont été mis au point pour faciliter la création, la gestion et l'interopérabilité des métadonnées des archives ouvertes compatibles OAI. L'appellation uniforme est rendue possible par la normalisation de la sémantique des métadonnées. Une syntaxe commune, XML, permet de représenter et de transférer des ensembles de métadonnées spécifiques à ces archives. La plupart des logiciels courants permettent de générer des enregistrements XML.

6 Les fournisseurs de services : améliorer l'accès

Le réseau AGRIS a aujourd'hui plusieurs fournisseurs de services, dont la base de données AGRIS à la FAO, la base de données commerciales AGRIS auprès de *Ovid Technologies*, le service AGRI2000 de *Sistema de Informacion y Documentacion Agropecuaria de las Americas* (SIDALC), le centre d'information virtuelle du *Consultative Group on International Agricultural Research* (CGIAR), ainsi que de nombreux autres petits fournisseurs régionaux. La nouvelle architecture présente un environnement d'une plus grande simplicité et d'une meilleure efficacité pour la mise au point de services spécifiques. Selon la portée et la nature de l'information souhaitée, c'est à partir de plusieurs archives ouvertes que les fournisseurs de services AGRIS récoltent des enregistrements qui ne renferment que des métadonnées, et non des

documents en texte intégral. Le succès d'un fournisseur de services dans sa moisson de données réside dans les schémas des métadonnées, l'état des collections et la fréquence de mise à jour des archives ouvertes, mais aussi la durée de la disponibilité des archives, les conditions d'utilisation des enregistrements, la mention de la source et le lien du fournisseur de services vers le fournisseur de données. En 2003, Foulonneau et Dawson (Foulonneau et Dawson, 2003) ont décrit les aspects clés à prendre en compte dans le travail du fournisseur de services, la plupart de ces aspects étant déjà pris en considération par le réseau AGRIS, comme expliqué ci-après. La collecte de données et l'analyse : la plupart des membres du réseau AGRIS collaborent déjà au niveau de la base de données AGRIS ; le schéma commun de métadonnées : le PA AGRIS est la norme établie d'échange de l'information agricole ; la granularité des descriptions : le PA AGRIS fournit des jeux de métadonnées appropriées qui permettent d'enrichir la description des documents agricoles. Il propose des schémas spécifiques à l'agriculture ; des passerelles vers les métadonnées : il existe des passerelles depuis le PA AGRIS vers d'autres schémas de métadonnées ; les problèmes de terminologie : plusieurs systèmes d'organisation des connaissances en sciences agricoles ont été mis au point et sont aujourd'hui utilisés ; un environnement multilingue : c'est l'une des principales caractéristiques du réseau AGRIS.

Les métadonnées récoltées peuvent être associées à des textes entiers dont le fournisseur de services accepte d'autoriser l'accès. Il est possible d'avoir différents fournisseurs de services : portails spécialisés, bibliothèques électroniques, passerelles ou pages web n'exposant que des métadonnées. Les services que l'on peut mettre au point avec les données récoltées sont les suivants : l'indexation de données ; la recherche avancée et la navigation ; les statistiques de consommation ; l'analyse syntaxique des citations ; la possibilité de faire des commentaires sur les documents ; l'assistance multilingue ; l'authentification des utilisateurs ; de nouveaux fils d'actualité agrégés ; des groupes de discussion (forums) électronique ; des alertes automatiques par courrier électronique (mél), c'est-à-dire, des services de diffusion sélective de l'information (DSI) ; des fils RSS sur les conférences, événements et cours de formation dans différentes disciplines ; des services personnalisés (par l'utilisateur).

7 La mise en réseau pour favoriser l'interopérabilité : le site Internet d'AGRIS et le répertoire d'archives ouvertes AGRIS

Le site Internet d'AGRIS se veut un canal de communication entre le fournisseur de données et services et l'utilisateur. Il servira de passerelle vers différentes ressources d'information, y compris un entrepôt/réservoir des productions des groupes d'étude, le répertoire d'archives ouvertes AGRIS, des normes et outils de gestion de données, ainsi que d'autres informations pertinentes. Ce site joue un rôle important dans la nouvelle architecture réseau AGRIS en gestation et répond parfaitement à la priorité accordée au plaidoyer. Le site est un outil majeur de promotion d'une approche globale plus cohérente de l'information agricole, dans le cadre de cette nouvelle initiative AGRIS.

Le répertoire des archives ouvertes d'AGRIS est une plate-forme où sont stockées toutes les archives AGRIS. Il sera interrogeable/consultable depuis le site Internet d'AGRIS et permettra d'identifier l'ensemble des fournisseurs de données de la communauté. Le répertoire va catégoriser et lister toute la panoplie d'archives ouvertes existantes, à l'aide d'un jeu de métadonnées qui en fournira les informations pertinentes. Il sera interrogeable en fonction de plusieurs critères (pays, lieu, thème, format de métadonnées, logiciel, type, documents concernés, entre autres).

8 Garantir la qualité dans la création des métadonnées

La description de contenu par la création de métadonnées de haute qualité et l'utilisation de terminologies standard est la condition de base d'une gestion efficace de contenus et d'une conception de services à valeur ajoutée. Les normes de métadonnées, à l'instar du format Dublin Core (DC) et du profil d'application AGRIS (PA AGRIS), fournissent les mécanismes nécessaires au partage de l'information de manière normalisée, en recommandant l'utilisation d'une sémantique commune et de syntaxes interopérables. Les normes OAI exigent l'utilisation du format Dublin Core non qualifié pour l'exposition de métadonnées via le Protocole d'échange OAI-PMH, mais cette condition reste insuffisante pour la communauté scientifique agricole. En effet, l'utilisation du format Dublin Core non qualifié empêche le développement de services efficaces à valeur ajoutée du fait,

notamment, de la médiocre qualité des données et de la perte d'informations. Comme indiqué plus haut, le PA AGRIS, norme richement définie, a été créé pour améliorer la qualité et l'échange des métadonnées au sein de la communauté des sciences agricoles. Une composante métadonnées de haute qualité facilite beaucoup l'agrégation de contenus et, partant, leur extraction. Du point de vue de la syntaxe et de la sémantique utilisées, les métadonnées normalisées facilitent également l'interopérabilité. En revanche, une mauvaise qualité de métadonnées signifie qu'une ressource restera en grande partie invisible dans le répertoire et donc inutilisée (Barton, Currier et Hey, 2003). Les fournisseurs de services les plus spécialisés se serviront du PA AGRIS comme format de base pour leur moisson de métadonnées.

Pour garantir un niveau minimal de qualité, il est nécessaire de convenir d'un certain nombre de champs de métadonnées et de l'usage de vocabulaires pertinents et contrôlés. Le PA AGRIS satisfait à cette exigence, car il rend obligatoires les cinq éléments requis et promeut l'utilisation de vocabulaires contrôlés spécifiques à l'agriculture, comme AGROVOC, les codes CABI, etc.

8.1 Le Profil d'application (PA) AGRIS : répondre aux exigences agricoles

Dublin Core est un format de métadonnées dont la structure est née d'un consensus international. Dans un ensemble de métadonnées Dublin Core (DCMES), quinze éléments y sont définis pour la description et la découverte primaire de ressources, dont tous sont recommandés mais non obligatoires. Le format Dublin Core (DC) a été enrichi grâce à l'ajout d'autres éléments optionnels, de qualificatifs, d'identifiants et de terminologies. Dublin Core non qualifié est le format de métadonnées requis pour une interopérabilité basique des protocoles d'échange OAI-PMH. Toutefois, dans des domaines spécialisés et au sein des communautés scientifiques, d'autres spécifications de métadonnées peuvent être requises. Par exemple, il est possible que l'on doive décrire des ressources de structures complexes et de manière spécialisée, comme dans le cas du réseau AGRIS (Liang, Salokhe, Sini et Keizer, 2006). Le PA AGRIS a été conçu pour le secteur agricole, qui a des caractéristiques et des exigences spécifiques. C'est une norme créée dans le seul but de faciliter la description,

l'échange et l'extraction de documents ressemblant à des objets d'information (*Document-like information objects – DLIO*) sur l'agriculture. Le PA AGRIS permet la mise en commun d'informations réparties dans de nombreux systèmes bibliographiques. Non seulement il fournit des qualificatifs/attributs supplémentaires au jeu d'éléments du format Dublin Core simplifié, mais encore il permet les extensions locales, considérées comme nécessaires à une description exhaustive des ressources d'information agricole. Mise au point à la faveur de la nouvelle vision AGRIS, cette norme améliore la qualité de la description des ressources d'information agricole, offrant ainsi au fournisseur de services davantage de possibilités de traitement.

L'utilisation du PA AGRIS par un fournisseur de données au sein de la communauté AGRIS présente certains avantages pour le fournisseur de services : enrichissement de la couche d'interopérabilité pour aider à la découverte de ressources secondaires et tertiaires, grâce à l'utilisation d'un ensemble d'éléments reconnus ; flexibilité dans la fourniture de réponses plus pertinentes aux questions posées, grâce à une recherche ciblée ; des métadonnées enrichies et reconnues par tous ; une information suffisante et appropriée sur le contenu des ressources, grâce à l'usage de qualificatifs spécifiques à l'agriculture (schémas de classifications, thésaurus, etc.) ; et enfin, possibilité de naviguer à travers les ressources par thème, pays, année, etc. Bien que pour être « officiellement » compatible OAI, le fournisseur de données doive se conformer aux normes Dublin Core simplifié, s'agissant du réseau AGRIS, qui fournit des « services spécialisés », il est recommandé que le fournisseur de données soit également compatible PA AGRIS.

8.2 Systèmes d'organisation des connaissances : aide à la navigation sémantique

Les Systèmes d'organisation des connaissances sont des structures bien agencées, dont les exemples les plus courants sont : les fichiers autorité, les systèmes/schémas de classification, les cartes conceptuelles, les vocabulaires contrôlés, les dictionnaires, les fiches descriptives/signalétiques, les glossaires, les ontologies (ensemble structuré de concepts), les en-têtes de sujet (mots-matières), les répertoires toponymiques. Ces ressources sont associées à un continuum (ensemble de variables attestées dans une communauté linguistique), en fonction de leur explication sémantique et de leur

interprétation par la machine. Les systèmes d'organisation des connaissances sont des mots ou expressions extraits d'un ensemble de termes normalisés, qui permettent de résoudre deux principales problématiques : celle de deux ou plusieurs termes pouvant renvoyer au même concept, comme les mots anglais *Fishing vessels/Fishing boats* (bateaux de pêche/navires de pêche) ou *Health risks/Health hazards* (risques sanitaires/menaces sanitaires) ; et celle de deux ou plusieurs mots ayant la même orthographe mais des sens différents, par exemple, vaisseau (sanguin) et vaisseau (navire de pêche). L'utilisation des systèmes d'organisation des connaissances permet de créer des services davantage spécialisés, comme la navigation par mot-clé et par pays/zone géographique, la recherche par type de document (brevets, ouvrages, etc.) et de limiter les résultats de la recherche à une ou plusieurs années spécifiques ou à une seule langue, ainsi que la navigation sémantique à l'intérieur d'un ensemble de résultats en fonction des mots-clés identifiés dans les vocabulaires contrôlés. Ces services sont opérationnels grâce à l'utilisation de vocabulaires contrôlés et de leur mention explicite dans les données échangées, une possibilité qu'offre la norme PA AGRIS. Pour ce qui concerne le secteur agricole, l'utilisation de formats lisibles par ordinateur, comme le XML, et les vocabulaires normalisés d'échange de métadonnées représentent l'avenir du « Web sémantique ». Ce sont les fondements sur lesquels peuvent s'appuyer les applications permettant de tirer parti des avantages de l'expressivité sémantique dans les métadonnées bibliographiques, afin d'améliorer l'accès à l'information agricole.

9 Conclusion et prochaines étapes

Le présent document propose l'architecture adaptée aux archives ouvertes du réseau AGRIS. Cette architecture associe l'expérience et l'histoire du réseau AGRIS, au nouveau modèle de publication en libre accès et à l'Initiative internationale pour des archives ouvertes. La mise en oeuvre d'une telle architecture va libérer sur Internet un espace très important dans lequel des milliers de fournisseurs de données pourront mettre en ligne leurs publications et à partir duquel de nombreux services seront offerts aux utilisateurs. Cela promet d'être une grande réussite, comparable à celle d'autres grands projets comme PubMed dans le secteur médical. Qui plus est, cela nécessitera beaucoup moins d'investissements pour mettre au point des outils et standards communs, des outils *open source* (en libre accès) de gestion de données

et de métadonnées pour les fournisseurs de données, des schémas de données, des vocabulaires et tout un ensemble d'autres outils garantissant la compatibilité OAI des systèmes existants. Dans tous ces domaines, beaucoup d'efforts ont déjà été faits, dont une partie non négligeable est à mettre à l'actif du réseau AGRIS, notamment : le schéma de métadonnées Dublin Core (DC) et les outils généraux OAI, le profil d'application AGRIS au format DC, les vocabulaires d'échange/systèmes d'organisation des connaissances tels que AGROVOC, les outils de gestion de métadonnées WebAGRIS, l'adaptation de DSpace aux besoins spécifiques du réseau AGRIS network et la mise au point d'un moteur de recherche *open source* AGRIS (via Lucene) pour les fournisseurs de services.

Pour les prochaines étapes, il s'agira de trouver un consensus pour son acceptation par les principales parties prenantes du réseau AGRIS, puis de planifier les différents projets touchant à des composantes spécifiques de l'architecture et, enfin, d'exécuter les projets mis en place dans les différents pays et régions membres du réseau.

Références

- [1] Deuxième Consultation AGRIS sur la gestion de l'information agricole (2002) - Une stratégie en faveur d'un réseau international d'information sur les sciences et la technologie agricoles dans le cadre du portail WAICENT (Centre mondial d'information agricole). <http://www.fao.org/DOCREP/MEETING/005/AC502E.HTM>
Dernière mise à jour en janvier 2007.
- [2] Initiative de Budapest pour l'accès ouvert (Budapest Open Access Initiative – BOAI). <http://www.soros.org/openaccess/read.shtml>
Dernière mise à jour en janvier 2007.
- [3] Eprints: Supporting Open Access. Selfarchiving FAQ. <http://www.eprints.org/openaccess/self-faq/#self-archiving> Dernière mise à jour en janvier 2007.
- [4] Van de Sompel, Herbert ; Lagoze, Carl (2002). The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine, 6(2). <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html> Dernière mise à jour en janvier 2007.

- [5] Site Internet de la FAO dédié à la gestion de l'information agricole (AIMS) : <http://www.fao.org/aims/>. Dernière mise à jour en janvier 2007.
- [6] Hochstenbach, Patrick ; Jerez, Henry ; Van de Sompel, Herbert (2003). The OAI-PMH Static Repository and Static Repository Gateway. JCDL 2003. <http://public.lanl.gov/herbertv/papers/jcdl2003-submitted-draft.pdf> Dernière mise à jour en janvier 2007.
- [7] Foulonneau, Muriel ; Dawson, David. (2003) Expert Report 3 – Open Archives Initiative – Protocol for Metadata Harvesting - Practices of cultural heritage actors. http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf Dernière mise à jour en janvier 2007.
- [8] Barton, Jane ; Currier, Sarah ; Hey, Jessie M. N. (2003). Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice. Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications Washington (USA). <http://eprints.rclis.org/archive/00001972/> Dernière mise à jour en janvier 2007.
- [9] Berners-Lee, T. et Connolly, D. (1995). Hypertext Markup Language - 2.0 http://www.w3.org/MarkUp/html-spec/html-spec_5.html Dernière mise à jour en janvier 2007.
- [10] Liang, Anita ; Salokhe, Gauri ; Sini, Margherita ; Keizer, Johannes (2006) "Towards an infrastructure for semantic applications: Methodologies for semantic integration of heterogeneous resources" Publication prévue dans le bulletin trimestriel *Cataloging and Classification*.