



HAL
open science

E-Science et professionnels de l'IST

Émilie Manon

► **To cite this version:**

Émilie Manon. E-Science et professionnels de l'IST. Sciences de l'information et de la communication. 2010. dumas-00511210

HAL Id: dumas-00511210

<https://dumas.ccsd.cnrs.fr/dumas-00511210>

Submitted on 24 Aug 2010

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.

Master 2 Professionnel 2009-2010

Spécialité Art, Culture et Médiations techniques

Option « Documentation scientifique et technique »



Mémoire de stage :
« E-Science et professionnels de l'IST »

Emilie MANON

UMS MI²S

sous la direction de :

Joanna JANIK (Responsable de stage)

Emilie MONDOLONI (Responsable pédagogique)

Février – Juillet 2010



Je tiens à remercier ma tutrice de stage, Joanna Janik, ainsi que Nicolas Balacheff, directeur, de m'avoir donné l'opportunité de travailler sur un sujet plus que novateur et de mener un projet dans son intégralité. Je remercie également Claudine Alberti, Marie-Anne Delègue, Serge Rouveyrol et Jérôme Zeiliger pour leurs conseils et leur aide. Un grand merci enfin à l'ensemble du personnel MIPS pour leur accueil et leur convivialité qui ont fait de ce stage une véritable expérience tant professionnelle qu'humaine.

Sommaire

Introduction	3
I. Présentation du centre de documentation	5
a. L'unité mixte de service MPS	5
b. Le service Information scientifique	6
1. <i>La bibliothèque de recherche</i>	7
2. <i>Les ressources électroniques</i>	7
3. <i>Les réseaux professionnels</i>	8
c. Mission du stage	9
II. E-science : définitions, enjeux et difficultés.....	13
a. Qu'est-ce que la e-science ?.....	13
1. <i>Naissance de la e-science</i>	13
2. <i>E-science, e-research, cyberinfrastructure, e-infrastructures</i>	17
b. De nouveaux environnements de recherche.....	19
1. <i>E-Science et Données de recherche</i>	20
2. <i>E-Science et Web collaboratif</i>	22
3. <i>E-Science et Apprentissage</i>	25
c. France et e-science : Informer et impliquer	27
1. <i>Position actuelle de la France dans la e-science</i>	27
2. <i>De la prise de conscience à l'implication : informer et conceptualiser la e-science</i>	30
III. E-science et professionnels de l'IST	34
a. Nouveaux enjeux	34
1. <i>Du document aux données de la recherche</i>	35
2. <i>Du prestataire au partenaire de la recherche</i>	36
b. Les domaines d'intervention	38
1. <i>La gestion des données</i>	38
2. <i>Les communautés virtuelles</i>	40
3. <i>Promotion et formation</i>	41
c. Vers le « e-science librarianship »	43
1. <i>Quelle formation ?</i>	43
2. <i>E-Science librarianship : un nouveau métier ?</i>	44
Conclusion	46
Glossaire	47
Ressources	49
Annexes.....	I
I. Le site e-science	I
II. Mind-map.....	III

Introduction

Le domaine de la documentation, et plus particulièrement de la documentation scientifique, est confronté ces dernières années à certains bouleversements voire mutations liés au développement d'Internet. L'essor des bibliothèques numériques et des archives ouvertes a redéfini le champ d'action des professionnels de l'information scientifique et technique (IST) désormais amenés à gérer des espaces virtuels de partage et de diffusion de la production scientifique. Cette redéfinition du métier va de pair avec l'évolution toujours plus rapide des méthodes de recherche scientifique. Au cours des dernières décennies, le développement des technologies de l'information et de la communication et la nécessité constante de visibilité ont considérablement encouragé les chercheurs à investir les nouveaux outils offerts par le web (blogs, wikis, réseaux sociaux, archives ouvertes ...). Parallèlement, les nouveaux équipements scientifiques permettent des expériences inédites qui entraînent de fait un véritable déluge de données qu'il s'agit maintenant de partager et d'exploiter. Depuis le début des années 2000, on a pu voir émerger dans la littérature scientifique un nouveau concept : celui de « e-science ». Ce terme désigne les infrastructures permettant le partage et l'exploitation *in silico* des données de la recherche ainsi qu'une communication quasi immédiate entre les différents acteurs du monde scientifique via des communautés virtuelles permises par les nouveaux outils d'internet. Ces évolutions bousculent considérablement les contours du champ d'action des documentalistes scientifiques. À la publication scientifique traditionnelle succède un support hétérogène où se rencontrent textes, hypertextes et données. Ces données ont besoin d'être correctement annotées et gérées afin d'être réutilisables et convenablement mises à disposition dans des environnements numériques permettant le travail collaboratif. En tant que médiateur de l'information scientifique, quelle place peut tenir le documentaliste dans ce nouvel environnement ?

C'est dans ce nouveau contexte et ces nouveaux enjeux de l'IST que j'ai choisi, dans le cadre de cette deuxième année de Master professionnel Art, Culture et Médiations techniques, spécialité « Documentation scientifique et technique », d'effectuer mon stage de fin d'études dans le centre de documentation scientifique de l'unité mixte de service Mi²S. Seconde expérience dans une bibliothèque de recherche, il s'est agi pour moi de parfaire et de compléter mes compétences et d'affirmer un peu plus mon projet professionnel. Après une brève présentation du centre de documentation et de ma mission, nous donnerons une définition du concept d'e-science et étudierons les enjeux et difficultés inhérents à cette notion plutôt vaste. Enfin, nous réfléchirons aux implications que ces nouvelles méthodes de travail induisent pour les professionnels de l'information scientifique et donnerons les préconisations nécessaires pour que ces métiers abordent plus sagement ces évolutions.

I. Présentation du centre de documentation

Ce stage de fin d'études a débuté le 1^{er} février 2010 pour une période de six mois au sein de l'unité mixte de service MI²S (UMS 3042). Dans cette partie, nous présentons l'UMS ainsi que le service d'information scientifique auquel j'ai été intégrée. Enfin, nous détaillerons la mission qui m'a été proposée.

a. L'unité mixte de service MI²S

L'unité *Moyens Informatiques Multimédia et Information Scientifique* (MI²S)¹ est une unité mixte de service (UMS 3042) sous tutelle UJF-Grenoble I, Grenoble-INP et CNRS. Basée sur le campus grenoblois, à la Maison Jean Kuntzmann, cette unité connaît une histoire récente. Sa création en 2007 fait suite à la reconstruction de l'ensemble des laboratoires du secteur Mathématiques et Sciences et Technologies de l'Information et de la Communication (MSTIC) sur le site grenoblois. S'appuyant sur le fort héritage de l'ancienne infrastructure d'Informatique et de Mathématiques Appliquées de Grenoble (IMAG), l'UMS assure l'offre de services initialement dédiés à cette institution. Elle dessert donc neuf laboratoires :

- Circuits Multi-Projets (CMP, UMS 3040)
- Laboratoire Grenoblois de l'Image, de la Parole, du Signal et de l'Automatique (Laboratoire GIPSA-Lab, UMR 5216)
- Informatique et Création Artistique (ICA, EA 2934)
- Laboratoire des Sciences pour la Conception, l'Optimisation et la Production de Grenoble (Laboratoire G-SCOP, UMR 5272)
- Laboratoire d'Informatique de Grenoble (LIG, UMR 5217)
- Laboratoire Jean Kuntzmann (LJK, UMR 5224)
- Techniques de l'Informatique et de la Microélectronique pour l'Architecture des Systèmes Intégrés (TIMA, UMR 5159)
- Techniques de l'Ingénierie Médicale et de la Complexité – Informatique, Mathématiques et applications de Grenoble (TIMC-IMAG, UMR 5525)
- Laboratoire Verimag (Verimag, UMR 5104)

¹ <http://mi2s.imag.fr/index.html>

Ces laboratoires couvrent un prisme particulièrement varié de problématiques de recherche s'appuyant sur le recours aux MSTIC : par exemple, la compréhension et le contrôle des processus normaux et pathologiques en biologie et santé, les sciences des calculs, l'informatique pour l'éducation, les systèmes intégrés sur silicium etc. L'UMS a donc pour mission d'offrir des services transversaux à l'ensemble de ces laboratoires. Cette unité se scinde en deux services qui sont régulièrement amenés à travailler ensemble :

- le service Moyens Informatiques et Multimédia (MIM)² qui assure les services informatiques de l'UMS : en matière d'infrastructure réseau et services associés, sécurité, sauvegarde, nomadisme, services de communication (messagerie, webmail, listes de diffusion, ...) et multimédia (visioconférence) ;
- le service Information Scientifique (IS)³. C'est dans ce dernier service, que nous présentons ci-après, que j'ai pu effectuer mon stage.

Depuis 2008, la direction de l'UMS est assurée par Nicolas Balacheff, directeur de recherche au CNRS.

b. Le service Information scientifique

Le service Information Scientifique auquel j'ai été intégrée a pour mission la mise à disposition de l'information scientifique relative au secteur de l'Informatique et des Mathématiques Appliquées. Il est composé de quatre personnes ainsi qu'un informaticien spécialisé dans l'ingénierie documentaire. Joanna Janik, ingénieur de recherche, assure la responsabilité du service. Marie-Anne Delègue gère les périodiques ainsi que l'ensemble des ressources électroniques. Elle s'occupe également de l'interface web. Claudine Alberti a en charge la gestion des monographies et des thèses (papier et microfiches). Andrée Gheraibia (partie en retraite au cours de mon stage) assure l'accueil des usagers (prêts et retours) ainsi que le prêt entre bibliothèques. Enfin, Serge Rouveyrol a en charge l'informatique de la bibliothèque. Il a également mis en place un analyseur de proxy afin d'établir des statistiques quant à l'utilisation des ressources électroniques.

Ainsi, le service dessert en priorité les membres des neuf laboratoires du réseau (tout statut confondu : chercheur, doctorant, stagiaire, invité) mais s'adresse également aux étudiants grenoblois de l'ENSIMAG (Grenoble-INP), de l'UFR IMAG (UJF) ainsi qu'à

² <http://mi2s.imag.fr/mim.html>

³ <http://mi2s.imag.fr/is/>

l'ensemble des doctorants en « Mathématiques-Informatique et Ingénierie pour la Santé, la Cognition, l'Environnement » de Grenoble Universités.

1. La bibliothèque de recherche

Pour répondre aux différents besoins, la bibliothèque de recherche du service dispose d'un fonds de plus de 40 000 monographies et thèses (dont 18 000 thèses sur microfiches), 339 titres de journaux imprimés dont 77 abonnements en cours en 2009 et 151 fermés mais prolongés sous format électronique. La bibliothèque dispose également d'un fonds spécialisé en histoire de l'informatique et des mathématiques appliquées (fonds HIMA) ainsi qu'un fonds spécifique adapté aux besoins de l'enseignement et prioritairement réservé aux étudiants de Licence, Master et Doctorat (fonds LMD). La politique documentaire du service se fait également en étroite collaboration avec les laboratoires. Pour ce faire, chaque laboratoire a un représentant CORIST (anciennement « Correspondants Doc ») chargé de relayer les informations du service aux autres membres de son laboratoire, mais également, de faire remonter les éventuels besoins ou problèmes de leurs collègues en matière d'IST. De même, plusieurs membres des laboratoires désignés « Correspondants Acquisitions » se réunissent avec le service IS en une commission « Bibliothèque » chargée d'analyser et de décider des acquisitions à réaliser.

Le catalogue de la bibliothèque, mis aux normes internationales (UNIMARC-SUDOC) en 2008, est intégré au catalogue Rugbis du Service Interétablissements de Coopération Documentaire 1 (SICD 1) réunissant bibliothèques de l'Université Joseph Fourier et de l'Institut Polytechnique de Grenoble et au SUDOC. Pour l'année scolaire 2008-2009, environ 450 lecteurs étaient inscrits à la bibliothèque et 513 emprunts ont été effectués.

2. Les ressources électroniques

Outre ces services classiques propres à une bibliothèque (prêt, acquisition, abonnements, prêt entre bibliothèques etc.), le service porte un intérêt particulier à la documentation numérique. L'offre en la matière est particulièrement riche et variée. Les membres des laboratoires ont accès au portail de ressources BiblioST2i⁴ du CNRS-INIST qui propose un vaste choix de revues électroniques (plus de 4500). L'ensemble des usagers a également accès au service G@el⁵ (Grenoble Accès en Ligne) du SICD 1 (plus de 5300

⁴ Portail de ressources mutualisées pour la communauté des Sciences et Technologies de l'Information : <http://bibliost2i.inist.fr/>

⁵ http://sicd1.ujf-grenoble.fr/Ressources_Electroniques/Periodiques_Electroniques_Gael/PE_gael.php

revues) et aux ressources du Réseau National des Bibliothèques de Mathématiques (RNBM)⁶ qui propose quelques 200 revues. Enfin, Mi²S a également acquis 83 abonnements en nom propre. L'ensemble de ces ressources est accessible soit par reconnaissance IP soit par login et mot de passe. Pour plus de simplicité, les usagers ont accès aux différents titres via l'interface AtoZ (EBSCO) qui permet de présenter une liste abécédaire de l'ensemble des ressources électroniques disponibles.

Le service propose également des formations aux différents outils documentaires qu'elle met à la disposition des étudiants et des chercheurs : découverte de la bibliothèque et de ses ressources, rédaction bibliographique et gestion d'outils bibliographiques, outils de veille, formation aux archives ouvertes comme Hal (Hyper Articles en Ligne) ou Tel (Thèses en Ligne)⁷. Les usagers ont également accès via le site aux signets de la bibliothèque (sur le site de partage de signets Del.ici.ous⁸) régulièrement mis à jour et recensant diverses ressources internet utiles pour la recherche d'information et la veille.

Enfin, dans le cadre de ses missions, l'UMS a entrepris en 2009 le développement d'un outil de gestion de la production scientifique. Un prototype, l'outil HALOA, a été créé par le laboratoire Verimag. Répondant aux standards du protocole OAI-PMH, il permet la gestion bibliographique des documents numériques déposés sur l'archive institutionnelle Hal et de générer des listes dynamiques de publications pouvant être facilement intégrées dans des pages web, des rapports d'activité ou tout autre forme de support.

3. Les réseaux professionnels

Par ailleurs, l'UMS est particulièrement investie dans des réseaux professionnels. En 2007, elle s'est vu confier l'animation et la coordination du réseau Doc-ST2i⁹. Ce réseau professionnel regroupe documentalistes en Sciences et Technologies de l'Information et de la Communication (STIC) et Sciences pour l'Ingénieur (SPI) du CNRS. Il s'agit de créer un réseau d'échange entre les professionnels de l'IST du domaine et de relayer les questions, besoins et difficultés des chercheurs en matière d'accès à l'information auprès des institutions. Le réseau Doc-ST2i a également pour mission la mutualisation des outils – le portail BiblioST2i en est la parfaite illustration – et la formation des publics aux différentes méthodes documentaires.

⁶ <http://www.rnbn.org/>

⁷ Respectivement <http://hal.archives-ouvertes.fr/> et <http://tel.archives-ouvertes.fr/>

⁸ http://delicious.com/mi2s_favoris

⁹ <http://doc-st2i.cnrs.fr/>

De même, les membres du service IS sont impliqués dans d'autres réseaux : au niveau régional par le biais de multiples actions au sein du réseau ISIDORA¹⁰ réunissant les documentalistes CNRS de Rhône-Alpes ; et au niveau national au sein du réseau RENATIS¹¹.

Enfin, le service organise chaque année une journée d'étude à destination à la fois de la communauté scientifique et des professionnels de l'information scientifique et technique. En mars 2010, une journée a donc été organisée autour de la question du droit d'auteur et de son évolution dans l'ère numérique.

c. Mission du stage

Nous l'avons vu, l'UMS MI²S s'est particulièrement investie dans une conception moderne de la documentation scientifique en proposant des services et outils numériques fondés sur le partage et la collaboration.

Ainsi, au cours de l'année 2009, en cohérence avec ses missions, l'UMS MI²S s'est engagée dans le cadre du volet scientifique du projet PILSI (Pôle d'Innovation Logiciel et Systèmes Intelligents) émanant du plan Campus, en proposant une opération tournée vers les nouveaux outils pour la recherche (projet e-science). PILSI est un pôle de valorisation de l'innovation et de soutien au partenariat entre monde académique et industriel. Il s'agit de rassembler les moyens humains et technologiques grenoblois et extérieurs, issus des mondes académiques et industriels, dans le domaine des logiciels et systèmes intelligents au cœur d'un même pôle d'excellence. Au sein de ce pôle, le projet e-science a pour objectif d'apporter des outils collaboratifs adaptés à la communauté scientifique qui permettront, entre autres, une meilleure coopération scientifique et une meilleure visibilité du site grenoblois en matière de logiciels et de systèmes intelligents et feront de PILSI un pôle d'attractivité tant national qu'international pour les jeunes chercheurs et étudiants du domaine. Au sein de ce projet e-science, plusieurs axes ont été définis dont la mise en place d'une plateforme e-science fédérant outils et services collaboratifs sur le net et intégrant des outils de partage de ressources dont des services de publication, d'exploitation, de communication et de réseaux sociaux.

Certes, de tels outils existent déjà sur internet et sont proposés gratuitement. Citons les sites de partage de ressources tels Zotero, Mendeley, Slideshare ou encore Youtube ou Flickr. Des réseaux sociaux scientifiques ont également fait leur apparition ces dernières années tels

¹⁰ <http://isidora.cnrs.fr/>

¹¹ <http://renatis.cnrs.fr/>

ResearchGate ou SciLink. Google est également l'une des sources privilégiées des chercheurs en informatique ou mathématiques en matière de partage (GoogleAgenda, GoogleDoc etc.). Cependant, il est urgent de s'interroger sur la fiabilité et la pérennité de tels systèmes. En effet, en avril 2008, une note du Fonctionnaire de Sécurité Défense du CNRS¹² mettait l'accent sur la vulnérabilité de ces services dont on ignore tout de leur gestion des données. Il s'agit donc de créer des outils dont on a le contrôle. Pour autant, il ne s'agit pas d'apposer des formats propriétaires mais d'étudier les services déjà existants (comme les solutions open source) ou de développer des outils adaptés aux besoins de la communauté. Le projet e-science doit donc répondre à différents défis tant technologiques (questions de sécurité, pérennité, puissance et capacité de stockage) qu'éthiques (modalités de partage et protection de la propriété intellectuelle).

Aussi, forte de son expérience dans les outils collaboratifs – comme les forges menées par différents laboratoires de la communauté (LJK, LIG, etc.), les archives ouvertes, les bases de données bibliographiques ... – l'UMS pourra être maître d'ouvrage pour la mise en place d'une telle infrastructure répondant à l'ensemble de ces défis.

Ma mission de stage consistait donc en une étude de la problématique e-science¹³ afin d'en dresser un état de l'art recouvrant aussi bien les multiples définitions de ce concept, que les projets labellisés « e-science » et les différentes implications à prévoir pour les professionnels de l'information scientifique et technique.

La France ne prend pas encore toute la mesure du mouvement d'ouverture des données contrairement à d'autres pays qui ont annoncé l'ouverture de leurs données publiques à l'instar de l'initiative lancée l'an passé par le président Obama avec le site data.gov¹⁴. Il existe au niveau international une révolution qui s'est mise en marche au tournant des années 2000 avec l'apparition de nouveaux outils de communication, de publication, de partage et d'accès à l'information. Il s'agit donc de comprendre dès aujourd'hui quels seront les environnements de recherche de demain que l'on nomme déjà e-science.

La première phase de mon stage a donc consisté à me documenter sur le sujet et à fournir une bibliographie conséquente à la communauté. Peu de ressources en français ont été trouvées, la notion même d'e-science, née en Angleterre, étant encore trop récente pour être entrée dans la littérature scientifique française. J'ai mis en place un partage de signets sur le

¹² <http://www.dgdr.cnrs.fr/fsd/>

¹³ Pour plus de clarté, nous définirons ce concept dans la seconde partie.

¹⁴ <http://www.data.gov/>

site Del.ici.ous¹⁵ référençant les différentes ressources utiles à la compréhension de la notion d'e-science. Cet outil m'a permis de regrouper ces ressources par catégories (disciplines, pays, type de ressource) et de les partager au-delà de la seule communauté grenobloise. Au total, ce sont plus de 70 ressources qui ont été référencées. De même, j'ai référencé et mis à disposition les ressources bibliographiques, une soixantaine, sur le site de gestion bibliographique Zotero¹⁶. Je continue au fil de mes travaux à enrichir ces deux plateformes. La communication et le partage n'ont eu de cesse d'être primordiaux dans cette étude dont le sujet est particulièrement complexe et novateur.

Au terme du premier mois de recherche et de collecte d'informations sur le sujet, j'ai pu mettre en place un site internet nommé « E-Science » à l'aide du CMS (Content Manager System) Wordpress© et présentant une synthèse de mes recherches. Cet outil, facile d'utilisation, a permis une diffusion plus large de mon étude dépassant la seule communauté des MSTIC. Ce site est un sous-site du site de l'UMS et a été intégré à la présentation du projet PILSI¹⁷. Il a été divisé en plusieurs rubriques (Définitions, Objectifs, Challenges, Projets, E-Science et professionnels de l'IST, Ressources) afin de faciliter la navigation des visiteurs et de respecter au mieux les différents intérêts des communautés concernées. En effet, au-delà de la communauté scientifique, c'est aussi celle des professionnels de l'IST qui est ici visée. C'est pour cette raison que la rubrique « E-Science et professionnels de l'IST » a été créée afin de traiter des diverses implications que ces nouvelles infrastructures de recherche auront (ou ont déjà, comme nous le verrons plus tard) sur les métiers de la documentation scientifique. Régulièrement, paraissent des articles présentant un projet, une définition ou des pistes de réponses aux nouveaux challenges posés par la e-science. À ce jour, plus d'une quarantaine d'articles¹⁸ ont été publiés.

Nous avons fait paraître un éditorial annonçant l'ouverture du site en page d'accueil du site de l'UMS courant avril et incitant la communauté à participer à ce projet. En effet, au-delà de la diffusion d'information, il s'agit également de récolter des avis et d'identifier des besoins propres à la communauté desservie afin d'étudier, par la suite, la faisabilité d'une plateforme répondant à ces critères. Cette seconde étude sera l'objet de la poursuite de mon expérience au sein du service à la fin de ce stage. Nous avons d'ores et déjà mis en place un outil statistique d'analyse des visites afin de cerner le type de public et leurs centres d'intérêt. Un plugin Wordpress me permet donc en temps réel de savoir qui visite le site (adresse IP et

¹⁵ http://delicious.com/mi2s_Escience

¹⁶ http://www.zotero.org/emilie_manon/items

¹⁷ Voir Annexe I. Le site est également consultable à l'adresse : <http://mi2s.imag.fr/e-science/>

pays), comment il est arrivé sur notre site (mots-clés entrés sur le moteur de recherche) et les pages qu'il a visitées.

Le site, bien que parfaitement référencé, ne jouit pas encore de grande visibilité. Cependant, depuis son ouverture, le site a attiré quelques 200 personnes. Il est toutefois encore tôt pour tirer des conclusions de ces statistiques. De plus, afin de me renseigner sur certains projets français liés à la e-science, j'ai également pris contact avec l'Institut de l'Information Scientifique et Technique (INIST) du CNRS. Son directeur adjoint, M. Guichard, intéressé par notre étude, a donc souhaité collaborer avec le service. Cette collaboration est actuellement en cours.

Cette mission de stage a nécessité un travail conséquent et parfois ardu de définition et de conceptualisation de la notion d'e-science. Celle-ci regroupe une variété importante de conceptions notamment en raison de la nouveauté des notions utilisées, telles celles de cyber-infrastructures, de web des données, de grille informatique etc. (que nous définirons plus tard ici). La première catégorie créée sur le site a d'ailleurs été le glossaire, outil fondamental pour une bonne compréhension de ce domaine fort vaste et multidisciplinaire. Le site étant ouvert à tous sur internet, il a fallu vulgariser les concepts informatiques afin de toucher un large public. Le site E-Science a donc une visée avant tout pédagogique afin de retracer l'histoire de l'e-science, de son apparition à ses implications pour le monde actuel de la recherche.

La e-science soulève également de nombreux défis à relever. Comment convaincre les chercheurs de partager leurs données ? Comment assurer la pérennité et la sécurité de ces sources ? Comment protéger les droits des créateurs ? Quels standards adopter pour une collaboration internationale et pluridisciplinaire ? Ces questions sont donc abordées dans le cadre du site e-science qui propose des pistes de réponses aux nombreuses interrogations qu'il soulève. Ce sont ces définitions et résultats que nous proposons de détailler ci-après.

¹⁸ Voir un exemple d'article Annexe I.

II. E-science : définitions, enjeux et difficultés

Nos capacités de mesurer, stocker, analyser et visionner les données est de plus en plus la nouvelle réalité à laquelle la recherche devra s'adapter.¹⁹

Comme le remarque ici John Wilbanks, la recherche entre dans une nouvelle ère de partage et d'exploitation de données. Le développement des nouvelles technologies de l'information et de la communication (NTIC) a fait évoluer les pratiques des chercheurs et les conduisent désormais à investir les nouveaux outils offerts par le web (archives ouvertes, wikis, blogs, réseaux sociaux...) fondés sur une collaboration plus étroite et un partage plus immédiat des ressources. Parallèlement, les nouveaux instruments scientifiques permettent des expériences inédites qui entraînent un véritable déluge de données qu'il s'agit dès à présent de partager et d'exploiter. C'est dans ce nouveau contexte qu'est née la e-science²⁰.

a. Qu'est-ce que la e-science ?

Pour comprendre les différents enjeux de la e-science, il faut avant tout percevoir une image précise du contexte dans lequel celle-ci est apparue et saisir les nombreuses notions qu'elle recouvre.

1. Naissance de la e-science

Le temps des calculs à la main et des longues recherches dans les notices des bibliothèques est pour ainsi dire révolu. Avec l'informatique et Internet, l'échange de ressources, la communication et la mise en commun pour les chercheurs des résultats de leurs recherches sont amplement facilités. Mais justement dans cette nouvelle ère du « tout numérique », la masse d'informations circulant sur le net est « diluvienne » [en référence à l'expression devenue commune "*data deluge*"²¹]. On a généré plus de données scientifiques depuis 2005 qu'au cours de toute l'histoire de l'humanité !

¹⁹ Wilbanks, John. I Have Seen the Paradigm Shift, and It Is Us. *The Fourth Paradigm: Data Intensive Scientific Discovery*. Microsoft Research, 2009. p.209. [disponible sur :] <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>

²⁰ Ce terme n'ayant pas véritablement de traduction française, les auteurs emploient indifféremment le genre masculin ou féminin. On a pu trouver « la e-science », « l'e-science » ou parfois simplement « e-science » sans article. Ici, nous utilisons plus volontiers la forme féminine.

²¹ Hey, Tony, et Anne E. Trefethen. The data deluge: an e-science perspective. *Grid computing: making the global infrastructure a reality*, 2003, Wiley and Sons, pp.809–824. [disponible sur :] http://eprints.ecs.soton.ac.uk/7648/1/The_Data_Deluge.pdf

Pour prendre un exemple significatif, lorsque que le Large Hadron Collider du CERN sera mis en fonction, c'est plus de 15 millions de gigaoctets de données qui seront produites par an ! Les chercheurs sont donc particulièrement touchés par ce déluge informationnel. Ils passeraient même deux tiers de leur temps à trouver et analyser des données.

Depuis l'apparition des NTIC, les chercheurs veulent désormais aller plus loin que les capacités originelles du net. Ils veulent intégrer, fédérer, analyser des données de différentes sources. Mais, chaque communauté scientifique a ses habitudes et ses normes. Aussi, toutes ces données issues de la recherche sont souvent dans des formats différents, leur pérennité n'est pas garantie, les logiciels entre pays, communautés, laboratoires sont souvent incompatibles, la communication n'est pas instantanée (les temps de publication sont souvent très longs selon les disciplines). La nécessité de disposer d'outils spécifiques de visualisation, de fouille de données (*datamining*) et d'infrastructures informatiques qui soutiennent le tout est plus qu'urgente. Comment alors pallier à ce « déluge » de données hétérogènes et se servir des nouvelles technologies pour faire de la science autrement ?

Certes, depuis une dizaine d'années, le mouvement *open access* a permis la mise en commun et un accès instantané à nombre de ressources. Cependant, dans ce nouveau contexte, il faut aller au-delà de la seule consultation. Ce n'est plus seulement aux publications que l'accès doit être facilité mais bien à l'ensemble des données issues de la recherche. Parallèlement, le développement des outils dits web 2.0 a permis au monde scientifique d'investir de nouveaux canaux permettant une communication plus synchrone et interactive. Les chercheurs tendent de plus en plus à travailler en réseau, sur le net, à créer des communautés virtuelles où se partagent publications scientifiques, opinions (*peer-review*), et folksonomies²².

Ce nouveau mode de transfert de connaissances s'appuie sur un savoir collectif et communautaire. Il s'agit donc de créer de vastes répertoires avec les capacités managériales des bibliothèques numériques traditionnelles en y ajoutant une couche supérieure de service : outils spécifiques d'exploitation de ces données (chercher, déplacer, manipuler, personnaliser ...), d'indexation, de communication synchrone et asynchrone etc. Il faut donc créer un lien entre les diverses applications et services afin de créer un seul ordinateur virtuel pour l'utilisateur final. Cette vision d'un middleware²³ (entre le réseau et l'application) qui

²² La folksonomie est un néologisme désignant le phénomène croissant d'indexation par l'utilisateur de ressources numériques disponibles sur un service web. Pour plus d'informations, se reporter au Glossaire.

²³ Un *middleware* est une couche de logiciels permettant la communication entre plusieurs applications informatiques s'effectuant sur différentes machines. Il permet à ces applications d'échanger et d'interopérer sur un réseau commun.

permettrait de faire communiquer les applications entre elles est un élément fondamental dans la mise en place d'une infrastructure propre aux besoins spécifiques des communautés.

Cette nouvelle façon de « faire de la science » est nommée « e-science ». Ce terme a été imaginé au Royaume-Uni. C'est John Taylor, fin 2000, alors directeur général des conseils de recherche, qui fait le constat que progressivement la science devenait dépendante de coopérations multidisciplinaires nationales et internationales. Les différentes communautés scientifiques auraient beaucoup à gagner dans la mise en place d'une infrastructure commune qui leur permettrait l'accès à distance aux ressources et une manipulation des données en ligne dont elles ont besoin. The UK National e-Science program est alors lancé en novembre 2000. On y définit la e-science comme :

the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large-scale computing resources and high-performance visualization back to the individual user scientists.²⁴

Plutôt que e-electronic science, il s'agit en fait de e-nhanced science (augmentée, améliorée). Le recours à des outils informatiques puissants est incontournable dans la e-science. Aussi, le terme est étroitement lié au concept de grille informatique ou grille de calcul. Une grille informatique (en anglais, *computing grid*) est une infrastructure virtuelle constituée d'un ensemble de ressources informatiques hétérogènes et dispersées géographiquement. Elle permet un partage de ressources dans un environnement virtuel sécurisé accessible à une collection dynamique d'individus et d'institutions. Sur ces grilles, plusieurs ordinateurs alimentent un réseau et fournissent ainsi une grande puissance de calcul et une grande capacité de stockage. Tout utilisateur peut alors utiliser les ressources de la grille via une interface unique sur son propre ordinateur. La e-science est ainsi une science de gestion intensive de données :

e-science is used to describe computationally intensive science that is carried out in highly distributed network environments, or science that uses immense data sets that require grid computing; the term sometimes includes technologies that enable distributed collaboration.²⁵

Aux États-Unis, la National Science Foundation (NSF) fait le même constat. Selon un rapport de 2003, il est alors plus qu'urgent d'avoir recours à ces technologies afin de :

²⁴ <http://www.nesc.ac.uk/nesc/define.html>

²⁵ extrait de l'article « E-science » de l'encyclopédie Wikipedia, [disponible sur :] <http://en.wikipedia.org/wiki/E-Science>

build more ubiquitous, comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments and that operate at unprecedented levels of computational, storage, and data transfer capacity.²⁶

Cela a concerné tout d'abord le domaine des sciences exactes (surtout physique et biologie), domaine particulièrement demandeur en puissance de calcul et capacité de stockage afin de traiter efficacement de larges silos de données. Il s'agit d'ailleurs de la voie traditionnelle de la e-science, fondée sur d'importantes grilles informatiques. Cependant, la e-science a très vite recouvert d'autres champs et s'est répandu à toutes les disciplines.

En effet, s'il est vrai que le terme d'e-science est étroitement associé au concept de grille, il ne se réduit pas à cela. E-science est un terme fort vaste qui ne représente pas seulement la nouvelle puissance des outils informatiques sur lesquels s'appuie désormais la recherche, mais aussi le concept de science collaborative, en réseau et distribuée. Au-delà de la seule performance informatique, il s'agit de permettre un accès à distance aux données, ressources et plus généralement à tout matériel permettant des collaborations inédites. John Taylor, « père » du concept e-science, le définit selon cette vision collective :

e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.²⁷

Il ne s'agit donc pas à proprement parler d'un nouveau domaine scientifique mais plutôt de la représentation formelle de l'ensemble d'outils et de technologies qui permettent une science collaborative. La e-science favorise les communautés multidisciplinaires, les espaces de travail collaboratif, un peer-review plus immédiat et interactif. En effet, elle permet de faire tomber les frontières entre les disciplines en favorisant de nouvelles collaborations. Les Digital Humanities, par exemple, ont également compris les enjeux de la e-science et on a vu récemment se multiplier les projets e-science dans ce domaine. C'est donc l'idée d'une « big science » et d'équipes de recherche virtuelles permises par des infrastructures informatiques puissantes.

²⁶ Ce rapport est devenu une référence en matière de e-science. Dans la littérature spécialisée, on le nomme « rapport Atkins » Voir D.E. Atkins, K.K. Droegemeier, S.I. Feldman, H. Garcia-Molina, M.L. Klein, D.G. Messerschmitt, P. Messina, J. Ostriker, et M.H. Wright, *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, National Science Foundation, 2003, p.3

²⁷ Taylor, John. Présentation du programme e-science. [disponible sur :] <http://www.nesc.ac.uk/nesc/define.html>

2. *E-science, e-research, cyberinfrastructure, e-infrastructures ...*

Il existe un flou terminologique autour du concept d'e-science. En effet, selon les pays ou les disciplines, on peut trouver des termes comme « cyberinfrastructure », « e-research » ou encore « e-infrastructure ». E-Science est un terme plutôt européen, tandis que « cyberinfrastructure » se trouvera plus généralement dans la littérature américaine.

Le terme « cyberinfrastructure » (on trouve aussi le terme « e-infrastructure » en Europe) désigne de manière générale les nouveaux environnements de recherche permis par les nouvelles technologies de l'information et de la communication. Les termes e-science et cyberinfrastructure sont presque synonymes. Le premier est plus largement utilisé en Europe et se place d'un point de vue scientifique; le second est plus répandu aux USA et est plus centré sur la prouesse informatique. Les deux termes font référence à l'utilisation d'outils informatiques puissants et en réseau permettant de nouvelles collaborations scientifiques et des méthodes inédites de recherche.

C'est le rapport Atkins de la NSF qui a popularisé le terme « cyberinfrastructure » en 2003. Celui-ci :

refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy.²⁸

Il s'agit d'une infrastructure numérique commune fondée sur les notions de collaboration, de libre accès aux données brutes de la recherche et d'interopérabilité. Chaque communauté a ses usages et ses normes. Les données, ressources ou logiciels développés par la recherche sont souvent non-interopérables. Une utilisation efficace d'une cyberinfrastructure par la communauté scientifique permettrait donc de faire tomber les frontières artificielles entre les disciplines que des pratiques trop individualistes ont tendance à isoler.

Si e-science est la théorie, nous pouvons dire que la cyberinfrastructure est la pratique, l'infrastructure informatique permettant ces nouvelles méthodes de recherche. Pour mieux comprendre les différents services que recouvre une cyberinfrastructure, on peut se référer aux travaux du TGE-ADONIS²⁹ qui synthétise ici les éléments fournis par le rapport Atkins :

²⁸ Rapport Atkins, *op.cit.*, p.5.

²⁹ Chartron Ghislaine et Elisabeth Caillon, Veille internationale - Infrastructures numériques SHS, Présentation, TGE-Adonis Université d'été, 3-5 septembre 2008 [disponible sur :] http://www.tge-adonis.fr/IMG/pdf_diaposPresentationVeilleAdonis.pdf

Environnements de recherche				
Adaptation par discipline/applications spécifiques à des projets				
Services intégrés d'une cyberinfrastructure				
Services de calcul haute performance	Services de gestion de données	Services de production, de mesure, d'observation	Services de visualisation, interfaces	Services de collaboration
Réseau, «operating systems », middleware,				
Technologie de base: calcul, stockage, communication				

Figure 1. Tableau extrait de la présentation "Veille internationale - Infrastructures numériques SHS" du TGE-ADONIS

On observe qu'une cyberinfrastructure est composée de différentes couches. Sur la couche purement matérielle (*hardware*), vient s'ajouter la couche réseau qui permet aux différents composants informatiques de communiquer entre eux (*middleware*). Enfin, une dernière couche orientée utilisateur permet d'imbriquer les différents services d'exploitation de données et de communication.

Si le terme « cyberinfrastructure » était au début plutôt réservé aux domaines des sciences dures (chimie, biologie et physique), il s'est très rapidement répandu à l'ensemble du paysage scientifique. Aussi, l'American Council of Learned Societies Commission a publié en 2006 un rapport sur l'usage des cyberinfrastructures pour les sciences humaines. Ce rapport définit plus largement le terme :

Une strate faite d'information, d'expertise, de normes, de politiques, d'outils et de services, qui sont largement partagés dans les communautés concernées mais développés pour des objectifs académiques spécifiques : une cyberinfrastructure est quelque chose de plus précis que le réseau lui-même, mais plus général qu'un outil ou une ressource développée pour un projet particulier, une série de projets, ou même plus largement, une discipline toute entière³⁰

Enfin, le terme « e-research » est plus spécifique aux champs des sciences sociales, arts et humanités, domaines qui de plus en plus s'investissent dans des projets de cyberinfrastructures. On trouve aussi en France le terme « e-recherche ».

³⁰ Commission on cyberinfrastructure for the humanities and social sciences, *Our cultural commonwealth: The report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences*, American Council of Learned Societies Commission, 2006. Traduction proposée sur http://unite-numerique.ens-lsh.fr/Pages_publicques/Humanites_numeriques_et_cyberinfrastructure

Même si nous sommes encore à la préhistoire des cyberinfrastructures, on observe déjà, à travers plusieurs projets, un certain élan de démocratisation de la science permis par ces structures. La e-science est en passe de « révolutionner » - pour reprendre le terme du rapport Atkins - la recherche et les échanges informationnels qui la sous-tendent.

b. De nouveaux environnements de recherche

Une fois les différentes ressources disponibles sur le réseau, il s'agit donc de permettre aux chercheurs de les exploiter directement. La puissance des outils permet de passer aisément de la recherche de données (*datamining*) à la simulation et visualisation des données, puis d'y intégrer la possibilité d'échange communicationnel sur l'utilisation de ces ressources. Cet échange favorise les découvertes scientifiques et installe une proximité entre chercheurs, et éventuellement, permet de nouveaux environnements pour l'enseignement. Le champ de la e-science a donc de multiples objectifs recouvrant l'ensemble du cycle de la recherche scientifique : exploitation de données, découverte scientifique, partage, communication et transfert de connaissances (enseignement). Ce constat est parfaitement illustré³¹ par les créateurs de l'un des projets e-science britanniques e-Bank³² :

³¹ Illustration extraite de Gutteridge, Christopher, et al. eBank UK Linking Research Data, Scholarly Communication and Learning. *Third UK e-Science Programme All Hands Meeting (AHM 2004)*, Nottingham, UK, 2004, p.2. [disponible sur :]

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.1925&rep=rep1&type=pdf>

³² Le projet e-Bank a pour objectif la mise en place d'archives permettant le partage et le traitement de données cristallographiques. Le site est disponible sur <http://www.ukoln.ac.uk/projects/ebank-uk/>

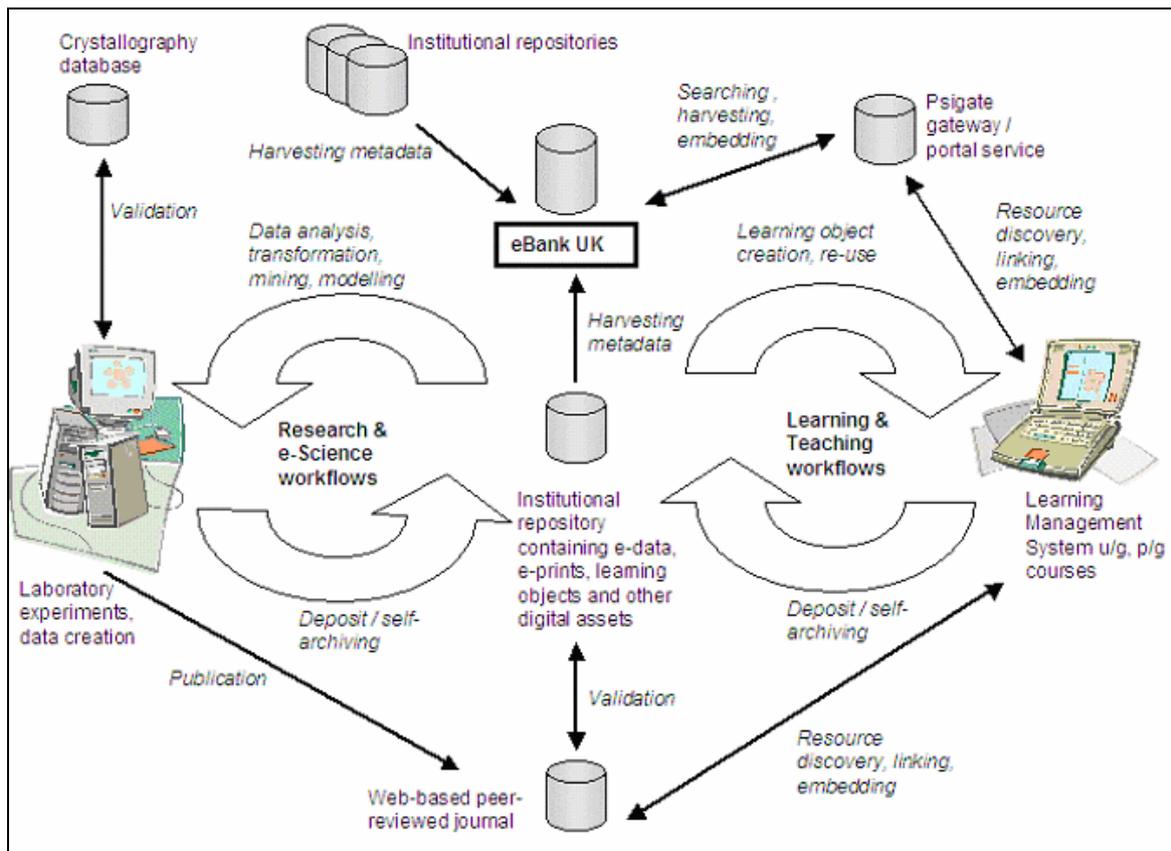


Figure 2. "The scholarly knowledge cycle for research and teaching" selon le projet e-Bank

1. E-Science et Données de recherche

Avec la naissance du mouvement vers l'*Open Access* dans les années 90, les chercheurs sont de plus en plus amenés à déposer leurs publications dans des bases de données ou archives institutionnelles. Cependant, ces dépôts soulèvent l'un des premiers enjeux de la e-science. En effet, de plus en plus de publications sont accompagnées de données (images, graphiques, tableaux etc.) dont l'accès direct et l'exploitation sont entravés par les pratiques individualistes de la plupart des communautés scientifiques (par l'apposition de formats propriétaires par exemple). Mais ces données scientifiques sont également une « production scientifique » dans le sens où elles résultent de l'activité de recherche. Aussi, comment avoir accès à ces données puis comment les exploiter, le challenge premier étant bien évidemment l'interopérabilité des systèmes ?

Dans ce nouveau contexte, il s'agit donc de maîtriser ces données, de les rendre accessibles, et de les rendre réexploitables. Cette idée générale meut l'ensemble du web et plus particulièrement son fondateur, Tim Berners-Lee, qui appelle l'ensemble de la communauté internautes au partage de données par son slogan :

Raw. Data. Now.³³

Il existe déjà des exemples de plateformes présentant articles et données brutes associées (BioLit³⁴ par exemple). Dans des disciplines telles que la physique ou la bioinformatique, le traitement, le stockage et la diffusion des données a pris une ampleur certaine. Ces disciplines jouissent déjà d'une tradition bien établie d'open access (citons ArXiv³⁵ pour les publications ou le Protein Data Bank³⁶ pour les données en bioinformatique). Dans ces disciplines, des normes de description ont alors été établies et soutenues par les politiques éditoriales des revues qui incitent à déposer certaines données dans des archives disciplinaires, les articles pouvant y faire référence via un identifiant unique.

De même, le projet anglais Integrative Biology³⁷ réunit des chercheurs anglais spécialisés dans l'activité électrique des cellules cardiaques et des chercheurs néozélandais spécialisés dans la modélisation des battements du cœur. Ce laboratoire virtuel permet d'interconnecter ces deux groupes et faciliter les échanges de données et le partage des ressources informatiques. À travers ce projet, il s'agit surtout de permettre d'établir des liens entre différentes anomalies génétiques du cœur qui affecteraient l'activité électrique des cellules cardiaques. Aussi, ces nouvelles modalités de partage permettent déjà des avancées scientifiques significatives.

Pour répondre à ces nouvelles exigences de partage, il est plus qu'urgent de développer des systèmes assez performants permettant à la fois d'interconnecter différentes bases de données, d'effectuer des requêtes complexes et d'assurer la pérennité de ces données.

La e-science repose donc sur cette nécessité intrinsèque de concevoir des outils informatiques assez puissants pour permettre la consultation et la manipulation à distance de ces données et de tout autre type de ressource nécessaire à la recherche : publications officielles, cours, workshop, espace de discussion etc. Aussi, les infrastructures conçues pour ces projets reposent sur des grilles informatiques et des superordinateurs.

En effet, le calcul sur grille informatique permet un partage dynamique et sécurisé de ressources émanant de systèmes hétérogènes indépendamment des administrations (départements, laboratoires, institutions et entreprises ...). Un chercheur lambda aura accès,

³³ « Donnée. Brute. Maintenant » Phrase prononcée par Tim Berners-Lee lors de la conférence TED de février 2009 [disponible sur :] http://www.ted.com/talks/lang/eng/tim_berniers_lee_on_the_next_web.html

³⁴ <http://biolit.ucsd.edu/doc/>

³⁵ <http://arxiv.org/>

³⁶ <http://www wwpdb.org/>

³⁷ <http://www.integrativebiology.ox.ac.uk/>

via un portail, aux ressources nécessaires à ces recherches mais pourra aussi les exploiter en ligne grâce à des outils de visualisation de données, de simulation etc. Connecté, ce chercheur pourra donner son avis sur les recherches de ces confrères (peer-review quasi immédiat) et échanger avec des spécialistes du monde entier sur ses thématiques de recherche.

Cette structure en grille permet donc l'interrogation de données de sources multiples. Par exemple, National Virtual Observatory³⁸, projet américain, est une plateforme permettant le dépôt, l'interrogation et l'acquisition de données astronomiques de nombreux télescopes du monde. Cet observatoire est souvent décrit comme « le meilleur télescope au monde alors qu'il n'en est pas un ».

Mais pour être efficaces, ces grilles fonctionnent sur des réseaux haute performance tel le réseau GEANT³⁹ (réseau mondial). Ce réseau très haut débit relie plus de 3000 centres de recherches dans plus de 30 pays. En Europe, il a permis de mettre en place des collaborations dans des domaines de recherche comme le changement climatique, la radioastronomie et les biotechnologies. GEANT sert également de base au réseau de communication mondial pour le Large Hadron Collider devenu la plus grande expérience scientifique jamais entreprise.

Le recours aux grilles informatiques capables de traiter et d'interconnecter de larges silos de données constitue véritablement la voie traditionnelle de la e-science, née dans ces milieux scientifiques particulièrement demandeurs de puissance de calcul et de capacité de stockage.

2. E-Science et Web collaboratif

Nous l'avons dit, la e-science est une science distribuée et avant tout collaborative. Elle permet la création de laboratoires virtuels (*e-laboratory* ou *e-lab*). Ce sont des espaces de collaboration, supportés par une cyberinfrastructure importante, où peuvent se faire des expériences *in silico* et où les diverses disciplines scientifiques construisent un savoir communautaire. Par exemple, un sociologue veut analyser la possibilité de causes socio-culturelles à l'obésité. Il récolte un ensemble de données sur un panel représentatif de la population et les partage dans une cyberinfrastructure. Plus tard, un médecin veut analyser l'éventualité d'un déterminisme génétique à l'obésité. Il pourra récupérer ces données, les

³⁸ <http://www.us-vo.org/>

³⁹ <http://www.geant.net/>

annoter, fournir ses conclusions et repropoter ce nouveau jeu de données à d'autres chercheurs⁴⁰.

La sphère e-science se place à la croisée des différentes notions de collaboration et de communauté que l'on retrouve dans le concept fort vaste de web 2.0. Le monde de la recherche a peu à peu investi ces lieux sans pour autant y avoir un recours systématique. En effet, certains outils pâtissent d'une réputation peu sérieuse (comme Facebook). Pourtant, malgré les controverses qui subsistent autour du terme lui-même, certains projets e-science affirment leur démarche dans le mouvement web 2.0 et ce qu'on a appelé bien vite la Science 2.0.

On a alors vu des sites s'afficher comme des « réseaux sociaux pour les chercheurs ». Par exemple, le site européen ResearchGATE a même pour slogan :

ResearchGATE is for scientists built by scientists driven by the concept of Science 2.0⁴¹.

Les chercheurs peuvent déposer leurs publications, créer des groupes et séminaires virtuels etc. Aux États-Unis, le site NanoHub⁴² fait figure de référence dans le domaine d'intégration d'outils web 2.0 dans une cyberinfrastructure. Plateforme dédiée aux nanosciences, NanoHUB a été créée par la Network for Computational Nanotechnology (fondée par la NSF). C'est un parfait exemple de plateforme e-science & e-learning, de réseau social et d'espace collaboratif (téléchargement / dépôt de publications, notes, présentations, vidéos, applications ...) proposant de nombreux outils (classés par thèmes) de recherche, de simulation, de calcul etc.

Les outils web 2.0 peuvent donc améliorer les collaborations via des laboratoires virtuels. En effet, il existe de nombreux services web collaboratifs grâce au mash-ups⁴³ qui s'essaient aux domaines du partage et de visualisation de données numériques et le site anglais myExperiment⁴⁴ en est certainement la parfaite illustration. Son objectif est de créer un environnement virtuel de recherche pour le partage et le travail collaboratif sur les *workflows*⁴⁵ et autres productions scientifiques. Les créateurs de myExperiment formalisent dans ce projet leur vision d'un cycle e-science : production, publication, réutilisation, évaluation et préservation.

⁴⁰ Cet exemple est l'objet du portail ObesityLab disponible sur <https://www.nibhi.org.uk/obesitylab/default.aspx>

⁴¹ <http://www.researchgate.net>

⁴² <http://nanohub.org/>

⁴³ Le mash-up (ou mashup ou encore application composite) est une application qui agrège du contenu de sources hétérogènes, et donc crée un service nouveau. Pour plus d'informations, se reporter au Glossaire.

⁴⁴ <http://myexperiment.org/>

⁴⁵ Pour une définition de workflow, se reporter au Glossaire.

Il y a donc une forte tendance de nouveaux projets labellisés « e-science » à s'écarter volontairement de la voie traditionnelle qui prévaut en e-science pour adopter des techniques clairement identifiées comme services web 2.0. Ces services permettent une communication et un *peer-review* quasi immédiats et plus interactifs. Parallèlement, des plateformes de vidéoconférences intégrées aux outils de partage de données ont fait leur apparition. AccessGrid⁴⁶ propose un système de vidéoconférence et de messagerie instantanée. Les participants se rencontrent dans des espaces virtuels grâce à différentes technologies multimédia et y partagent des applications (échange de fichiers, présentations ...). Dans ce nouveau paysage de la e-science, on observe deux tendances : certaines plateformes partent du groupe pour partager des documents (comme NanoHub, par exemple), d'autres s'étayent sur les documents pour construire les communautés (myexperiment). D'une prouesse technologique, la e-science est alors en passe de devenir surtout un nouveau mode de communication et de publication de résultats conçu autour de cette notion de communauté. Les services « Grille » purs s'étiolent au profit d'Open Grids 2.0, capables de colliger des flux de données issues de différentes sources (: des bases de données aux billets de blogs, signets en ligne et documents des archives ouvertes) et sur lesquels viennent se construire des organisations virtuelles d'utilisateurs. C'est bien l'ensemble de l'activité des chercheurs qu'il convient de connecter. Cela vient ajouter une étape supplémentaire au cycle de la recherche scientifique :

Work. Finish. Publish. *Release*⁴⁷

Textuellement : « Travailler. Finir. Publier. Libérer ». Ce concept de libération fait référence au mouvement open source et à la nécessité pour les chercheurs de rendre leurs travaux accessibles et surtout réexploitables. Un projet n'est pas véritablement terminé si le code et les données utilisées et obtenues ne sont pas publiés sur le web. On peut cependant comprendre les réticences des chercheurs à adopter ce modèle : perte de temps, peur de se faire « voler » ses données, doute sur la pérennité des documents mis en ligne ... Cependant, on observe un mouvement dynamique dans tous les champs de la recherche à tendre vers ce réflexe – le premier bénéfice étant bien évidemment une démocratisation de la science quasi inédite.

⁴⁶ <http://www.accessgrid.org/>

⁴⁷ Gezelter, Dan, *What, exactly, is open science ?* billet du blog openscience.org, posté le 28/07/2009 [disponible sur :] <http://www.openscience.org/blog/?m=200907>

3. *E-Science et Apprentissage*

Dans son cycle e-science vu précédemment, le projet anglais e-Bank définit l'apprentissage en ligne (le *e-learning*⁴⁸) comme la dernière étape de la chaîne :

experimentation, analysis, publication, research, and **learning**⁴⁹.

Alors que le monde scientifique semble s'emparer progressivement du concept de cyberinfrastructure pour la recherche, peu de projets ont mis en place une cyberinfrastructure commune pour la recherche et l'apprentissage. Les organisations virtuelles sont alors confinées dans des espaces réservés ne permettant que peu l'interaction étudiant/professeur.

Pourtant, les technologies développées par la e-science présentent des opportunités inédites pour mettre en place de nouvelles méthodes d'enseignement et d'apprentissage. L'accès aux données brutes de la recherche peut permettre à un étudiant, par exemple, de comprendre un peu mieux l'ensemble des concepts parfois obscurs mobilisés pour une expérience. Il s'agit donc maintenant de dépasser les frontières entre e-science et e-learning, de créer des liens entre les communautés de chercheurs, enseignants et étudiants pour établir des collaborations.

L'interactivité et les possibilités de communication (synchrone et asynchrone) des cyberinfrastructures sont un atout pour la compréhension et la participation des étudiants au processus de recherche. Après lecture d'un cours comprenant des exemples d'expériences, un étudiant pourra aller chercher les jeux de données relatifs à ces expériences, les réutiliser selon les principes qu'il aura appris dans ce cours, et poser directement des questions à la communauté scientifique. L'interactivité des nouveaux outils (visualisation en 3-d ...) permet une visualisation de concepts scientifiques parfois obscurs. Le site Nanohub propose des ressources en ligne sur les nanotechnologies (cours, présentation, forum, articles, séminaires ...) dont un grand nombre d'outils de simulation et de modélisation. Les utilisateurs peuvent laisser des commentaires et noter l'outil, ou poser des questions quant à son utilisation. C'est là finalement la vraie nature de la recherche scientifique : investigation, recherche, analyse et découverte. L'étudiant pourra donc prendre part à ce processus et ainsi en comprendre un peu mieux les enjeux.

Basées sur des cyberinfrastructures, les nouvelles plateformes d'e-learning développent de véritables communautés numériques d'apprenants. Citons par exemple le site

⁴⁸ Pour une définition d'e-learning, se reporter au Glossaire.

iplantcollaborative⁵⁰ qui a pour objectif de créer une communauté virtuelle en sciences végétales. Il s'agit de promouvoir une communauté multidisciplinaire de scientifiques, enseignants et étudiants et de participer à l'élaboration d'une cyberinfrastructure commune qui permettra une meilleure compréhension et des avancées significatives dans ce domaine de recherche.

Mais comment faciliter l'adoption de ces cyberinfrastructures par des étudiants tout à fait étrangers à ces technologies ? Les technologies actuelles ont modifié la façon dont les élèves ou étudiants apprennent. Avec la multiplication des blogs scientifiques vulgarisateurs, les nombreux réseaux sociaux, les cours en ligne ... tout étudiant a la possibilité d'apprendre seul, de chez soi et ce, 24h/24. Ces nouveaux espaces d'e-learning investis par les étudiants sont de véritables espaces personnalisables avec affichage du profil, possibilité d'enregistrements de données, export personnel vers son ordinateur etc. Dans ces réseaux sociaux, les étudiants peuvent créer de la connaissance dans des groupes contribuant à une intelligence collective. La e-science doit donc s'adapter aux nouveaux moyens d'apprentissage et de communication utilisés par les étudiants. Les projets e-science doivent donc profiter conjointement des cyberinfrastructures et des outils web 2.0. Cela permettrait de réduire le fossé entre recherche et éducation. À l'heure actuelle, il semble important de familiariser les étudiants à la e-science et de les préparer à ce déluge de données déjà bien amorcé. Ce sont eux surtout qui seront concernés. Il faut donc permettre aux étudiants de profiter de ces cyberinfrastructures mais en les proposant dans des environnements qui leur sont familiers tels que les réseaux sociaux.

La notion de *learning centre* (« centre de ressources pour l'information et la recherche »⁵¹) est intrinsèquement liée à ce contexte. C'est en fait la représentation physique de ce que peuvent être ces infrastructures informatiques. De nombreux projets voient le jour. Citons par exemple le Rolex Learning Center de Lausanne (Suisse) inauguré au début de l'année 2010. L'Université Paul Sabatier (Toulouse) mène un projet similaire mais principalement axé sur le développement d'une bibliothèque multimédia multi-usages. Qu'ils

⁴⁹ Tony Hey et A.E. Trefethen, Cyberinfrastructure for e-Science, *Science*, vol. 308, Mai 2005, p. 817.

⁵⁰ <http://www.iplantcollaborative.org/>

⁵¹ Traduction proposée par Suzanne Jouguelet. *Les Learning centres : un modèle international de bibliothèque intégrée à l'enseignement et à la recherche*, Inspection générale des bibliothèques, 2009, p.1. [disponible sur :] http://media.enseignementsup-recherche.gouv.fr/file/2009/33/6/Rapport_Learning_Centers_7-12_RV_131336.pdf

soient matériels ou virtuels, ces nouveaux espaces d'*e-learning* sont partie prenante du monde naissant de la e-science.

La e-science recouvre donc parfaitement le spectre de la recherche scientifique. Mais les études sur le sujet sont encore fort récentes et les différentes communautés scientifiques ne sont que relativement investies ou seulement conscientes de ces nouvelles méthodes de recherche. Dans cette phase de balbutiement théorique, seule une communication claire et pédagogique pourra lancer un mouvement d'adoption de la e-science d'ampleur assez significative pour perdurer dans le temps.

c. France et e-science : Informer et impliquer

Au vu de ces différentes observations sur l'état de la e-science, force a été de constater que la France n'a que peu de connaissance du sujet. Le site mis en place dans le cadre de ce stage a donc une double vocation : informer et provoquer une prise de conscience de la part de l'ensemble des acteurs concernés, chercheurs et documentalistes scientifiques.

1. Position actuelle de la France dans la e-science

Le terme « e-science » n'est apparu que tardivement en France. Il faut attendre 2005 pour que le terme commence à se déployer dans la littérature scientifique française. Cependant, de nombreux projets de recherche démontrent un engagement profond de la France pour le partage de ressources afin de permettre de nouvelles expériences scientifiques et de motiver de nouvelles collaborations. Cet engagement a été pris en compte dans la mise en place du site. Il s'est agi de montrer à la communauté que la e-science n'est pas un concept nouveau, obscur et réservé à quelques grandes institutions internationales. Au contraire,

e-Science is not a new scientific discipline in its own right : e-Science is a shorthand for the set of tools and technologies required to support collaborative, networked science⁵²

En France, ces outils supportant une science collaborative et distribuée existent déjà. L'exemple significatif est bien évidemment le déploiement de grilles informatiques qui allient partage et puissance de calcul. La France est l'un des pays leaders en matière de grilles informatiques. Grâce à internet, il est possible aujourd'hui d'interconnecter des machines du monde entier afin d'exploiter leur puissance et leur capacité de stockage de données. Cette

évolution scientifique et technologique a donné naissance aux grilles informatiques. Les Français se sont rapidement essayés à ce nouveau concept. Avec e-Toile⁵³, premier projet de grille à haute performance en France et soutenu par le CNRS, la recherche française crée une émulation autour de cette révolution technologique. C'est avec Grid5000⁵⁴ surtout, que la grille informatique française connaît son apogée.

Projet lancé en 2003, son objectif est de créer une infrastructure informatique pour réaliser des expériences dans le domaine des systèmes distribués à grande échelle. Cette grille expérimentale a pour but d'interconnecter 5000 processeurs – objectif atteint durant l'hiver 2008-09 – répartis sur neuf sites géographiques différents : Bordeaux, Grenoble, Lyon, Lille, Nancy, Rennes, Toulouse, Sophia-Antipolis, Orsay.

Au cours de notre étude, nous avons également pu constater que l'une des particularités françaises dans le champ de la e-science est sans doute le fait que l'appropriation de ces notions de services intégrés tout au long du processus de recherche s'est surtout opérée dans le domaine des Digital Humanities. Avant la diffusion du terme e-science, ce champ disciplinaire avait déjà pris toute la mesure de l'importance du partage de ressources en ligne. En 1999, le portail dédié aux SHS Revues.org voit le jour. Développée par le Centre pour l'édition électronique ouverte (Cléo), cette plateforme d'édition électronique publie en ligne des revues en texte intégral de quelques deux cents périodiques. Elle intègre également la plateforme de publication de carnets de recherche Hypothèses (des carnets de fouille archéologique aux carnets de terrain ou d'enquêtes en passant par des blogs de projets en cours etc.). Enfin, Calenda annonce les prochains événements scientifiques en SHS. Il s'agit donc véritablement d'une plateforme complète et collaborative d'information scientifique et de diffusion des résultats de la recherche : de la soumission d'un article, au suivi d'un projet, et à l'annonce d'événements scientifiques à venir. Revues.org a par ailleurs été récompensé du titre de meilleur projet e-science en 2003 pour la France dans le cadre du concours « World Summit Awards. The best in e-content and e-creativity ».

Cette ouverture des ressources s'affirme encore un peu plus quand, en mars 2007, le CNRS rassemble une équipe pour concevoir un « Très Grand Equipement » dédié aux sciences humaines et sociales afin de permettre un accès simplifié aux ressources utiles aux chercheurs de la discipline. Ce projet d'« Accès unifié aux données et documents numériques

⁵² Hey, Tony, et Jessie Hey. E-Science and its implications for the library community. *Library Hi Tech*, vol.24, n° 4, 2006. p.516. [disponible sur :] http://conference.ub.uni-bielefeld.de/2006/proceedings/heyhey_final_web.pdf >

⁵³ <http://www2.cnrs.fr/presse/communiqu207.htm>

⁵⁴ <https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home>

des SHS » (ADONIS) répondra à plusieurs objectifs : accès unifié aux données de la recherche, réseau social SHS, partage de référence, gestion de *workflow*, calendrier, traduction, outils de manipulation d'images et vidéo, géo-localisation de ressources etc.

Ce projet s'interroge également sur les questions d'interopérabilité, de mutualisation, et de collaboration pour permettre aux SHS de se développer tant à l'échelle nationale qu'internationale. Pour ce faire, le TGE-ADONIS travaille étroitement avec le Centre de Calcul CC-IN2P3 qui a développé une expertise dans les technologies de grille et le Centre pour la communication scientifique directe (CCSD), une unité propre de service du CNRS.

Cependant, dans la littérature spécialisée sur la e-science, la France n'apparaît que peu. On trouve d'ailleurs ce type de discours :

Le débat sur l'e-science est beaucoup moins mature en France, même si, individuellement, les laboratoires s'équipent de logiciels spécialisés gérant les flux de tâches en leur sein. A rebours de ce que l'on constate en Allemagne et au Royaume-Uni où des politiques et des programmes « e-Science » ont été définis, ce qui semble faire défaut en France est une réflexion centrale sur les « infrastructures » réseaux/logiciels de la recherche publique conçue comme un système global. Paradoxalement, la situation française s'apparente plus sur ce point à celle constatée aux États-Unis, où (sauf sur les évolutions futures d'Internet, qui font l'objet d'un programme national) l'initiative en matière d'e-Science est largement laissée aux organismes de recherche eux-mêmes.⁵⁵

La France n'a pas pour le moment lancé de programme national comme le Royaume-Uni l'a fait avec « e-science program ». Il faut également ajouter que certains projets sont encore à leur état embryonnaire. Ce que nous avons pu constater est que les différents projets rencontrés ne portent pas de labellisation « e-science ». Mais outre ces considérations terminologiques, un certain nombre de projets français regroupent les différentes problématiques inhérentes à la e-science : partage de ressources (que ce soit puissance des outils ou libre accès à la production scientifique), création d'organisations virtuelles pour une collaboration plus étroite entre les acteurs du monde de la recherche ... Ainsi donc, la recherche française, Jourdain des temps modernes, fait depuis longtemps de « la e-science sans le savoir ». L'enjeu est donc de montrer à la communauté l'ensemble des outils et des compétences qu'elle possède déjà pour éveiller une prise de conscience et faciliter l'adoption *a posteriori* des notions clés de la e-science.

⁵⁵ Communiqué de presse i-expo 2006 du 20 mars 2006 [disponible sur :] http://www.i-expo.net/documents/2006/ist_5.htm >

2. De la prise de conscience à l'implication : informer et conceptualiser la e-science

La difficulté est donc d'informer la communauté française sur ce nouveau concept. On imagine aisément les réticences de la communauté scientifique à adopter ces nouvelles méthodes. Nous avons pu constater que la e-science se décrit beaucoup en termes issus du monde informatique : grille, *middleware*, logiciels etc. Pour le scientifique non informaticien (pensons surtout et au risque de réveiller quelques clichés aux chercheurs en arts et sciences humaines), voilà bien de quoi le rendre réfractaire à ces nouvelles méthodes de travail. Le concept même de grille est souvent perçu comme un équipement informatique complexe permettant de résoudre des problèmes scientifiques de grande échelle et donc réservé à des disciplines comme la physique et la biologie, pionnières dans ce domaine. Si cela a été vrai jadis, il reste à convaincre des communautés scientifiques plus petites ou appartenant aux champs des SHS, Arts et Humanités, des avancées significatives qu'elle permet déjà dans leur domaine : recherche, partage et visualisation de données, *peer-review* interactif, communautés virtuelles de travail etc.

Pour toucher un large public, il a donc été nécessaire de conceptualiser les méthodes e-sciences en prenant en considération les différences de pratiques inhérentes à chaque communauté, leur propre culture, leur propre « langage » parfois bien éloigné des considérations informatiques et leur propre intérêt à adopter ces technologies. Il faut informer et vulgariser pour ces disciplines qui ne sont pas « data-intensive ».

Le site a toujours été pensé selon ces différentes exigences. Par exemple, nous avons opté pour une représentation visuelle⁵⁶ des notions recouvertes par la e-science en page d'accueil du site (voir image ci-dessous) :

⁵⁶ Pour un aperçu de la page d'accueil du site, voir Annexe I. Cette visualisation a été réalisée à partir du site Manyeyes disponible sur <http://manyeyes.alphaworks.ibm.com/manyeyes/>

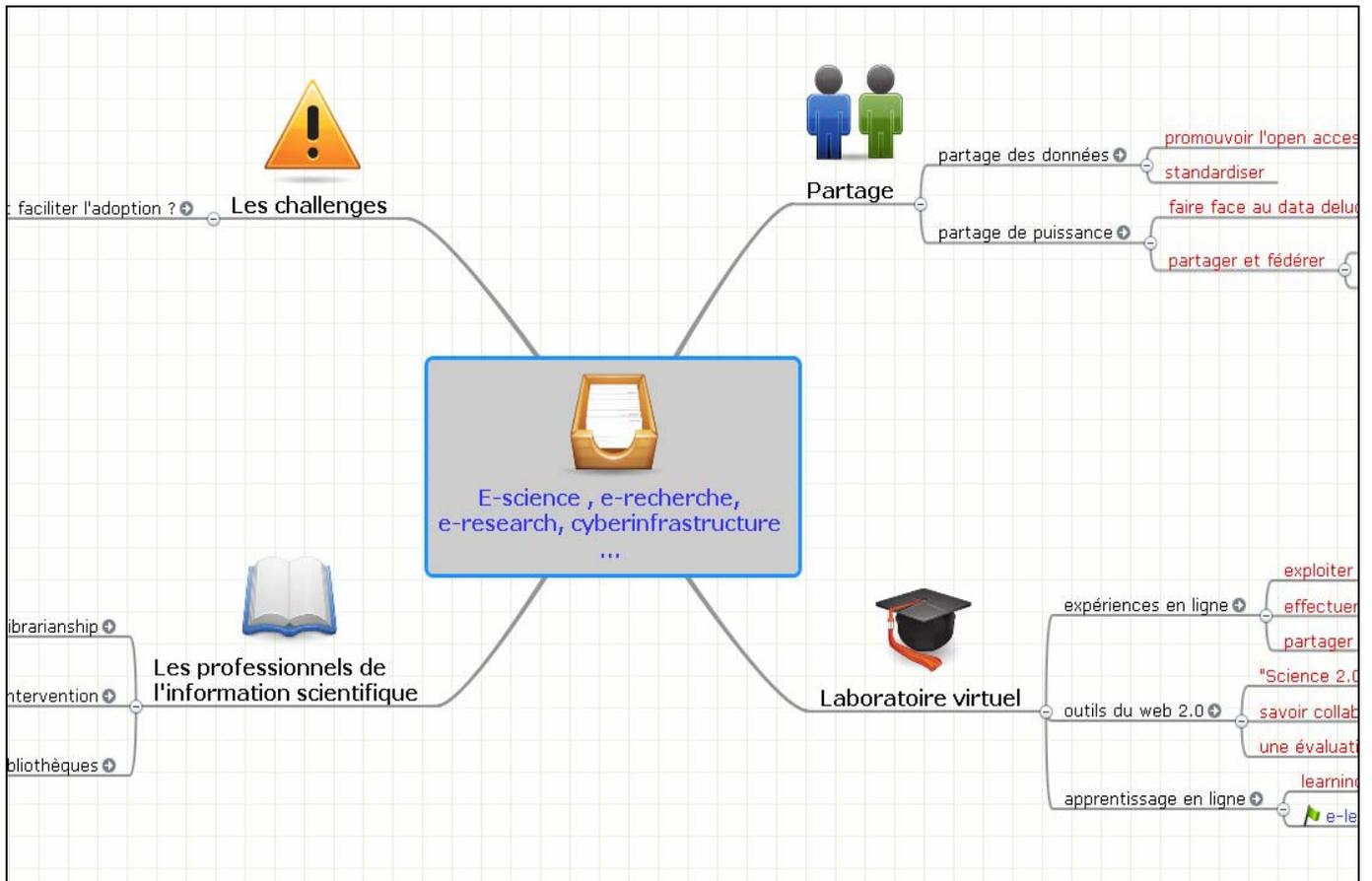


Figure 3. Mind map

Il nous a semblé important au-delà de la conceptualisation, d’accompagner nos propos d’exemples de projets plus évocateurs que la théorie pure. En effet, comme le rappelle la Commission “e-infrastructures européennes” :

scientists are people, and people like stories⁵⁸.

Ce rapport recommande simplement de mettre en lumière des « *success stories* » en matière de e-science dans des disciplines voisines du public visé. Une dissémination efficace doit ainsi promouvoir le voisinage entre disciplines pour qu’émergent des communautés interdisciplinaires stimulant de fait une adoption plus rapide de la e-science. C’est donc ce que nous avons choisi de faire en répertoriant différents projets e-science et en choisissant d’en mettre certains en lumière dans une catégorie nommée « Zoom sur ... ». Nous travaillons actuellement à la construction d’un formulaire permettant aux internautes de chercher

⁵⁸ Collectif. *Report of the 7th e-Infrastructure Concertation Meeting e-Infrastructures for e-Science*. Bruxelles, Belgique : European Commission – DG INFSO/F3 GÉANT & e-Infrastructure Unit, 2009, p.43. [disponible sur :] <http://www.beliefproject.org/docs/7th-eConcertation%20Meeting-Report-Final.pdf/view>

directement sur notre site des projets selon le pays ou la discipline et d'en avoir une description synthétique.

Le site E-Science créé a donc dû répondre à un nombre important d'exigences. Tout en adoptant une visée pédagogique, il s'est agi de conceptualiser des concepts parfois obscurs et à la croisée de différents champs disciplinaires (sciences de l'informatique et des réseaux, sciences de l'information ...). Ce travail de définition, de présentation de projets, de questionnement nous a permis de prendre conscience des enjeux que ces nouveaux environnements de recherche représentent particulièrement pour les métiers de l'IST. Mais comment préparer dès aujourd'hui les professionnels à ce nouveau bouleversement ?

III. E-science et professionnels de l'IST

Les différentes ressources que nous avons pu trouver sur la e-science ont toujours intégré les professionnels de l'information scientifique comme acteurs primordiaux de ce bouleversement. Considérés comme experts de la structuration de l'information, les documentalistes scientifiques doivent prendre part à cette « révolution » e-science imaginée par Dan Atkins, sans quoi ces nouvelles méthodes scientifiques ne sauraient perdurer dans le temps.

a. Nouveaux enjeux

Un service d'information – quel qu'il soit – doit sans cesse réadapter ses services en fonction des évolutions technologiques et de leurs impacts sur les pratiques communicationnelles et informationnelles de la communauté qu'il dessert. L'avènement de la e-science que nous connaissons depuis peu (bien qu'il s'agisse d'une réalité encore bien peu tangible en France) est en passe de révolutionner le monde de la recherche et des services qui lui sont associés. Les mouvements initiés ces dernières années par les bibliothèques telles que l'open access et le protocole OAI-PMH ont permis un nouvel espace de circulation d'informations plus dynamique et universel. Mais la e-science, e-research, cyberinfrastructure ou peu importe le terme qu'on lui donne, recouvre un mouvement bien plus ample de circulation de données qu'il s'agit de gérer, de rendre accessibles et bien évidemment de préserver.

L'une des conséquences de la e-science est donc la mise à disposition imminente de larges bases de données issues de la nouvelle génération d'expériences et d'instruments scientifiques. Ce « data deluge » amorcé depuis quelques années entraîne un changement de paradigme radical dans la capture et la diffusion de l'information scientifique et implique de fait de nombreux défis pour les scientifiques qui auront besoin des compétences et des technologies issues de l'informatique et du monde des bibliothèques pour gérer ces données. Il faudra également mettre en place des outils adéquats de téléchargement, d'exploitation et de visualisation pour l'utilisateur final afin de créer un environnement de travail transparent.

1. *Du document aux données de la recherche*

Les bibliothèques ont toujours été les garantes d'expertise en matière d'organisation, d'enrichissement et de diffusion de l'information mais représentent encore un bien lourd héritage du document en tant qu'unité cloisonnée. L'ère numérique n'a pas encore changé cette représentation. On parle toujours de document numérique en tant qu'objet délimité et fermé. Or, de plus en plus de publications intègrent des éléments hétérogènes : tableau de données, extraits audio ou vidéo... La e-science tend à décroisonner ces objets numériques afin de rendre l'ensemble des éléments de la recherche exploitables et liés entre eux. Auparavant, l'échange de données scientifiques pour différentes expériences était difficile : problème d'accès, de droit, d'interopérabilité etc. Les cyberinfrastructures se fondent elles sur un partage sécurisé de ces données pour un travail collaboratif bien plus interactif et efficace.

Ces nouvelles opportunités de services de post-publication vont permettre aux bibliothèques de s'éloigner de la représentation statique du document et de créer grâce aux technologies du web un réseau d' « artefacts »⁵⁹ de recherche interconnectés. C'est certainement la partie la plus importante dans la gestion de ces données : les liens qui connectent l'ensemble des éléments d'une recherche (auteur, données, publications, institutions, recherches similaires etc.).

Cette nouvelle dynamique nécessite la mise en place d'annotation et d'ontologies afin de créer du sens entre ces données. Cela est depuis peu possible grâce au mouvement du web des données. Ce web des données a pour base le projet *Linking Open Data* du W3C mené par Tim Berners-Lee. Le mouvement d'open data a permis des initiatives importantes (d'entreprises, de gouvernements ou d'individus via les réseaux sociaux) qui vont dans ce sens. L'objectif premier du projet est donc de publier des répertoires de données ouvertes sur le web et de créer des liens entre les données des différentes bases. Cela représente une opportunité certaine pour les projets e-science. En effet, le mouvement Linked data a établi des lignes pour faciliter la connexion entre des données qui ne l'étaient pas auparavant et beaucoup de banques de données suivent cette méthode. Ce web des données s'appuie sur le format RDF (Resource Description Framework). Il s'agit d'un langage de description dont la grammaire est extrêmement simplifiée et se repose sur des « triplets » : sujet-prédicat-objet. Cela permet d'écrire des phrases suffisamment simples pour que les machines puissent

⁵⁹ Dans la littérature spécialisée sur la e-science, on parle volontiers d'artefacts plutôt que de données ou de publications tant les objets issus de la recherche sont hétérogènes et protéiformes.

échanger des données. Ces phrases (« déclarations ») permettent d'ajouter du sens aux pages web. Par exemple, la déclaration : « ce site a pour titre « E-Science » » donne :

sujet : site identifié par son URL; **prédicat** : titre; **objet** : « E-Science »

En 2009, à l'occasion de la conférence TED, Tim Berners-Lee promeut ce projet « Linked data » dont les fondements sont le partage et l'interconnexion des données sur le web. Ce ne sont plus les documents qui sont liés (des sites internet pointant entre eux par des liens hypertextes) mais les données qui sont interconnectées. Il en appelle à l'ensemble des internautes – avec son fameux « Raw. Data. Now » pour que tous nous partagions nos données et que nous les liions entre elles. De nombreuses initiatives ont déjà adopté le standard RDF. DBpedia est un excellent moyen de comprendre RDF. Ce projet extrait les nombreuses données de Wikipédia et les structure grâce à RDF en créant des liens sémantiques entre elles⁶⁰.

C'est dans toute la complexité de ce nouveau concept que les professionnels de l'IST doivent aborder le partage de données.

2. Du prestataire au partenaire de la recherche

Dans ce nouvel environnement virtuel de recherche, le bibliothécaire, l'archiviste ou le documentaliste devra intervenir tout au long du processus de recherche pour décrire au mieux les données selon leurs différentes phases de traitement. Or la plupart des bibliothécaires sont plus familiers avec la phase de post-recherche (rapports, communication, publications) qu'au spectre de la démarche de recherche proprement dit. Dans la e-science, les données ne sont plus considérées comme des produits intermédiaires, oubliées une fois que la publication finale paraît. Au contraire, les différentes phases permettent d'apprécier différentes informations capturées à des étapes d'exploitation précises. Au cycle de la publication scientifique bien connu des professionnels de l'information, il faudra rajouter celui de la phase de pré-publication correspondant à l'expérimentation et à la validation de l'hypothèse scientifique :

⁶⁰ Voir par exemple la transcription de l'article Berlin disponible sur <http://dbpedia.org/page/Berlin>



Data and publication life cycles⁶¹

Au vu de ce schéma, c'est bien à toutes les étapes de la recherche que devront intervenir les professionnels de l'IST pour s'assurer que les différents composants de la recherche (données, articles, workshops, logiciels) sont correctement annotés et enregistrés dans des formats pérennes :

[D]ata-services will at some point be a model for library services in general as libraries deal increasingly with digital materials and become more of a lab than a warehouse....This opens up the possibility of the data librarian working with researchers at the earliest stages of research: helping with the documentation process and ensuring that the data will be **preservable, usable, and re-usable for the long-term. (nous soulignons)**.⁶²

Au-delà de la préservation des données, il faudra prendre en compte la nécessité intrinsèque à la e-science de proposer des services collaboratifs ou dits « web 2.0 » afin de répondre aux exigences du travail en réseau et du partage de connaissances. De plus, la e-science introduit un changement d'échelle radical en se fondant sur les grilles informatiques et sur la puissance de calcul de plusieurs centaines (si ce n'est milliers) de processeurs. Il faudra composer avec ce changement d'échelle en terme d'offre de contenu mais également de public visé. En effet, les chercheurs créent et utilisent des jeux de données de plus en plus importants et travaillent de plus en plus en équipe interdisciplinaire et multi-institutionnelle. Les bibliothèques de recherche sont souvent conçues autour d'un axe disciplinaire ou d'une institution. Les projets e-sciences remettent en cause cette politique en ayant pour fondement principal un accès universel à la connaissance. Il ne s'agira donc plus de desservir une certaine communauté mais de proposer des services communs à des disciplines aux pratiques pourtant hétérogènes.

⁶¹ Anna Gold, Cyberinfrastructure, Data, and Libraries, Part 1&2. *D-Lib Magazine*, vol.13, n° 9/10, Octobre 2007. [partie 1] <<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>>

⁶² James Jacobs et Charles Humphrey, « Preserving research data », *Communications of the ACM*, Vol. 47, n°9, 2004, p. 27.

b. Les domaines d'intervention

La e-science introduit une ouverture progressive des services proposés par les bibliothèques afin que ceux-ci soient plus intégrés au travail de recherche scientifique. L'environnement e-science est dynamique, multidisciplinaire et hétérogène et transforme les outils et les méthodes de recherche. Ces challenges, bien loin d'effrayer le monde des bibliothèques, s'avèrent être de véritables opportunités. Certes, si nous en sommes encore à la préhistoire de la e-science, des essais, des initiatives, des recommandations à l'attention des bibliothèques de recherche voient le jour afin de mieux aborder ce bouleversement. De grandes organisations professionnelles dont l'American Society for Information Science & Technology, l'American Library Association (ALA) et l'Association of Research Libraries (ARL) mènent des activités de sensibilisation et d'investissement dans des projets e-sciences.

En 2007, l'ARL a organisé un groupe de travail « E-science Task Force » afin d'étudier les nouveaux champs d'action pour les bibliothèques de recherche ainsi que les nouvelles collaborations à entreprendre avec le monde scientifique. Le rapport⁶³ résultant de ce groupe de travail dresse donc une liste des domaines d'intervention et des compétences (acquises ou à développer) qui seront nécessaires dans la formation à ces nouvelles infrastructures. L'ARL organise le rôle des bibliothèques de recherche selon trois champs d'action que nous détaillons et enrichissons de nos lectures ci-après.

1. La gestion des données

Jusqu'à récemment, les journaux imprimés et les actes de conférence étaient le moyen le plus efficace pour la diffusion des résultats de la recherche. Mais l'arrivée du web a démultiplié les supports de communication et leurs modes de circulation, modifiant ainsi les enjeux de la communication scientifique. Aujourd'hui, les chercheurs peuvent diffuser leurs articles en pre ou post-publication directement sur leur page personnelle tandis que les éditeurs peuvent mettre en ligne leurs archives. Nous l'avons vu, la e-science induit plus qu'un simple dépôt et déverse des publications de plus en plus complexes intégrant de plus en plus de matériaux hétérogènes (tableau, audio, vidéo, jeux de données numériques etc.). La e-science soulève un nouveau défi pour la diffusion de l'information scientifique :

⁶³ Lougee, Wendy et al. *Agenda for Developing E-Science in Research Libraries*. Association of Research Libraries, 2007. 26p. <http://www.arl.org/bm~doc/ARL_EScience_final.pdf>

The traditional, linear, batch processing approach is changing to a process of continuous refinement as scholars write, review, annotate, and revise in near-real time using the Internet.⁶⁴

Le caractère collaboratif des environnements e-science introduit de nouvelles modalités d'écriture et de lecture : écriture collective, annotation, *review*, folksonomies etc. De nouveaux paramètres sont également à prendre en compte pour l'exploitabilité des résultats de la recherche : formats, versions, logiciels propriétaires. La e-science est donc en passe de modifier la définition traditionnelle et linéaire d'une publication scientifique – nous généralisons bien évidemment, les différences entre disciplines relativisant ce mouvement global. Cela induit une participation proactive des bibliothèques en amont de la recherche (processus de génération des données) et en aval (publication et préservation).

La création des données :

Anna Gold⁶⁵ insiste sur la standardisation des processus de génération de données et des métadonnées ainsi que sur une meilleure communication auprès des chercheurs pour une adoption plus rapide de ces nouvelles méthodes de travail. Il existe de grandes variations d'utilisation des formats et des types de données selon les disciplines et au sein même de celles-ci. Les bibliothèques doivent alors participer à l'élaboration de répertoires plus dynamiques supportant des *workflows* standards d'expérimentation et de génération de données. Un identifiant unique pourra ainsi être attribué à une donnée ou un jeu de données afin de les rendre explicitement repérables et exploitables.

La diffusion :

Les publications hybrides émergentes (articles et données) doivent être précisément annotées afin de participer à un réseau intelligent de diffusion du savoir. Ces nouveaux objets scientifiques sous-tendent des réseaux de liens qu'il faut rendre repérables. Cela nécessite en amont des ontologies et des systèmes d'annotation pertinents et en aval, des outils adéquats de recherche, de visualisation et de simulation. Les professionnels de l'information scientifique doivent également considérer le social bookmarking comme une aide à l'indexation tant le volume de données à traiter est important. Il s'agit de créer autour des publications de véritables communautés de savoir.

⁶⁴ Rapport Atkins, *op.cit.*, p.42.

⁶⁵ Anna Gold, *op.cit.*

La préservation :

Les professionnels de l'information sont garants d'une expertise en matière de gestion et de préservation de données. Cependant, le changement d'échelle induit par le *data deluge* introduit une complexité telle qu'il est urgent de définir de nouveaux modèles économiques pour la pérennité de ces nouveaux objets issus de la recherche. Les bibliothèques pourraient alors accueillir les centres de préservation numérique comme c'est déjà le cas avec l'établissement récent du Digital Data Curation Center⁶⁶ (D2C2) dans les bibliothèques de l'Université de Purdue. Dans ce domaine, on parle déjà de « *data-librarian* ».

2. Les communautés virtuelles

Les organisations virtuelles remettent en question les politiques actuelles de licence et d'accès à l'information. Les bibliothèques devront gérer cette nouvelle mise à disposition de l'information et son traitement par les différentes communautés d'utilisateurs (*reviews*, *folksonomie* etc.). Dans ces nouveaux environnements, il ne s'agit plus « simplement » d'un stockage de données. Les chercheurs ont besoin de bibliothèques virtuelles permettant des environnements collaboratifs dans lesquels la communication est continue et synchrone. Cela introduit une complexité supplémentaire à la mise à disposition des objets numériques. Il faut proposer des services de « co-laboratoires » communs pour l'ensemble des communautés. Pour répondre à cet objectif, la Cornell University Library a créé une communauté virtuelle en sciences de la vie. VIVO⁶⁷ regroupe les informations disponibles publiquement sur les personnels, départements, domaines de recherche et équipements de l'Université Cornell. Cela a permis la construction d'un réseau de scientifiques qui rendent publics leurs projets et publications et permet ainsi d'initier de nouvelles collaborations. VIVO sert à trouver des collaborateurs localisés dans différents services, ou, pour les étudiants, à trouver de nouvelles ressources. Sept universités américaines vont recevoir 12,2 millions de dollars pour mettre en place ce réseau « VIVOweb », qui sera par la suite étendu au reste des États-Unis.

Comme le mentionne Cliff Lynch⁶⁸, les nouveaux environnements virtuels de recherche peuvent sembler de prime abord ne pas s'adapter à la tradition plutôt passive des bibliothèques et en marge du processus même de recherche. Ces environnements rassemblent des chercheurs et des ressources hétérogènes dispersés géographiquement, ce qui soulève des difficultés pour les bibliothèques traditionnellement habituées à servir une communauté locale

⁶⁶ <http://d2c2.lib.purdue.edu/>

⁶⁷ <http://vivo.cornell.edu/>

ou une institution. Pourtant, des initiatives significatives sont en passe de modifier cette passivité. Par exemple, la bibliothèque Albert R. Mann de Cornell développe DataStar⁶⁹ (a Data Staging Repository) avec le soutien de la NSF. DataStar met à la disposition des chercheurs un environnement de partage de données astronomiques dans lequel il est également possible de créer des métadonnées préliminaires puis standardisées selon les normes de la discipline, de publier des jeux de données et d'obtenir à tout moment l'aide de professionnels de l'information scientifique.

3. *Promotion et formation*

Dans ce nouvel environnement e-science, les bibliothèques en tant que médiateurs de l'information se doivent de promouvoir ces nouvelles méthodes de recherche et de sensibiliser les chercheurs à l'importance de la gestion de leurs données tout au long de leur cycle de vie. Cette prise de conscience peut se faire par le biais de communications sur internet, de mise en place de groupes de travail réunissant l'ensemble des acteurs concernés, ou encore lors de conférences et autres *symposia* comme la conférence Librarians & e-Science de 2008 proposée par l'université de Purdue ou plus récemment le Librarian e-Science Symposium en avril 2010 (Massachusetts). Un certain nombre de bibliothèques ont donc mis en ligne un grand nombre d'informations à destination de la communauté scientifique et des professionnels de la documentation. Citons par exemple le e-Science Portal for New England Librarians⁷⁰. Ce portail hébergé à l'université de médecine du Massachusetts proposera un accès unifié à différentes ressources liées à la e-science : informations récentes sur le sujet, présentation du concept, tutoriels d'utilisation des outils, espace de discussion etc. L'enquête préliminaire à l'établissement de ce portail montre que les professionnels de l'information sont encore peu au courant de ces nouvelles évolutions et qu'un tel portail est plus que nécessaire. L'information sur la e-science doit donc s'accompagner d'une formation à ses méthodes. En ce sens, la bibliothèque du MIT a mis en place un site internet qui fournit des conseils aux chercheurs sur la gestion de données. La page d'accueil les avertit :

Managing your data before you begin your research and throughout its life cycle is essential to ensure its current usability and long-run preservation and access.⁷¹

⁶⁸ Lynch, Cliff. The Institutional Challenges of Cyberinfrastructure and E-Research. *EDUCAUSE Review*, vol.43, n°6. 2008, p.1.

⁶⁹ <http://datastar.mannlib.cornell.edu/>

⁷⁰ http://library.umassmed.edu/escience_symposium09.cfm

⁷¹ <http://libraries.mit.edu/guides/subjects/data-management/index.html>

Ensuite, le site dresse une liste de contrôle pour la gestion d'une donnée afin que celle-ci soit réutilisable et préservée dans le temps, et une série de conseils sur les formats de fichiers, sur l'évaluation des besoins, sur la sécurité etc.

À l'image des agences gouvernementales qui promeuvent le libre accès et la libre réutilisation des données obtenues sur des fonds publics, les bibliothèques de recherche doivent soutenir ce mouvement d'open access et participer à la diffusion de l'information autour des formes alternatives de licences (comme par exemple, ScienceCommons⁷²) afin de veiller à la protection intellectuelle du contenu scientifique (notamment pour les données à caractère privatif comme les données médicales). L'ARL, dans sa liste de recommandations à l'intention des bibliothèques de recherche, met l'accent sur ce rôle de médiateur. Elles doivent promouvoir l'*Open Access* principalement en participant à l'effacement des barrières qui empêchaient l'accès aux données de la recherche et qui limitaient les liens entre celles-ci.

Ce mouvement de partage ne peut cependant être accompli sans collaboration. Les bibliothèques doivent collaborer dans des projets multi-institutionnels et interdisciplinaires afin de mettre en place des infrastructures de partage de données, métadonnées et publications. Il s'agit de travailler ensemble pour limiter le déséquilibre de services entre bibliothèques. Enfin, elles doivent assurer une formation continue des utilisateurs mais aussi de leur personnel afin de développer le capital humain nécessaire pour la gestion de ces nouveaux environnements de recherche. Parallèlement, elles doivent participer aux initiatives visant à la diffusion de la e-science pour une compréhension accrue de la part des publics et professionnels.

Le rapport de l'ARL démontre une volonté active de positionner les bibliothèques de recherche dans ce nouveau prisme des e-sciences. Cependant, cela implique pour les professionnels de l'information scientifique une certaine connaissance du domaine de recherche concerné afin de comprendre l'ensemble des mécanismes et des enjeux du processus scientifique. C'est ce « *Go native* » qui martèle la plupart des recommandations : les professionnels de l'information doivent intervenir dès le début de la recherche afin d'encadrer au mieux la génération et la diffusion du contenu scientifique. Les différentes études insistent également sur la nécessité pour les chercheurs de reconnaître cette figure d'autorité en matière d'information scientifique qu'incarnent les bibliothèques de recherche pour une collaboration plus étroite et plus productive.

⁷² <http://sciencecommons.org/>

c. Vers le « e-science librarianship »

« *Data librarian* », « *data manager* », « *e-science librarianship* » ... tels sont les différentes appellations que prennent ces « nouveaux » métiers de l'IST dans la littérature spécialisée sur la e-science. Quelle formation est-il nécessaire aujourd'hui d'assurer pour répondre aux nouvelles exigences des environnements de recherche ? Cependant, si le contexte a certes grandement évolué, les métiers des professionnels de l'IST n'ont-ils pas déjà les compétences nécessaires pour affronter ces changements ?

1. *Quelle formation ?*

Nous avons pu voir que les nouveaux environnements de recherche vont nécessiter une mobilisation forte des professionnels de l'IST. Le rôle traditionnel des bibliothèques d'organisation, de fournisseur d'accès et de préservation de la connaissance devra être enrichi, à la lueur de la e-science, par de nouvelles compétences en matière de gestion de ces nouveaux objets de recherche et des espaces collaboratifs qui les créent. Il s'agit de comprendre le concept d'e-science ainsi que les méthodologies qui lui sont associées.

Les nouvelles missions qui incomberont à ces professionnels répondent déjà au nom générique de « *e-science librarianship* » (on retrouve même un article de Wikipédia sur le sujet⁷³ !). Les métiers de l'IST connaissent un véritable changement de paradigme : il s'agit de passer de la gestion de collections spécialisée (l'ARL le nomme « *the branch library model* ») à un modèle proactif, multidisciplinaire et engagé tout au long du processus de recherche.

Il n'existe pour le moment pas de formation adéquate à ce changement d'implications pour le métier de professionnel de l'information qu'il soit documentaliste, archiviste, ou bibliothécaire, encore peu confrontés pour le moment à la question de la gestion de données brutes scientifiques et des environnements virtuels de recherche. Cependant, certains cours sont apparus aux USA notamment le Digital Library Curriculum Project⁷⁴ construit selon une collaboration entre Virginia Tech et l'Université de Caroline du Nord. Mais citons surtout Frank Olke et Fredric Gey qui proposent un cursus universitaire pour la formation des étudiants au « *social science data librarianship* »⁷⁵. Celui-ci comprend des cours en matière

⁷³ http://en.wikipedia.org/wiki/E-Science_librarianship

⁷⁴ <http://www.ariadne.ac.uk/issue55/henty/#11>

⁷⁵ Olken, Franck et Fredric Gey. *Social Science Data Librarianship – A university Curriculum*. 11 janvier 2006, 19p. [disponible sur :] https://hpcrd.lbl.gov/staff/olken/ssdl/iassist_ssdl_curriculum.pdf

de *digital humanities*, gestion de bases de statistiques, sémantique, apprentissage des notions liées au réseau informatique etc.

Le professionnel de l'information, s'il continue de faire figure d'autorité en matière de gestion de publications, semble devoir se préparer à la gestion et à la préservation de ces nouveaux environnements de gestion intensive de données.

2. *E-Science librarianship : un nouveau métier ?*

Nous avons dressé une esquisse du nouveau contexte de la recherche scientifique avec lequel les professionnels de l'IST auront à composer. Mais, ce bouleversement n'est-il pas déjà arrivé ? Si le terme e-science est récent, les bibliothèques n'ont-elles pas déjà les capacités dans leur histoire pour affronter d'ores et déjà ces nouvelles méthodes scientifiques ?

Certaines bibliothèques de recherche sont déjà depuis quelques années confrontées à la gestion de données brutes. Elles mettent à disposition de leur public des images, sons, cartes géographiques, jeux de données statistiques etc. Il est vrai, cela concerne un nombre restreint de domaines scientifiques. Il s'agit des sciences sociales (avec par exemple la base ICPSR⁷⁶ de l'université du Michigan référençant plus de 50 000 jeux de données), des systèmes d'information géographique (SIG) (avec la base TIGER⁷⁷ en ligne depuis 1990 !) et de la bioinformatique.

De manière plus générale, les bibliothèques détiennent une expertise incontestable, profitable à la e-science. Avec le mouvement général de l'*open access*, les bibliothèques ont développé une expertise en matière de politiques et principes de partage libre de la production scientifique (archives institutionnelles ou disciplinaires). Au sein de ces archives, les professionnels de l'IST ont participé au développement d'outils interopérables (travail sur les métadonnées, standardisation, OAI-PMH) et ont donc les compétences pour affronter un nouvel effort de standardisation des descriptions de données. De plus, elles bénéficient d'une expérience indéniable de stratégies d'archivage à long terme, et sont plus que n'importe quel organisme sensibles au cycle de vie d'une publication scientifique et donc de l'importance de son accès et de sa préservation dans le temps.

En somme, les actions prioritaires relèvent de la prise de conscience. La e-science va bouleverser les carcans habituels des bibliothèques de recherche dont la tradition a longtemps été d'axer leurs services autour d'une discipline ou d'une institution et de gérer les

⁷⁶ <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

publications scientifiques en bout de chaîne du cycle de la recherche. Mais les méthodes e-science ne correspondent finalement qu'à une évolution attendue du mouvement *open access* qui les avaient conduites à s'investir dans des projets d'archives ouvertes. À la communauté des professionnels de l'IST de s'approprier ces nouveaux enjeux et de mutualiser des compétences déjà existantes au sein des corporations pour s'investir à nouveau dans cette continuelle révolution numérique.

⁷⁷ <http://www.census.gov/geo/www/tiger/>

Conclusion

La notion d'e-science, complexe et fort vaste, recouvre un prisme particulièrement large d'outils de recherche intégrés dans des plateformes interopérables permettant un travail collaboratif. S'appuyant sur des infrastructures informatiques puissantes, ces plateformes permettent la mise à disposition ainsi que l'exploitation de la production scientifique (de la donnée brute à l'article final). L'essor d'internet a rendu possible des collaborations inédites au sein des diverses communautés scientifiques qui permettent déjà des découvertes scientifiques majeures. L'accès aux données facilite le transfert de connaissances et engendre par là même une émulation inédite dans le champ de la recherche scientifique. Dans ces nouveaux environnements, les professionnels de l'information scientifique sont les garants d'une expertise en matière de description et de préservation des données. Cependant, ils devront se confronter à de nouvelles modalités de partage de ces données et s'impliquer dans des environnements collaboratifs tout au long du processus de recherche. Si des formations existent déjà pour ce métier déjà nommé « *e-science librarianship* », les métiers de la documentation scientifique disposent déjà dans leur histoire des compétences nécessaires pour affronter sagement ces nouvelles méthodes scientifiques.

L'étude menée tout au long de ce stage s'est surtout concentrée sur une conceptualisation des notions recouvertes par le champ e-science ainsi qu'un travail important de vulgarisation et de communication. Le site mis en place permet de toucher une communauté diverse et d'informer les professionnels de l'information scientifique et techniques de cette révolution en marche. C'est par l'étude des projets e-sciences menés dans les différents pays et par une écoute attentive des besoins de la communauté MSTIC que nous pourrons, à terme, proposer une infrastructure complète et viable.

Ce stage s'est avéré être une expérience plus qu'enrichissante pour moi. L'opportunité de mener un projet dans son intégralité m'a obligé à me fixer des objectifs et des contraintes tant temporelles que matérielles. Surtout, cela m'a permis de prendre conscience et de me confronter aux nouveaux enjeux de la documentation scientifique. En effet, les métiers des professionnels de l'information scientifique ont connu des évolutions importantes au cours de la dernière décennie. Au vu de ce nouveau paysage e-science, gageons qu'ils en connaîtront de similaires au cours des dix années à venir.

Glossaire

- **E-learning** : étymologiquement l'apprentissage par des moyens électroniques. En Français, la traduction correcte est apprentissage en ligne, mais le terme anglais est plus largement utilisé. Cela regroupe la formation aux élèves, étudiants mais aussi aux professionnels (formation à distance). Le terme se généralise dans les années 90 en Amérique du Nord et désigne alors l'enseignement à distance via l'internet. C'est au début des années 2000 que son acception se diversifie pour désigner le champ des nouvelles technologies multimédia et d'internet utilisées en classe ou à distance.
- **Folksonomie** (contraction de folk et taxonomy) : néologisme qui désigne le phénomène croissant d'indexation par l'utilisateur de ressources numériques disponibles sur un service web. L'internaute indexe la ressource à l'aide de termes qu'il aura librement choisis et autrement appelés « tags », et partage parfois cette indexation avec les autres internautes. Cela permet une organisation plus personnalisée de l'information présente sur le web mais également de faire des découvertes en fonction de ces centres d'intérêt. Exemples de sites ayant recours à la folksonomie : Wikipedia, Flickr, Delicious etc.
- **Grille informatique** (en anglais, *computing grid*) : infrastructure virtuelle constituée d'un ensemble de ressources informatiques hétérogènes et dispersées géographiquement. Elle permet un partage coordonné de ressources dans un environnement sécurisé par différents individus.
- **Mash-up** (ou **mashup** ou encore **application composite**) : application qui agrège du contenu de sources hétérogènes, et donc crée un service nouveau. Pour le cas de sites internet, on peut citer par exemple les sites de comparaison de prix ou les pages d'accueil personnalisables telles que iGoogle. Ce nouveau contenu peut être mis à disposition de l'utilisateur par le biais d'API (Application programming interface) ouvertes qui autorisent l'extraction d'informations et permettent un nouveau traitement de ces données.
- **Middleware** : couche de logiciels permettant la communication entre plusieurs applications informatiques s'effectuant sur différentes machines. Il permet à ces applications d'échanger et d'interopérer sur un réseau commun. On trouve la traduction française – quoique peu utilisée – *intergiciel*.
- **Mind-map** (ou carte heuristique en bon français) est une représentation topographique d'idées reliées entre elles par des liens sémantiques et hiérarchiques. Cela permet d'organiser

et de représenter ses idées de manière plus visuelle. Le **mind-mapping** (c.à.d la réalisation de cartes heuristiques) peut être réalisé avec tout support mais il s'est surtout répandu sur le net ces dernières années. Certains sites proposent de réaliser ses cartes en ligne (ex : MindMeister) ou sur son ordinateur personnel (ex : Freemind).

- **Workflow** : flux de travail au sein d'une organisation. Le workflow décrit les différentes tâches à accomplir entre les différents acteurs d'un processus-métier. Il s'agit d'automatiser des procédures de circulation de documents, de traitement d'information, d'exécution de tâches etc. Pour la recherche scientifique, il s'agit par exemple de l'ensemble des étapes et traitements pour une expérience spécifique.

Ressources

▪ Projets e-science cités :

[FR] Grid5000 : <https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home>

[FR] Très Grand Equipement Adonis : <http://www.tge-adonis.fr/>

[International] Réseau GEANT : <http://www.geant.net/>

[International] Protein Data Bank : <http://www wwpdb.org/>

[UK] AccessGrid : <http://www.accessgrid.org/>

[UK] e-Bank : <http://www.ukoln.ac.uk/projects/ebank-uk/>

[UK] Integrative Biology : <http://www.integrativebiology.ox.ac.uk/>

[UK] Myexperiment : <http://myexperiment.org/>

[UK] National E-Science Centre : <http://www.nesc.ac.uk/nesc/define.html>

[UK] Obesity e-lab : <https://www.nibhi.org.uk/obesityelab/default.aspx>

[US] Accès aux données publiques américaines : <http://www.data.gov/>

[US] Biolit : <http://biolit.ucsd.edu/doc/>

[US] Datastar : <http://datastar.mannlib.cornell.edu/>

[US] Digital Data Curation Center : <http://d2c2.lib.purdue.edu/>

[US] Inter-University Consortium for Political and Social Research (ICPSR) :
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

[US] iplantcollaborative : <http://www.ipiantcollaborative.org/>

[US] Nanohub : <http://nanohub.org/>

[US] National Virtual Observatory : <http://www.us-vo.org/>

[US] Topologically Integrated Geographic Encoding and Referencing system :
<http://www.census.gov/geo/www/tiger/>

[US] VIVO : <http://vivo.cornell.edu/>

▪ Ressources bibliographiques :

- [1]. AMBASSADE DE FRANCE AU ROYAUME-UNI. *Le programme e-Science au Royaume-Uni*. Ambassade de France au Royaume-Uni-Service Science et Technologie, Mars 2007. <http://www.ambafrance-uk.org/IMG/pdf/rdp-200703Rapport1.pdf>.
- [2]. ATKINS, Daniel, DROEGEMEIER, Kelvin, FELDMAN, Stuart, GARCIA-MOLINA, Hector, KLEIN, Michel, MESSERSCHMITT, David, MESSINA, Paul, OSTRICKER, Jeremiah, WRIGHT, Margaret. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. National Science Foundation, Janvier 2003. [disponible sur :] <http://www.nsf.gov/od/oci/reports/atkins.pdf>. (consulté le 23/04/2010).
- [3]. ATKINS, Daniel. *The NSF Cyberinfrastructure Initiative: Vision and Implementation Towards Learning and Discovery without Barriers*. Vidéo de la conférence donnée au MIT, 2006. [disponible sur :] <http://mitworld.mit.edu/video/434/> (consulté le 17/04/2010).
- [4]. BERNERS-LEE, Tim, HENDLER, James, LASSILA, Ora. The semantic Web. *Scientific American*, vol.284, n° 5, 2001, pp.28-37.
- [5]. BORGMAN, Christine L. The role of libraries in e-Science. (diaporama). *European Conference of Medical and Health, Helsinki*, Finlande, 27 Juin 2008. [disponible sur:] http://www.terkko.helsinki.fi/bmf/EAHILppt/Christine_L_Borgman.pdf (consulté le 14/04/2010).
- [6]. BORGMAN, Christine L. What can Studies of e-Learning Teach us about Collaboration in e-Research? Some Findings from Digital Library Studies. *Comput. Supported Coop. Work*, vol.15, n° 4, 2006, pp.359-383.
- [7]. BRANDT, Scott. Librarians as partners in e-research. *College & Research Libraries News*, vol.68, n°6, Juin 2007, pp.365-396. [disponible sur :] <http://crln.acrl.org/content/68/6/365.full.pdf+html> (consulté le 17/05/2010).
- [8]. BUETOW, Kenneth H. Cyberinfrastructure: Empowering a « Third Way » in Biomedical Research. *Science*, vol.308, n° 5723, 06 mai 2005, pp.821-824.
- [9]. COLLECTIF. 2009. *Report of the 7th e-Infrastructure Concertation Meeting e-Infrastructures for e-Science*. Bruxelles, Belgique : European Commission – DG INFSO/F3 GÉANT & e-Infrastructure Unit <http://www.beliefproject.org/docs/7th-eConcertation%20Meeting-Report-Final.pdf/view> (consulté le 27 Mai 2010).
- [10]. COMMISSION ON CYBERINFRASTRUCTURE FOR THE HUMANITIES AND SOCIAL SCIENCES. *Our cultural commonwealth: The report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences*. American Council of Learned Societies Commission, 2006, 55 p. [disponible sur :] <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf> (consulté le 23/04/2010).

- [11]. CUMMINGS, Jonathon, FINHOLT, Thomas, FOSTER, Ian, KESSELMAN, Carl, LAWRENCE Katherine A. *Beyond being there : a blueprint for advancing the design, development, and evaluation of virtual organizations*. Final report from workshops on building effective virtual organizations. National Science Foundation, Mai 2008, 88 p.
- [12]. DAVID, Paul A., BESTEN, Matthijs, SCHROEDER, Ralph. How open is e-science ?. *Second IEEE International Conference on e-Science and Grid Computing*, Amsterdam, Pays-Bas. IEEE, 2006, 33-40.
- [13]. DE ROURE, David, GOBLE, Carole, STEVENS, Robert. The design and realisation of the Experimentmy Virtual Research Environment for social sharing of workflows. *Future Generation Computer Systems*, vol.25, n° 5, 2009, pp.561-567.
- [14]. FOX, Geoffrey C., GUHA, Rajarshi, MCMULLEN, Donald F., MUSTACOGLU, Ahmet Fatih, PIERCE, Marlon E., TOPCU, Ahmet E., WILD, David J. Web 2.0 for grids and e-science. In *Grid Enabled Remote Instrumentation, Signals and Communication Technology*. Springer US, 2009. pp. 409-431 [disponible sur :] <http://grids.ucs.indiana.edu/ptliupages/publications/INGRIDFinal.pdf> (consulté le 26/02/2010).
- [15]. FREY, Jeremy G., DE ROURE, David, CARR Leslie. Publication At Source: Scientific Communication from a Publication Web to a Data Grid. *Euroweb 2002*, Oxford, UK, 2002. [disponible sur :] <http://eprints.ecs.soton.ac.uk/7852/> (consulté le 01/06/2010).
- [16]. GALLEZOT, Gabriel, LE DEUFF, Olivier. Chercheurs 2.0 ? *Les Cahiers du numérique*, vol.5, n° 2, 2009, pp.15-32. [disponible sur :] http://archivesic.ccsd.cnrs.fr/docs/00/39/62/78/PDF/Chercheur2.0-Gallezot_LeDeuff-2009.pdf. (consulté le 07/04/2010).
- [17]. GOLD, Anna. Cyberinfrastructure, Data, and Libraries, Part 1&2. *D-Lib Magazine*, vol.13, n° 9/10, Octobre 2007. [disponible sur :] [partie 1] <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html> [partie 2] <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>. (consultés le 17/05/2010).
- [18]. GUTTERIDGE, Christopher, DUKE, Monica, COLES, Simon, LYON, Liz, HEERY, Rachel, HURSTHOUSE, Mike, FREY, Jeremy, DE ROURE David. eBank UK Linking Research Data, Scholarly Communication and Learning. *Third UK e-Science Programme All Hands Meeting (AHM 2004)*, Nottingham, UK, 2004. [disponible sur :] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.1925&rep=rep1&type=pdf> (consulté le 22/02/2010).
- [19]. HALEY, Benjamin, KLIMECK, Gerhard, LUISIER, Mathieu, VASILESKA, Dragica, PAUL, Abhijeet, SHIVARAJAPURA, Swaroop, BEAUDOIN, Diane. Computational nanoelectronics research and education at nanoHUB.org. *Journal of Computational Electronics*, vol.8, n° 2, 1er juin 2009, pp.124-131.
- [20]. HEY, Tony, TREFETHEN, Anne E. Cyberinfrastructure for e-Science. *Science*, 308, n° 5723, 6 mai 2005, pp.817-821.
- [21]. HEY, Tony, TREFETHEN, Anne E. The data deluge: an e-science perspective. In *Grid computing: making the global infrastructure a reality*, Wiley and Sons, 2003, pp.809–824. [disponible sur :] http://eprints.ecs.soton.ac.uk/7648/1/The_Data_Deluge.pdf (consulté le 23/04/2010)

- [22]. HEY, Tony, HEY, Jessie. E-Science and its implications for the library community. *Library Hi Tech*, vol.24, n° 4, 2006. pp.515 – 528. [disponible sur :] http://conference.ub.uni-bielefeld.de/2006/proceedings/heyhey_final_web.pdf. (consulté le 10/05/2010).
- [23]. HUNTER, Jane. Scientific Publication Packages – A Selective Approach to the Communication and Archival of Scientific Output. *International Journal of Digital Curation*, vol. 1, n°1, 2006. [disponible sur :] <http://www.ijdc.net/index.php/ijdc/article/view/8> (consulté le 03/06/2010).
- [24]. JACOBS, James A., HUMPHREY, Charles. Preserving research data. *Communications of the ACM*, vol.47, n°9, 2004. pp.27-29. [disponible sur:] http://3stages.org/jj/w/preserving_research_data.html (consulté le 17/05/2010)
- [25]. JANKOWSKI, Nicholas W. Exploring e-Science: An Introduction. *Journal of Computer-Mediated Communication* 12, n° 2, 2007, article 10. [disponible sur :] <http://jcmc.indiana.edu/vol12/issue2/jankowski.html> (consulté le 23/04/2010)
- [26]. JEANNOT, Emmanuel, D'ANFRAY, Philippe. GRID'5000 une plate-forme d'expérimentation pour les systèmes distribués à large échelle, *JRES 2007*, Strasbourg, 2007. [disponible sur :] <http://2007.jres.org/planning/pdf/62.pdf> (consulté le 12/03/2010).
- [27]. JONES, Elisabeth, LOUGEE, Wendy, RAMBO, Neil,CELESTE Eric. *E-Science : talking points for ARL deans and directors*. Association of Research Libraries, 2008. 12p. [disponible sur :] <http://www.arl.org/bm~doc/e-science-talking-points.pdf>. (consulté le 17/05/2010).
- [28]. JOUGUELET, Suzanne. *Les Learning centres : un modèle international de bibliothèque intégrée à l'enseignement et à la recherche*. Inspection générale des bibliothèques, Décembre 2009. [disponible sur :] <http://www.enssib.fr/bibliotheque-numerique/notice-48085> (consulté le 16/03/2010).
- [29]. *Journal of Computer-Mediated Communication*, numéro spécial e-science, vol. 12, n° 2, Janvier 2007. [disponible sur :] <http://jcmc.indiana.edu/vol12/issue2/> (consulté le 23/04/2010)
- [30]. LICHNEWSKY, Alain. Les e-infrastructures en France (diaporama). Session d'information du PCN Infrastructures de recherche, 18 juin 2008. [disponible sur:] http://www.eurosfair.prd.fr/7pc/documents/1213880854_2_expose_e_infra_pcn_juin_2008_v4pdf_mode_de_compatibilite_.pdf (consulté le 23/04/2010)
- [31]. LOUGEE, Wendy et al. 2007. *Agenda for Developing E-Science in Research Libraries*. Association of Research Libraries, 2007. 26p. [disponible sur :] http://www.arl.org/bm~doc/ARL_EScience_final.pdf (consulté le 17/05/2010).
- [32]. LUCE, Richard E. A New Value Equation Challenge : The Emergence of eResearch and Roles for Research Libraries.” dans *No Brief Candle : reconceiving research libraries for the 21st century*, CLIR publication. Washington: Council on Library and Information Resources (CLIR), 2008. pp.42-50. [disponible sur :] <http://www.clir.org/pubs/reports/pub142/luce.html> (consulté le 11/05/2010).
- [33]. LYNCH, Clifford. The Institutional Challenges of Cyberinfrastructure and E-Research. *EDUCAUSE Review*, vol. 43, n° 6, 2008. pp.74-76, [disponible sur :] <http://net.educause.edu/ir/library/pdf/ERM0867.pdf> (consulté le 19/03/2010)

- [34]. LYNCH, Clifford. The Shape of the Scientific Article in The Developing Cyberinfrastructure. *CT Watch Quarterly*, vol. 3, n°3, 2007. [disponible sur :] <http://www.ctwatch.org/quarterly/articles/2007/08/the-shape-of-the-scientific-article-in-the-developing-cyberinfrastructure/> (consulté le 19/03/2010)
- [35]. MACDONALD, Stuart, URIBE, Luis Martinez. Libraries in the converging worlds of open data, e-research and web 2.0. *Online*, vol.32, n° 2, Avril 2008. pp.36-40.
- [36]. MADHAVAN, Krishna P.C., GOASGUEN, Sebastien. Integrating cutting-edge research into learning through web 2.0 and virtual environments. *International Workshop on Grid Computing Environments*. Reno, Nevada (US), 2007. [disponible sur :] <https://pubcentral.rit.edu/oajournals/index.php/gce/article/viewFile/94/55> (consulté le 23/04/2010).
- [37]. MCLENNAN, Michael, Mark LUNDSTROM, George ADAMS, et Gerhard KLIMECK. nanoHUB.org – online simulation and more materials for semiconductors and nanoelectronics in education and research. *Other Nanotechnology Publications*, 18 Août 2008. [disponible sur :] <http://docs.lib.purdue.edu/nanodocs/144> (consulté le 22/04/2010).
- [38]. MESSERSCHMITT, David G., STRAUCH, Roger A. Opportunities for Research Libraries in the NSF Cyberinfrastructure Program. *ARL: A Bimonthly Report*, n° 229, Août 2003. pp.1-7. [disponible sur :] <http://www.arl.org/resources/pubs/br/br229/br229cyber~print.shtml> (consulté le 04/02/2010).
- [39]. NELSON, Bryn. Data sharing: Empty archives. *Nature*, vol.461, n° 7261, 10 Septembre 2009. pp.160-163.
- [40]. NSF CYBERINFRASTRUCTURE COUNCIL. *Cyberinfrastructure vision for 21st century discovery*. National Science Foundation, Mars 2007. [disponible sur :] http://www.nsf.gov/od/oci/CI_Vision_March07.pdf (consulté le 23/04/2010).
- [41]. OLKEN, Franck, GEY, Fredric. Social Science Data Librarianship – A university Curriculum. 11 janvier 2006, 19p. [disponible sur :] https://hpcrd.lbl.gov/staff/olken/ssdl/iassist_ssdl_curriculum.pdf (consulté le 18/06/2010).
- [42]. PATKAR, Vivek, CHANDRA, Smita. E-Recherche : les bibliothèques numériques du futur, omniprésentes et sur une grille en open access. *World Library and Information congress : 72nd IFLA general conference and council*. 20-24 août 2006. Séoul, Corée, 2006. [disponible sur :] <http://www.ifla.queenslibrary.org/IV/ifla72/papers/140-Patkar Chandra trans-fr.pdf> (consulté le 23/04/2010).
- [43]. PIERCE, Marlon, TOPCU, Ahmet E., FOX, Geoffrey C. Web 2.0 for E-Science Environments. In *Third International Conference on Semantics Knowledge and Grid SKG 2007*, pp.1-6.
- [44]. SCHALLIER, Wouter. E-science et le rôle des bibliothèques de recherche (diaporama). *7e conférence de la DIS*, Genève, 14 Janvier 2010. [disponible sur :] http://www.unige.ch/biblio/sinformer/conferences/100114_Schallier.pdf (consulté le 23/04/2010).
- [45]. SCHROEDER, Ralph, FRY, Jenny, DE BEER, Jennifer A.. E-Research infrastructures and scientific communication. *IATUL Annual Conference Proceedings*. Stockholm, Suède,

2007 [disponible sur :] http://people.oii.ox.ac.uk/schroeder/wp-content/uploads/2008/06/iatul2007_schroeder.pdf (consulté le 01/06/2010).

[46]. SHEARER, K. et al. *Comblent les lacunes de données de recherche Examen de nouveaux services de bibliothèques*, communiqué du groupe de travail « Données de recherche-Canada », Institut canadien de l'information scientifique et technique, 2010. 16p. [disponible sur :] http://www.carl-abrc.ca/about/working_groups/pdf/library_roles-final_f.pdf (consulté le 24/05/2010).

[47]. TGE-ADONIS. *Cahier des charges de la plateforme ISIDORE*. 2009, 64 p. [disponible sur :] http://www.tge-adonis.fr/IMG/pdf_pdf_TGE-ADONIS_PlateformeISIDORE-CCTP_2009.pdf (consulté le 20/04/2010).

▪ **Pour plus d'informations, se reporter également :**

Site e-science : <http://mi2s.imag.fr/e-science/>

Page Zotero : http://www.zotero.org/emilie_manon/items

Page Delicious : http://delicious.com/mi2s_Escience

Annexes

I. Le site e-science

Page d'accueil

INFORMATIQUE - MATHÉMATIQUES - AUTOMATIQUE - SIGNAL ACCUEIL
UMS 1042

Mi²S **E-SCIENCE**

CONTEXTE ET DÉFINITIONS OBJECTIFS CHALLENGES INITIATIVES ET PROJETS E-SCIENCE ET PROS DE L'IST RESSOURCES

Accueil

Vous trouverez sur ce site toutes les informations relatives aux e-sciences (définitions, enjeux, difficultés, projets ...). Tous les projets évoqués au fil des billets sont référencés dans la rubrique [Initiatives et Projets](#).

Au-delà de sa vocation informative, ce site se veut un espace d'échange et de dialogue. N'hésitez donc pas à commenter les billets.

[Edit this entry](#)

Recherche

S'ABONNER

CATÉGORIES
Choisir une catégorie

ARTICLES RÉCENTS

- > La publication scientifique à l'ère de la e-science
- > Comment faciliter l'adoption de la e-science par la communauté scientifique ?
- > Quelques exemples de bibliothèques investies dans la e-science
- > Mind map
- > Mind map sur la e-science

DELICIOS.IJS

- > Cyberinfrastructure | Bulletin des Bibliothèques de France 07/05/2010
- > CTWatch Quarterly » November 2005 06/05/2010
- > e-Science@IFCA » National Initiatives : Spanish Network of e-Science 12/04/2010
- > E-science et le rôle des bibliothèques de recherche - Présentation de W. Schallier, directeur de LIBER 08/04/2010
- > YouTube - Chaîne de eusproject 02/04/2010

MOTS-CLEFS

"e-Science librarianship"⁷e PCRD adoption
Allemagne Asie bibliographie bibliothèque contexte
corée cyberinfrastructure cyberscience
définition data deluge données
e-labs E-Learning E-Research e-science Europe
folksonomie France Grille informatique Japon
learning centre mash-up Middleware mind map NSF
open access open science partage projets
publication réseaux sociaux
Royaume-Uni USA web 2.0 web
des données Workflow

QUI SOMMES-NOUS ?

LIENS

Mi²S

CC BY Copyright © E-science | Powered by WordPress

Designed by: Business Web Hosting | Thanks to Buy Icons, travel tips and Used Cars

Exemple d'article

INFORMATIQUE - MATHÉMATIQUES - AUTOMATIQUE - SIGNAL **ACCUEIL**
DMS 3042





CONTEXTE ET DÉFINITIONSOBJECTIFSCHALLENGESINITIATIVES ET PROJETSE-SCIENCE ET PROS DE L'ISTRESSOURCES

Zoom sur ... myExperiment

avril 2nd, 2010  [Edit](#)

my experiment Site anglais, [myExperiment](#) se présente comme une plateforme de réseau social pour chercheurs. Son objectif est de créer un environnement virtuel de recherche pour le partage, la réutilisation et le travail collaboratif sur les workflows et autres productions scientifiques. Fondé par le [JISC](#), myExperiment est le fruit d'une collaboration entre l'Université de Southampton et l'Université de Manchester, et dirigée par David De Roure et Carole Goble. Il fait partie du consortium [myGrid](#) qui a développé le [Taverna workbench](#) afin de créer et d'exécuter des workflows. Cette plateforme est également construite sur le modèle [CombeChem](#) (voir l'article [Zoom sur ... Combechem](#)). myExperiment bénéficie donc de l'expérience de deux projets pilotes du UK e-Science programme et est aussi soutenu par l'[QMI-UK](#) et l'[EPSRC](#).

Lancé en 2007, ce site a été conçu surtout à destination des chercheurs en sciences de la vie mais commence à intéresser depuis peu les domaines de la chimie, des sciences sociales et statistiques, et de la musique. Le partage de méthodes scientifiques permet aux chercheurs de gagner en temps, en visibilité et en expertise dans leur domaine et par conséquent, permet à l'ensemble de la communauté scientifique d'acquérir de nouveaux savoir-faire. myExperiment est donc fondé sur l'idée de partage et de réutilisation comme le montre ce schéma :



myExperiment cycle- Image CC : BY-SA, dullhunk@flickr

Un scientifique télécharge un workflow, l'utilise, l'enrichit de nouvelles données ou métadonnées et le réinjecte dans myExperiment en laissant son avis ou autres commentaires. C'est là le but premier du site : faire une science :

*replayable, repeatable, reproducible and reliable.**

Les services

Page de profil : chaque utilisateur dispose d'une page de profil où il peut renseigner diverses informations personnelles (âge, adresse, hobbies ...) et professionnelles (emploi actuel, travaux de recherche ...). Cela permet de trouver d'autres membres dont les centres d'intérêt sont proches et assurer ainsi de nouvelles collaborations. De plus, tout utilisateur peut créer ou rejoindre un groupe, favorisant de fait l'esprit communautaire de ce site.

Partage et tag : tout utilisateur, une fois enregistré, peut partager des workflows avec les autres membres. myExperiment permet un système de versionnage des workflows qui permet de consulter toutes les versions disponibles. En outre, il est aussi possible de partager d'autres ressources tels que des fichiers simples ou des packs. En effet, en juillet 2008, le site introduit la notion de « pack » (on retrouve aussi le terme « package ») qui permet à l'utilisateur de regrouper plusieurs ressources numériques afin de faire d'un workflow, des données brutes, d'une présentation et d'une publication par exemple un « tout » téléchargeable en une fois.

Peer-review : pour chaque objet de recherche déposé, les autres membres ont la possibilité d'attribuer une note à l'objet, de laisser des commentaires ou de soumettre un rapport plus complet de son expérience d'utilisation.

Messagerie : myExperiment propose également une messagerie instantanée.

News Feed : un peu à la manière de Facebook ou autres réseaux sociaux, la page d'accueil de l'utilisateur l'informe des actualités : nouveaux membres dans un groupe auquel il a adhéré, nouveaux inscrits, nouvel objet en partage etc.

Voici une capture d'écran d'un workflow téléchargeable. On y retrouve toutes les métadonnées nécessaires, les notes attribuées par les autres utilisateurs ainsi que leurs commentaires :

 **S'ABONNER**

CATÉGORIES

Choisir une catégorie ▼

ARTICLES RÉCENTS

- > La publication scientifique à l'ère de la e-science
- > Comment faciliter l'adoption de la e-science par la communauté scientifique ?
- > Quelques exemples de bibliothèques investies dans la e-science
- > Mind map
- > Mind map sur la e-science

DELICIOUS

- > Cyberinfrastructure | Bulletin des Bibliothèques de France 07/05/2010
- > CTWatch Quarterly » November 2005 06/05/2010
- > e-Science@FCA » National Initiatives : Spanish Network of e-Science 12/04/2010
- > E-science et le rôle des bibliothèques de recherche - Présentation de W. Schallier, directeur de LIBER 08/04/2010
- > YouTube - Chaîne de eusproject 02/04/2010

MOTS-CLEFS

"e-Science librarianship" 7e PCRD adoption

Allemagne Asie bibliographie **bibliothèque** contexte

corée **cyberinfrastructure** cyberscience

définition data deluge données

e-labs E-Learning E-Research **e-science** Europe

folksonomie France **Grille informatique** Japon

learning centre mash-up Middleware mind map NSF

open access open science partage **projets**

publication **réseaux sociaux**

Royaume-Uni USA web 2.0 web

des données Workflow

QUI SOMMES-NOUS ?

LIENS



II. Mind-map

