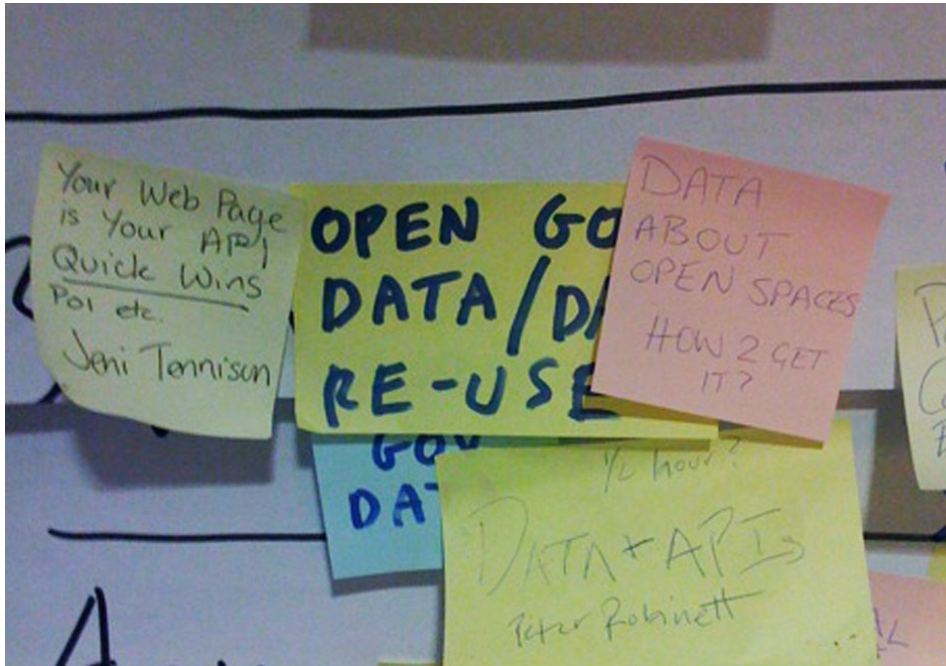


DATALIFT: UN CATALYSEUR POUR LE WEB DE DONNÉES

LE 7 JUILLET 2010 NICOLAS CYNOBER

Interview de François Scharffe, qui participe au projet DataLift, dont le but est de développer une plateforme pour publier et interconnecter des jeux de données sur le web de données.



Malgré **mon optimisme** naturel, **la polémique monte** sur la capacité de l'APIE à développer l'accès aux données gouvernementales en tant que service public. Robin Berjon a récemment **mis en garde** contre un modèle cherchant à en monétiser l'accès. Cela rajouterait en effet **une barrière considérable à l'innovation** nécessaire dans l'exploitation de ces données. Dans ce contexte, il est intéressant de noter le développement de projets parallèles sur le territoire français. Je pense au déjà connu **Regards Citoyens**, à **Data Publica**, dont je parlerai bientôt plus en détail, et également au tout jeune projet **DataLift**, dont **François Scharffe** a accepté de nous parler.

Nicolas Cynober : **Salut François, est-ce que tu peux nous présenter DataLift en quelques mots ainsi que ton rôle dans le projet ?**

François Scharffe : DataLift est un projet visant à créer une plateforme permettant à des éditeurs de données de publier leurs données sur le web de données. Le projet est découpé en tâches correspondant à des problèmes à résoudre le long de la chaîne de publication. Il faut tout d'abord sélectionner les ontologies qui permettront de décrire les données brutes. Ensuite il faut convertir les données dans le format du web sémantique, RDF. Puis il faut publier ces données selon les principes du web de données, ce qui suppose en parallèle d'interconnecter ces données avec d'autres jeux de données existants.

La plateforme d'outils qui sera construite au sein du projet a un caractère expérimental. C'est-à-dire que d'une part il y a dans DataLift une composante de recherche : nous allons développer de nouvelles techniques permettant d'automatiser le processus de publication. D'autre part nous allons expérimenter ces techniques pour effectivement publier des jeux de données. Le projet vient d'être retenu par l'**ANR** au sein de l'appel CONTINT et commencera en octobre. Je suis le coordinateur scientifique du projet.

Que signifie "interconnecter les données", à quoi cela sert-il ?

Lorsque l'on publie un jeu de données, un principe du web de données est qu'il faut interconnecter ce jeu de données à d'autres jeux existant. Cela revient à identifier les ressources déjà publiées sur le web de données qui correspondent à des ressources dans le jeu de données à publier. Par exemple si l'IGN publie un jeu de données sur les localités en France, il sera utile d'indiquer les équivalences entre ces localités et les ressources leur correspondant dans **DBPedia**.



Interconnecter les jeux de données est une étape cruciale.



C'est une tâche ardue, mais sans laquelle il n'y aurait pas de web de données. Ce serait un peu comme avoir le web sans liens entre les pages.

Peux tu nous parler un peu plus de vos fournisseurs de données, qu'attends-tu de l'APIE ?

Nous commençons le projet avec deux importants partenaires fournisseurs de données : l'IGN et l'INSEE. Ce seront donc les premiers à bénéficier des services de la plateforme. À côté de cela nous avons contacté un certain nombre d'institutions, associations et entreprises qui souhaitent rejoindre le projet et expérimenter la plateforme DataLift. Parmi eux je peux citer l'association Regards citoyens, le projet Data Publica, la DILA, la Fédération des parcs régionaux de France. La plateforme est ouverte et j'invite les lecteurs à **me contacter** s'ils sont intéressés pour nous rejoindre.

Nous sommes aussi en contact avec l'APIE qui souhaite développer un portail de données gouvernementales. L'APIE effectue un travail de sensibilisation auprès des ministères pour obtenir l'ouverture de leurs données. Les données que l'APIE pourra rassembler pourront être enrichies par la plateforme DataLift. **La discussion reste aujourd'hui ouverte** sur la manière dont l'APIE construira son portail.

Comment va se passer l'interconnexion de données ayant des licences très hétérogènes ?

C'est une très bonne question. Au sein de DataLift, l'équipe Edelweiss de l'INRIA va étendre les langages de représentation et les mécanismes d'interrogation des données afin de prendre en compte les licences et les informations de provenance attachées aux données. Des techniques vont donc être développées afin que rien n'empêche un fournisseur de données ayant un modèle de revenus basé sur une consultation payante de publier ses données, mais aussi d'attacher des informations de qualité et de confiance aux données publiées afin que les consommateurs de données s'y retrouvent.

DataLift est un projet de recherche, des points communs avec le projet du Tetherless World ?

Oui plusieurs. Tout d'abord nous sommes issus de la même communauté de recherche autour du web sémantique et croisons régulièrement **Jim Hendler** lors de la conférence internationale **ISWC**. Cela dit, les Américains ont la chance de pouvoir s'appuyer sur un catalogue de données gouvernementales déjà constitué. Nous devons directement contacter les institutions en attendant l'émergence d'un éventuel données.gouv.fr. Je pense que nous allons aussi plus loin en proposant une plateforme permettant à chaque fournisseur de données de publier ses données sur un serveur interne de façon décentralisée. Nous allons travailler afin que la plateforme devienne une référence mondiale en matière d'outils de publication de données.

À ton avis, quand pourrons-nous voir les premiers mashup basés sur les données de DataLift ?

Les premiers jeux de données publiés sont prévus à six mois et la première version de la plateforme est prévue à un an du début du projet. À côté du développement technique sur la plateforme, nous sommes contents d'avoir à bord la FING avec laquelle nous allons constituer une communauté de développeurs autour des données de la plateforme. Plusieurs événements sont prévus mais je préfère garder le suspense de ce côté-là. À côté de cela, la société Atos Origin va travailler à une interface de programmation web de données pour smartphones. Tout cela devrait permettre l'émergence d'applications innovantes autour de la plateforme. Nous encourageons les propositions de projets et communiquerons bientôt sur le sujet.

Le projet a été financé pour trois ans par l'Agence Nationale de la Recherche, peux-tu nous parler un peu de vos ressources et de ton équipe ?

Le projet est actuellement constitué de sept partenaires : les équipes EXMO et Edelweiss de

l'INRIA, Eurecom, la société Mondeca, Atos Origin Integration, l'IGN, L'INSEE, et la FING. Un huitième partenaire, l'équipe Tadoo du LIRMM va nous rejoindre sous peu. La plateforme est ouverte et nous accueillerons de nouveaux partenaires notamment fournisseurs de données. Le projet est assez gros pour un projet ANR et son budget total dépasse les trois millions d'Euros. C'est beaucoup d'argent et nous remercions l'ANR de nous faire confiance et de porter un projet sur ce thème. Nous devons maintenant montrer que cet argent sera utilisé à bon escient. Les subventions accordées vont notamment nous permettre de recruter des étudiants, chercheurs et ingénieurs. Avis aux amateurs, des positions sont ouvertes chez les divers partenaires.

L'équipe EXMO est située à l'INRIA Grenoble Rhône-Alpes, elle est dirigée par Jérôme Euzenat. Nous travaillons sur le web sémantique et en particulier sur l'alignement d'ontologies. Quand à moi je vais m'en aller vers Montpellier à partir de la rentrée comme maître de conférence au LIRMM. Je resterai tout de même rattaché à EXMO, on ne quitte pas une si bonne équipe comme ça !

Merci François, bonne continuation à toi et ton équipe !

—

Billet initialement publié sur **all about web**, le blog de Nicolas Cynober ; le Twitter de Nicolas : <http://twitter.com/cyno/>

À consulter aussi, **son pearltree** sur la publication de données en France

Image CC Flickr **illustir**


AGENCE WEB GRENOBLE

le 9 juillet 2010 - 16:28 • ; SIGNALER UN ABUS - PERMALINK



Ce sujet m'a fortement intéressé surtout le travail sur l'interface de programmation web de données pour smart-phone. A bientôt

VOUS AIMEZ  0

VOUS N'AIMEZ PAS  0


LUI RÉPONDRE


JOSEPHINEABBOTT

le 6 septembre 2010 - 8:59 • ; SIGNALER UN ABUS - PERMALINK



The loans are very useful for guys, which would like to ground their business. By the way, it is very easy to get a bank loan.

VOUS AIMEZ  0

VOUS N'AIMEZ PAS  0


LUI RÉPONDRE


ARAÇ BORÇ SORGULAMA

le 3 mai 2011 - 13:07 • ; SIGNALER UN ABUS - PERMALINK



thank you for informations

VOUS AIMEZ  0

VOUS N'AIMEZ PAS  0


LUI RÉPONDRE


COMPUTER REPAIR MIAMI

le 11 novembre 2011 - 1:18 • ; SIGNALER UN ABUS - PERMALINK



After studying a few of the weblog posts on your web site now, and I really like your approach of blogging. I bookmarked it to my bookmark website record and can be checking back soon. Pls take a look at my site as nicely and let me know what you think.

VOUS AIMEZ  0

VOUS N'AIMEZ PAS  0

LUI RÉPONDRE

2 pings

Les tweets qui mentionnent DataLift: un catalyseur pour le web de données »
Article » OWNI, Digital Journalism -- Topsy.com le 7 juillet 2010 - 18:38

*[...] Ce billet était mentionné sur Twitter par Mehdi Lamloum. Mehdi Lamloum a dit:
[#owni] DataLift: un catalyseur pour le web de données <http://goo.gl/fb/GeJPP> (via
@owni) [...]*

L'opendata dans tous ses états – Juillet II « le 12 juillet 2010 - 15:16

[...] DataLift: un catalyseur pour le web de données [...]