

La fusion annoncée de DuraSpace et Lyris devrait booster le développement des technologies Open Source au service des bibliothèques de recherche



Analyse I/IST-n°13- Février 2016

En bref

Deux organisations américaines à but non-lucratif et développant des applications pour les bibliothèques, Lyris et DuraSpace ont annoncé le 27 janvier leur fusion. Le principal intérêt de ce rapprochement est l'accélération probable du développement et de la pénétration des applications Open Source Fedora, DSpace et Vivo dont l'audience a vocation à s'internationaliser.

L'analyse d'Intelligence IST

Lyris et DuraSpace sont deux organisations dont le nom est à peu près inconnu en Europe. Lyris est une « coopérative » fournissant outre-Atlantique des services aux bibliothèques, essentiellement autour des problématiques de numérisation de masse, de gestion d'archives et de capacités d'hébergement. Ses revenus (cotisations des membres + vente de services à ces derniers) ont été en 2014 de 74,7 M\$ et ses réserves financières se montent à 15 M\$. Créée il y a vingt ans Lyris a une assise (1200 membres) limitée aux Etats-Unis et est positionnée sur une gamme de services solides mais peu innovants (essentiellement axés sur la préservation des collections). Le seul intérêt du « cas Lyris » pour un observateur européen est l'illustration qu'il donne de la vitalité du modèle coopératif aux Etats-Unis dans l'offre de services aux bibliothèques¹.

Le mariage annoncé de Lyris et de DuraSpace est le mariage du nain et du géant : DuraSpace, fondé en 2009 affiche des revenus 2014 de seulement 1,63 M\$. Mais il est positionné sur une offre de services très innovante, qui surtout se base sur les technologies Open Source et connaît déjà une certaine pénétration en Europe. L'intérêt de Lyris pour DuraSpace tient donc à sa volonté de rajeunir sa palette de services, de se doter d'un nouveau foyer de croissance plus centré sur les besoins des bibliothèques de recherche et de s'ouvrir une fenêtre de visibilité sur l'international. Pour DuraSpace, l'adossement à Lyris lui garantit l'accès aux financements nécessaires à son développement et lui permettra de bénéficier de la notoriété et de la bonne pénétration de ce grand frère. Les développements dans le domaine du logiciel libre sont aux Etats-Unis souvent amorcés par des crédits provenant des agences de financement de la recherche (NIH, NSF) ou de fondations. Mais il s'agit là de financements non récurrents et la pérennité des développements n'est acquise que si une communauté d'utilisateur suffisante arrive à se former rapidement. C'est cette pérennité que l'adossement de DuraSpace à Lyris devrait garantir. Tout l'enjeu de ce rapprochement – et cette fois l'Europe est concernée puisque DuraSpace dispose de solides références sur le vieux continent – est la probable accélération du développement des outils collaboratifs de mise à disposition d'archives institutionnelles mais aussi la création de bases de connaissances interopérables sur les domaines d'expertise des chercheurs.

DuraSpace gère essentiellement trois projets collaboratifs de développement de logiciels ad hoc, tous trois fondés sur des technologies Open Source, DSpace, Fedora et Vivo. En conformité avec les logiques de l'Open Source et du logiciel libre, DuraSpace ne développe pas les différentes releases successives de ces outils puisque ce développement est collaboratif (tout membre d'une communauté centrée sur l'une ou l'autre des applications DuraSpace peut participer à son développement), DuraSpace n'ayant qu'un rôle de « stewardship » (d'organisation et d'animation) des communautés de développeurs.

¹Une vitalité que démontre aussi année après année les résultats d'OCLC (organisation « not-for-profit » de droit américain, mais très internationalisée) qui a enregistré en 2015 202,7 M\$ de revenus et dispose de réserves financières de 213,5 M\$.

Fedora (cf. : https://en.wikipedia.org/wiki/Fedora_Commons) est une « boîte à outils » de composants logiciels permettant aux institutions de recherche de développer des archives institutionnelles en mobilisant des technologies « Open Source » orientées objets. Fedora, initialement développé par Cornell University a été utilisé pour développer environ 1000 archives institutionnelles dans 90 pays. L'une de ses références parmi les plus prestigieuses est la National Library of Medicine, mais de très nombreuses universités de recherche ont bâti des applications autour de Fedora.

DSpace (initialement développé par HP et le MIT) est une suite logicielle « clé-en-main » permettant la création et l'exploitation d'archives sans développement propre (« turnkey solution »). L'application est construite en Open Source et mobilise des technologies objets. Le téléchargement de la suite Dspace est gratuite ce qui a assuré une large diffusion de cette outil. En France, l'INSEE, l'INIST-CNRS et l'INRA (entre autres) ont développé des services sur DSpace. Un des intérêts de DSpace est que toutes les archives institutionnelles développées sous cette application sont interopérables.

Vivo (www.vivoweb.org/) est une application développée à l'origine (2003-2006) à *Cornell University* comme une base de données relationnelles centrée sur la description des compétences des chercheurs en sciences de la vie. L'application deviendra ensuite multidisciplinaire. Entre 2010 et 2012 un financement de 12,2 M\$ du NIH² permet à Vivo de franchir une nouvelle étape de son développement et devient une application ouverte multiinstitutionnelle permettant un maillage des réseaux de compétences. En 2013, Vivo passe sous la houlette de l'« incubateur » DuraSpace avec pour mission d'élargir la communauté ouverte de développeurs Vivo. Vivo est entièrement basée sur des technologies *Open Source* et sur les technologies du web sémantique (Vivo est associé à une ontologie qui lui est propre. Vivo sert à construire par moissonnage (des publications, des archives institutionnelles, etc.) des bases de connaissances sur les domaines d'expertise des chercheurs, des labos, des institutions. Vivo est utilisé par une centaine d'institutions de par le monde. L'un des développements actuels les plus intéressants de Vivo est son expérimentation large dans les milieux de la recherche agronomique. L'USDA (US Department of Agriculture) s'appuie sur Vivo pour construire des bases de connaissances sur les 5000 chercheurs qui travaillent dans les différentes agences de recherche de l'USDA³. Le site dédié www.vivo.usda.gov résume bien l'intérêt de ce type d'outil : « *As research becomes more interdisciplinary, it can be hard to find collaborators outside your own area. USDA VIVO provides a powerful Web search tool for connecting researchers, research projects and outcomes, and others with relationships to the research. The idea is to link researchers with peers and potential collaborators. VIVO makes it possible to quickly identify scientific expertise to address a variety of food- and agriculture-related issues or to rapidly mobilize a response on a scientific issue* ». Un projet international AgriVIVO (www.agrivivo.net), soutenue par la FAO⁴, vise dans une logique collaborative à construire une base de connaissance internationale sur les chercheurs en agronomie. En France l'INRA utilise Vivo dans une application web sémantique de Linked Open Publication⁵, dont l'un des objectifs est d'enrichir les bases de connaissance sur les chercheurs de l'institut.

L'intérêt des bases de connaissances sur les chercheurs fondées sur des technologies Open Source est de rendre à terme interopérables des répertoires relevant de différentes organisations de recherche mais travaillant sur des problématiques voisines ou complémentaires.

Les fournisseurs commerciaux de « *Research Information Management Systems* » (RIMS, voir note EPRIST N°8⁶), à savoir Elsevier, Thomson Reuters et Symplectic suivent de près le développement de Vivo (Thomson et Symplectic sont d'ailleurs membres de la communauté Vivo), dans la mesure où ces développements collaboratifs ouverts et basés sur des briques logicielles libre d'accès et gratuites peuvent concurrencer certaines de leurs offres. L'application la plus aboutie de construction de bases de connaissances sur les chercheurs est l'application Elsevier SciVal Experts, dont les objectifs sont assez similaires à ceux de Vivo, à cette différence près que SciVal Experts, à cette différence près que SciVal Experts est payant - et cher.

On bute ici sur un constat : les outils de productivité proposés par les éditeurs pour gérer les informations sur la recherche sont des outils « propriétaires », payants, non interopérables, basés sur des logiciels non « ouverts » qui les rendent incompatibles avec une logique d'« Open Process » : ne peuvent communiquer entre elles que les institutions qui ont opté pour un même outil propriétaire, par exemple SciVal Experts. Or la science est de plus en plus collaborative, avec des réseaux de coopération multidisciplinaires très évolutifs. Permettre grâce à des technologies

²National Institutes of Health, (www.nih.gov) principale agence de financement de la recherche biomédicale aux Etats-Unis

³USDA's Agricultural Research Service (ARS), Economic Research Service (ERS), National Institute of Food and Agriculture (NIFA), National Agricultural Statistics Service (NASS), and U.S. Forest Service (USFS)

⁴Food and Agriculture Organization, agence de l'ONU en charge des questions agraires

⁵<http://docslide.fr/software/realiser-une-application-web-semantique-grace-a-loutil-vivo-cas-pratique-de-lop-inra.html>

⁶http://www.eprist.fr/wp-content/uploads/2015/12/I-IST_8_RIMS.pdf

« Open Source »⁷ de décrire finement des « cartes de compétences » des chercheurs, des labos, des organismes de recherche, et les rendre interopérables (donc susceptibles d'être exploitées dans une approche multiinstitutionnelle) en un enjeu important pour l'eScience. Si la plupart des développements « Open Source » liés aux « Research Information Management Systems » sont comme Vivo loin d'avoir acquis une pleine maturité, cette voie de l'Open Source doit être cependant explorée avec tenacité par le monde de la recherche publique.

L'expérimentation Vivo à l'INRA

En France, l'INRA www.inra.fr a utilisé VIVO dans le cadre d'un projet expérimental LOP (Linked Open Publications). L'objectif principal du projet pour la délégation à l'IST était de monter en compétences sur les technologies du web sémantique et d'explorer les fonctionnalités de VIVO. Il s'agissait d'intégrer plusieurs sources de données INRA sur la recherche (Prodnra, l'archive institutionnelle de l'Inra, les référentiels de personnes, structures, et activités), d'en publier le résultat en "Linked data" (ensemble de bonnes pratiques pour publier et connecter des données structurées sur le Web), et fournir une interface Web dotée de fonctions de recherche, de navigation et d'accès à ces données. Ces fonctions devaient permettre de répondre à des questions telles que « Combien de personnes à l'Inra travaillent sur le blé? Qui sont-elles? Quelles sont leurs spécialités/expertises? Sur quels projets travaillent-ils aujourd'hui? Qu'est-ce qu'elles publient? Comment les contacter? ».

Le projet a conduit à l'adaptation de l'ontologie de VIVO pour répondre aux besoins identifiés et au déploiement en intranet d'une application VIVO contenant les données des différentes sources précitées, liées entre elles et transformées en RDF (Resource Description Framework) conformément à l'ontologie de VIVO adaptée.

L'équipe projet a tiré de cette expérimentation trois principales leçons sur VIVO :

- ▶ VIVO est une application pour lier, partager et découvrir les données et répond correctement à des questions de type « qui fait quoi, dans quelle structure » conformément à son modèle de données de base ;
- ▶ L'extension du modèle, pour répondre à des questions plus complexes est risqué surtout car elle met évidence une exigence très grande dans la qualité des données que souvent nous n'avons pas.
- ▶ L'interface utilisateur de VIVO propose par défaut des fonctions de recherche, de navigation et de présentation des données. Si ces fonctions de base s'avèrent trop éloignées des fonctionnalités attendues, il peut valoir le coup de considérer une autre approche telle que utiliser VIVO comme base de connaissances et développer une nouvelle interface utilisateur pour la présentation des données.

⁷http://duraspace.org/why_open_source