

serval

serveur académique lausannois

Rapport de Concept

Version 2.8

10 mai 2007

Jeannette Frey
(Cheffe de Projet)

Leire Urcelay
(Informaticienne spécialiste)

avec la collaboration de
Hubert Villard, Pascal Jacot-Guillarmod
Claude Borgeaud, Silvio Corsini, Philippe Gardel, Jacques Guélat
Isabelle de Kaenel, Pierre Keller, Danielle Mincio

Table des matières

RÉSUMÉ.....	6
1 GÉNÉRALITÉS.....	10
1.1 ÉVOLUTION DU CONTEXTE.....	10
1.2 RAPPEL DES OBJECTIFS DU PROJET.....	11
1.3 ÉVOLUTION DU MANDAT.....	12
1.4 ACTIVITÉS DE PROJET DURANT LA PHASE CONCEPT.....	13
1.5 LE CONCEPT SERVAL: UN PROJET D'INTÉGRATION.....	14
2 TYPES DE DOCUMENTS SERVAL.....	16
2.1 TYPES DE DOCUMENTS DÉPOSABLES.....	16
2.2 FORMATS ENVISAGÉS.....	16
3 MÉTADONNÉES SERVAL.....	18
3.1 MÉTADONNÉES BIBLIOGRAPHIQUES.....	18
3.1.1 <i>Métadonnées bibliographiques pour les documents scientifiques prioritaires.....</i>	<i>18</i>
3.1.2 <i>Métadonnées bibliographiques pour les autres types de documents prioritaires:</i>	
<i>MODS.....</i>	<i>33</i>
3.2 MÉTADONNÉES DE CONSERVATION.....	36
3.3 MÉTADONNÉES RELATIONNELLES.....	37
4 COLLECTIONS SUR SERVAL.....	40
4.1 COLLECTION SCIENTIFIQUE UNIL/CHUV.....	40
4.2 COLLECTION DE LA BCU.....	44
4.3 COLLECTIONS DE LA BIUM ET DU CEMCAV.....	45
5 LA CLASSIFICATION SERVAL.....	46
6 DROITS D'AUTEUR : GESTION ET CONSEIL.....	47
6.1 SITUATION JURIDIQUE ACTUELLE POUR LES PUBLICATIONS EN COURS.....	47
6.1.1 <i>La LDA suisse.....</i>	<i>47</i>
6.1.2 <i>Creative Commons.....</i>	<i>49</i>
6.2 GESTION DES DROITS D'AUTEUR DANS SERVAL.....	51
6.3 INFORMATION ET CONSEIL POUR LES AUTEURS.....	52
6.3.1 <i>SHERPA/RoMEO.....</i>	<i>52</i>
6.3.2 <i>Conseil en droit d'auteur.....</i>	<i>54</i>
6.4 SITUATION JURIDIQUE POUR LES DOCUMENTS RÉTRONUMÉRISÉS.....	54
7 PAC SERVAL.....	55
8 SERVAL ET E-ARCHIVING.....	58
8.1 ETAT ACTUEL DES EFFORTS DE E-ARCHIVING EN SUISSE.....	58
8.2 ETAT ACTUEL DU PROJET URN DE LA BN.....	59
8.3 OBJECTIFS SERVAL EN MATIÈRE DE E-ARCHIVING.....	60
9 FONCTIONNALITÉS SERVAL.....	61
9.1 SAISIE DES PUBLICATIONS.....	61
9.1.1 <i>Saisie de publications via BIB.....</i>	<i>61</i>
9.1.2 <i>Saisie de publications via FEZ.....</i>	<i>63</i>
9.2 VALIDATION DES PUBLICATIONS.....	65
9.2.1 <i>Validation via BIB.....</i>	<i>65</i>

9.2.2	Validation via FEZ.....	65
9.3	INTERFACES DE RECHERCHE.....	65
9.3.1	FEZ en tant qu'interface de recherche.....	65
9.3.2	Nouvelle application Web pour la recherche SERVAL.....	67
9.4	METADATA HARVESTING	67
10	WORKFLOWS DE DÉPÔT DE DOCUMENTS ET DE MÉTADONNÉES	69
10.1	WORKFLOW POUR LE DÉPÔT DE LA PRODUCTION SCIENTIFIQUE PAR LES AUTEURS.....	69
10.2	WORKFLOW POUR LE DÉPÔT DE LA PRODUCTION UNIL, SCIENTIFIQUE OU NON-SCIENTIFIQUE, PAR LES INSTITUTIONS PARTENAIRES.....	70
10.3	WORKFLOW POUR LA SAISIE DE E-COLLECTIONS NON UNIL PAR LES PARTENAIRES.....	71
10.4	WORKFLOW D'IMPORTATION	72
11	UTILISATEURS, AUTHENTIFICATION ET SÉCURITÉ	74
11.1	LES ADMINISTRATEURS DU DÉPÔT INSTITUTIONNEL.....	74
11.2	LES UTILISATEURS FINAUX DANS LEUR RÔLE DE LECTEURS ET CHERCHEURS D'INFORMATION ..	75
11.3	LES UTILISATEURS FINAUX DANS LEUR RÔLE DE PRODUCTEURS D'INFORMATION.....	75
11.4	LES UTILISATEURS INSTITUTIONNELS DANS LEUR RÔLE DE CHERCHEURS D'INFORMATION	76
11.5	LES UTILISATEURS INSTITUTIONNELS DANS LEUR RÔLE DE PRODUCTEURS D'INFORMATION ...	76
11.6	LES AGRÉGATEURS DE CONTENUS	76
11.7	LES MÉTA-UTILISATEURS	77
12	ARCHITECTURE	78
12.1	CHOIX TECHNOLOGIQUES	78
12.1.1	Principales caractéristiques de l'architecture FEDORA	78
12.1.2	Principales caractéristiques de l'interface FEZ.....	83
12.2	VUE D'ENSEMBLE DE L'ARCHITECTURE SERVAL.....	84
12.3	OUTILS D'ADMINISTRATION.....	86
12.3.1	FEDORA Admin Tools.....	86
12.3.2	FEZ Admin Tools.....	86
12.4	INTERACTIONS ENTRE LES DIFFÉRENTS COMPOSANTS DU SYSTÈME ET ÉVOLUTION.....	87
12.4.1	BIB <-> Dépôt institutionnel FEDORA.....	87
12.4.2	FEZ <-> Dépôt institutionnel FEDORA.....	88
12.4.3	Dépôt institutionnel FEDORA <-> Unisciences.....	88
12.4.4	Dépôt institutionnel FEDORA <-> ADIFAC.....	89
12.4.5	Dépôt institutionnel FEDORA <-> Sites Web institutionnels	90
12.4.6	Dépôt institutionnel FEDORA <-> Interface de recherche SERVAL.....	91
12.5	DONNÉES EXPOSÉES PAR LE DÉPÔT.....	91
13	MODÈLE D'OBJETS	92
13.1	STRUCTURE D'UN OBJET FEDORA	92
13.2	STRUCTURE D'UN OBJET SERVAL	93
13.2.1	Structure d'un objet de contenu.....	94
13.2.2	Structure d'un objet container	96
14	COMMUNICATION DU PROJET	98
14.1	IDENTITÉ SERVAL.....	98
14.2	COMMUNICATION EN PHASE RÉALISATION.....	98
14.2.1	Communication interne UNIL/BiUM/BCU.....	98
14.2.2	Communication externe	100
14.3	COMMUNICATION EN PHASE INTRODUCTION	100
14.3.1	Communication interne	100
14.3.2	Communication externe.....	101

14.4	COMMUNICATION EN PHASE FINALISATION	102
14.4.1	<i>Communication interne</i>	102
14.4.2	<i>Communication externe</i>	102
15	RISQUES ENCOURUS ET PROBLÈMES CONNUS.....	105
15.1	MONTÉE EN CHARGE	105
15.2	INSÉCURITÉS ACTUELLES DU E-ARCHIVING	105
15.3	MÉTHODE COLLABORATIVE DE L'ACQUISITION DES DOCUMENTS	105
15.4	SERVEURS PEU FRÉQUENTÉS - OBLIGATION DE DÉPOSER	105
15.5	FALSIFICATION DE DOCUMENTS EN FORMAT WORD	105
16	BUDGET	106
16.1	PERSONNEL	106
16.2	MATÉRIEL	107
16.3	LOGICIELS	107
17	ORGANIGRAMME.....	108
18	PLANIFICATION	109
18.1	RÉALISATION (IMPLÉMENTATION)	109
18.2	RAPPEL EN GRANDES LIGNES DES PHASES ULTÉRIEURES DU PROJET	110
18.2.1	<i>Introduction</i>	110
18.2.2	<i>Finalisation</i>	110
19	CONCLUSION.....	112
20	ANNEXES	113
20.1	FORMULAIRE DE L'ENQUÊTE SERVAL	113
20.2	DÉTAIL DE L'AUSTRALIAN STANDARD RESEARCH CLASSIFICATION DANS FEZ :	122
20.3	GLOSSAIRE	141
20.4	CONTACTS.....	165

Résumé

Le travail sur le projet SERVAL s'est intensifié durant la phase Concept (juillet 2006 – avril 2007). Une informaticienne spécialiste a été engagée au Ci en octobre 2006 pour les développements, un serveur-test a été mis en place et des premières saisies-test effectuées.

Comme les institutions participantes proposent déjà beaucoup de produits et de services en ligne, SERVAL se révèle être avant tout un projet complexe d'intégration. SERVAL sera à la fois un serveur institutionnel pour la collection des publications scientifiques UNIL/CHUV et le serveur pour la mise à disposition des collections numériques de la BCU et de la BiUM.

Du point de vue des types de documents intégrables (chap. 2), la priorité sera mise, au moment du lancement de SERVAL, sur les thèses et articles scientifiques UNIL/CHUV ainsi que sur les données des projets de numérisation de la BCU.

Les documents déposés sur SERVAL seront décrits par un ensemble très complet de métadonnées. Le groupe de travail BCU et le Ci ont élaboré des ensembles de métadonnées bibliographiques pour les types de publications prioritaires qui seront déposés sur SERVAL (chap. 3). Pour les autres types de documents, nous proposons d'utiliser le standard MODS (*Metadata Object Description Schema*). Les métadonnées bibliographiques sont complétées par des métadonnées de conservation, récupérées automatiquement dans l'en-tête des fichiers grâce au logiciel JHOVE, et par des métadonnées relationnelles que les auteurs ou les administrateurs sont en mesure de générer grâce à la souplesse du logiciel FEDORA. Les métadonnées relationnelles sont encodées en format XML en utilisant le standard RDF.

Les métadonnées et documents SERVAL seront structurés par la création de collections (chap. 4). Chaque institution participante possède son dossier principal de collection. Comme un chercheur fait toujours partie d'une unité de recherche et que nous proposons d'utiliser l'interface BIB d'Unisciences pour la saisie par les chercheurs, la collection scientifique UNIL/CHUV sera structurée de la même manière qu'Unisciences. Les bibliothèques sont libres de structurer leur collection comme elles le désirent.

Une classification (chap. 5) apparaît comme indispensable, mais ne doit pas représenter une création autonome SERVAL. Nous proposons donc de reprendre la classification existante dans FEZ, qui est l'*Australian Standard Research Classification*. Le choix de cette classification doit encore être confirmé ou infirmé par le comité de pilotage, car la CUS est engagée depuis peu dans des projets qui utilisent d'autres classifications. Une harmonisation est envisageable.

Les documents scientifiques déposés dans SERVAL sont soumis au droit d'auteur (chap. 6). Comme nombre de ces droits sont transférés à l'éditeur au moment de

la publication, SERVAL recueillera en priorité la dernière version d'un article avant publication. La plupart des éditeurs acceptent cette pratique. Pour aider les auteurs dans la clarification de leurs questions liées au droit d'auteur, SERVAL proposera des liens sur la banque de données SHERPA/RoMEO. Les auteurs devraient également pouvoir consulter un spécialiste du droit d'auteur en la personne de M. Fenter, moyennant un accord entre l'UNIL et les PPUR, pour lequel M. Fenter travaille déjà. Les projets de numérisation posent également des questions de droits d'auteur. Les responsables de ces projets doivent être bien au fait de la LDA et du Copyright et, soit choisir pour les programmes de numérisation des contenus libres de droits, soit négocier des accords avec les auteurs ou les sociétés de gestion les représentant. Les auteurs peuvent aussi intégrer une licence « Creative Commons » à leurs documents et signifier ainsi clairement ce qu'il est permis ou non de faire avec leurs oeuvres.

Tout document SERVAL sera assorti d'un code PAC (*Preservation And Conservation*), déterminant son niveau de conservation (chap. 7). Le niveau 0 sera donné aux Valdencia natifs numériques, donc en priorité à la production scientifique UNIL. Le niveau 1 sera donné notamment aux numérisations coûteuses pour protéger l'investissement, alors que le niveau 3 sera donné aux documents pouvant être remplacés. Les niveaux 2 et 4 sont réservés pour des extensions futures du PAC.

SERVAL aura un impact positif sur la question de l'archivage pérenne (chap. 8) des documents UNIL dans le sens que le projet clarifie les workflows et centralise la gestion des documents. De par ce fait, SERVAL permet d'avoir une vue globale sur les formats utilisés, et les codes PAC permettent de créer un sous-ensemble à surveiller plus précisément pour éviter les pertes. Mais SERVAL ne possède pas d'infrastructure d'archivage pérenne dans le sens d'une norme OAIS, car une telle infrastructure dépasserait largement le cadre financier alloué au projet. Les réflexions concernant l'archivage pérenne des documents natifs électroniques doivent se poursuivre au niveau régional et national.

SERVAL propose tout en lot de fonctionnalités pour la saisie, la validation et la recherche des publications (chap. 9). La saisie des publications se fera au moyen des interfaces de soumission BIB et FEZ, BIB pour la saisie des publications scientifiques par leurs auteurs et FEZ pour la saisie de publications scientifiques et e-collections par les institutions partenaires. L'interface BIB, basée sur le standard BibTeXML, est également intégrée à l'application Unisciences et contient des outils pour importer des références de publications depuis une base externe. Le DEE utilise BIB à satisfaction depuis plus d'une année. FEZ est une application Web conçue et implémentée par l'Université de Queensland pour combler le manque d'interfaces graphiques pour gérer un dépôt FEDORA. Dans FEZ, les types de documents et leurs métadonnées sont entièrement configurables par l'administrateur du dépôt. FEZ intègre par défaut les standards de métadonnées Dublin Core et MODS. La validation des documents par le chercheur se fera par l'interface BIB, la validation par la Faculté et/ou par l'administrateur SERVAL par l'interface FEZ. La recherche se fera par l'interface FEZ, ainsi que par une nouvelle application Web SERVAL. La visibilité des ressources contenues dans SERVAL sera soutenue par l'exposition de certains contenus au metadata

harvesting par OAI-PMH. Les contenus de SERVAL pourront ainsi être positionnés dans Google, ou intégrés par des agrégateurs.

Le groupe BCU, la BiUM et le Ci ont analysé les workflows (chap.10) à mettre en place pour la saisie des métadonnées et des documents SERVAL. Dans le cas des publications scientifiques UNIL/CHUV, l'auteur crée son document sur son dossier personnel MyUNIL en saisissant les métadonnées dans BIB. Le document peut ensuite transiter par plusieurs états associés à une visibilité de plus en plus large, pour être finalement validé et visible sur Internet. Les documents issus de la rétronumérisation ou faisant partie des e-collections des bibliothèques sont saisis directement par des interfaces FEZ. Des importations en masse seront possibles à partir des données d'Alexandrie (BiUM) et de RERO DOC.

Comme architecture (chap. 12) pour SERVAL, nous proposons d'utiliser FEDORA (*Flexible Extensible Digital Object and Repository Architecture*), un système fondé sur une architecture d'objets numériques extensibles et flexibles conçu conjointement par l'Université de Virginia et l'Université de Cornell. Le *Fedora Repository Core Service* offre les fonctionnalités de base du dépôt. Autour de ce service central gravitent d'autres services comme l'*OAI Provider*, l'*Ingest Service*, le *Generic Search*, le *Preservation Integrity Service* et *Preservation Monitoring Service*, l'*Event Notification*, le *Workflow Service*, ainsi que l'*ORE Access Point*.

L'architecture FEDORA ne comprenant pas d'interfaces graphiques de soumission, nous proposons l'utilisation de FEZ, un système de gestion du contenu et du workflow pour les dépôts numériques basés sur FEDORA. FEZ fournit les fonctionnalités pour la soumission, la validation et la recherche des ressources sur SERVAL. L'administrateur technique SERVAL aura également accès à des outils administratifs FEDORA.

SERVAL implique donc l'interaction entre différentes parties de système. Actuellement les informations relatives aux publications saisies dans BIB sont stockées dans une base XML. Notre proposition consiste à remplacer la base XML et le serveur de fichiers par le dépôt institutionnel FEDORA. La communication entre l'application BIB et le dépôt FEDORA se fera au moyen de webservices.

Du point de vue du modèle d'objets (chap. 13), on retrouve deux types d'objets sur SERVAL : des objets de contenu et des objets « containers ». Un objet de contenu représente une ressource numérique déposée dans SERVAL et doit stocker différents types d'informations, au minimum ses métadonnées et des données d'administration (état du workflow etc.). Les objets « container » représentent les collections et leurs hiérarchies.

Un concept de communication (chap. 14) a été établi pour SERVAL pour assurer une bonne circulation de l'information concernant le projet auprès des différents publics cible. SERVAL étant un projet d'intégration et représentant une infrastructure de base pour les institutions partenaires, on veillera en priorité à communiquer intensivement le projet aux Facultés et aux chercheurs. La communication externe passera essentiellement par le site Web SERVAL et l'annonce du projet auprès de certains organismes collectant les données sur les

serveurs institutionnels, accessoirement par des présentations aux conférences spécialisées.

Plusieurs centaines de serveurs institutionnels et thématiques ayant déjà vu le jour dans le monde, certains problèmes sont connus (chap. 15). Le problème principal est de motiver les chercheurs à déposer leur production. Nous proposons d'aller progressivement, par des réformes des règlements de Faculté, vers une obligation de déposer. Dans l'immédiat, l'intégration de l'interface BIB dans SERVAL devrait avoir un impact très favorable sur la quantité de documents déposés.

Le budget (chap. 16) alloué par la Direction UNIL sera, en 2007, investi pour la majeure partie pour la coordinatrice technique au Ci. La BCU prend à sa charge dès le mois d'octobre la création d'un nouveau poste d'administrateur des contenus SERVAL, les autres postes budgétaires étant assumés par les budgets courants du Ci, de la BCU et de la BiUM.

L'organigramme (chap. 17) intégrera dès la phase Réalisation les groupes-pilotes, c.-à-d. le DEE, le DPT, l'ISSP et l'ITB, ainsi que des validateurs des groupes de recherche et de la BCU. La phase Réalisation/Implémentation devrait s'achever comme prévu fin 2007.

Au vu de l'évolution favorable du contexte et du bon déroulement des travaux effectués jusqu'à l'heure actuelle, nous proposons au Comité de pilotage d'accepter le présent rapport de Concept et de donner le feu vert pour la phase Réalisation/implémentation du projet SERVAL.

Nous proposons à la Direction UNIL d'agender avec les Facultés, dans le courant de l'année 2007, des présentations du projet SERVAL, ainsi que des discussions concernant la modification des règlements de dépôt de thèses en faveur d'un dépôt automatique de la version électronique avec la version imprimée. La discussion pourrait également porter sur une obligation de dépôt de la version préprint des publications scientifiques UNIL sur SERVAL.

1 Généralités

1.1 Évolution du contexte

Le nombre de serveurs institutionnels et thématiques ne cesse d'augmenter dans le monde. Actuellement OpenDOAR¹ enregistre 853 serveurs, dont 6 pour la Suisse², dont le dernier en date est le serveur zurichois ZORA³ qui a été mis à disposition du public en octobre 2006. Du point de vue des contenus, les serveurs actifs en Suisse continuent leur développement de manière positive⁴.

Les serveurs institutionnels constituant une infrastructure de base pour la mise à disposition des documents scientifiques, le développement d'infrastructures nationales de mise en commun ainsi que d'infrastructures offrant des services secondaires avance à grands pas particulièrement en Angleterre⁵ et dans les pays du nord de l'Europe⁶. Le projet DRIVER⁷, lancé en juin 2006, teste la mise en commun au niveau européen des données provenant des serveurs régionaux et/ou nationaux ainsi que la mise au point de services avancés.

Les éditeurs commerciaux suivent ces développements avec plus ou moins d'inquiétudes⁸. Ils peuvent en tirer profit par le harvesting (en tant qu'agrégateurs de contenus), mais il existe tout de même une certaine crainte de voir surgir de nouveaux modèles de publication qui échapperaient totalement à leur contrôle. De tels modèles de publication sont actuellement étudiés intensivement par certains groupes, p.ex. celui autour de Herbert van de Sompel

¹ Voir sous: <http://www.opendoar.org/find.php> [vu le 10 avril 2007].

² Voir sous: <http://www.opendoar.org/countrylist.php?cContinent=Europe#Switzerland> [vu le 10 avril 2007].

³ Accès au serveur ZORA : <http://www.zora.unizh.ch/zora/> [vu le 1^{er} mai 2007].

⁴ L'EPFL vient de publier les premiers résultats du projet InfoScience : D. Aymonin, P. Crevoisier, F. Gobry, L'archive institutionnelle de l'Ecole Polytechnique Fédérale de Lausanne – Etat actuel et perspectives, en ligne sous : <http://ametist.inist.fr/document.php?id=294&format=print> [vu le 30 avril 2007].

⁵ Projets du JISC: <http://www.jisc.ac.uk/whatwedo/projects.aspx?page=0&name=All&status=Current> [vu le 10 avril 2007].

⁶ Voir p.ex. le projet SURF DARE: <http://www.surf.nl/smartsite.dws?ch=eng&id=10935> [vu le 10 avril 2007].

⁷ Projet DRIVER: http://www.driver-repository.eu/index.php?option=com_frontpage&Itemid=1

⁸ Inquiétudes énoncées lors de la récente conférence "*Scientific Publishing in the European Research Area - Access, Dissemination and Preservation in the Digital Age*" organisée par la Commission Européenne le 15-16 février 2007 à Bruxelles, par une déclaration des éditeurs STM (voir: <http://www.stm-assoc.org/brussels-declaration/> [vu le 10 avril 2007]) ; déclaration elle-même violemment contestée par les scientifiques et les bibliothécaires lors des discussions. La Communication de la Commission au Conseil, au Parlement Européen et au Comité Economique et Social Européen: http://eur-lex.europa.eu/LexUriServ/site/fr/com/2007/com2007_0056fr01.pdf

(Los Alamos National Laboratory, Research Library) et Carl Lagoze (Cornell University, Computing and Information Science)⁹ ou le projet RIOJA¹⁰.

Plusieurs études ont été menées concernant la fréquentation des serveurs institutionnels¹¹. Elles révèlent en général de grandes disparités suivant les domaines académiques. Étant donné les usages de publication très divergents, cela n'étonne guère. Il est cependant utile de relever que les serveurs institutionnels ne se peuplent pas spontanément. Pour obtenir des collections complètes, une obligation de déposer est indispensable.

1.2 Rappel des objectifs du projet

Dans le rapport d'analyse préliminaire, par rapport aux publics cibles, les objectifs du projet SERVAL ont été décrits de la manière suivante:

1. Mise en place d'un serveur permettant le dépôt et l'archivage pérennes des documents produits par les utilisateurs finaux producteurs d'information ainsi que par les institutions productrices d'information.
2. Création des outils et des interfaces permettant le dépôt des documents par les utilisateurs finaux producteurs d'information, les institutions productrices d'information, ainsi que par les administrateurs du dépôt.
3. Création des outils et des interfaces permettant la mise à disposition telle qu'elle est requise pour satisfaire les besoins des administrateurs du dépôt, des utilisateurs finaux lecteurs/chercheurs d'information, des institutions en tant que consommatrices d'information, des agrégateurs de contenus et des méta-utilisateurs.
4. Création de workflows permettant l'acquisition de toutes les publications scientifiques de l'UNIL, notamment par le dépôt obligatoire des thèses électroniques et des articles scientifiques.
5. Création d'un réseau d'administrateurs de SERVAL au sein du Ci, de la BCU, de la BiUM et dans les différentes facultés, départements et instituts de l'UNIL.

⁹ Voir Herbert van de Sompel, Carl Lagoze, Jeroen Bekaert, Xiaoming Liu, Sandy Payette, Simeon Warner : *An Interoperable Fabric for Scholarly Value Chain*, in D-Lib Magazine vol. 12 (10), octobre 2006, en ligne sous : <http://www.dlib.org/dlib/october06/vandesompel/10vandesompel.html> [vu le 25 avril 2007].

¹⁰ Site du projet d'*Overlay Journals* RIOJA : http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/rioja.aspx [vu le 25 avril 2007].

¹¹ Voir p.ex. la récente étude « *Institutional Repositories – Evaluationg the Reasons for Non-use of Cornell University's Installation of Dspace* », in D-Lib Magazine vol. 13 (3/4), mars/avril 2007, en ligne sous : <http://www.dlib.org/dlib/march07/davis/O3davis.html> [vu le 25 avril 2007].

6. Création d'un concept de *Preservation And Conservation* (PAC) pour SERVAL et réalisation du concept par le biais des accords supra-régionaux, nationaux et internationaux nécessaires.
7. Création des services annexes de conseil et d'information aux auteurs (droit d'auteur, information sur les formats de documents et métadonnées, information sur les procédures).
8. Création d'un concept de communication SERVAL.

1.3 Évolution du mandat

La phase Concept est l'une des plus importantes du projet, car elle doit fournir comme résultat une proposition de produits et de développements nécessaires à la mise en place du serveur, ainsi que l'organisation nécessaire à sa mise en œuvre, les groupes-pilotes et les workflows envisagés.

Le crédit de la Direction UNIL a permis d'engager une informaticienne pour les développements SERVAL. Depuis le mois d'octobre 2006, Leire Urcelay, informaticienne EPFL, s'attelle à cette tâche. Des serveurs-test ont été mis en place pour explorer les possibilités du produit FEDORA et tester les interfaces utilisateur disponibles, comme FEZ. Dès le mois d'octobre le Ci a contacté les utilisateurs et développeurs du produit et participé activement aux échanges de la communauté FEDORA.

Le Ci, la BiUM et la BCU ont procédé à des tests de saisie simples sur l'interface FEZ, et ont testé les possibilités de FEDORA/FEZ en termes de structuration des métadonnées et des collections. Des tests de montée en charge ont également été effectués au moyen des données des projets de numérisation de la collection précieuse de la BCU.

Le comité de projet a contacté différentes équipes de chercheurs UNIL pour participer à la réalisation de SERVAL dans le cadre de groupes-pilotes. Quatre équipes participeront : un groupe du DPT, le DEE ainsi que 2 groupes de SSP, l'ISSP et l'ITB.

Le comité de projet a également décidé de procéder à une enquête auprès des chercheurs, dans un premier temps ceux qui font partie des groupes-pilotes, en élargissant ensuite à un échantillon de chercheurs de toutes les facultés UNIL. Cette enquête a débuté fin mars et est encore en cours (formulaire d'enquête en annexe 1).

À la BCU, les projets de numérisation se poursuivent à un rythme accéléré. De grandes quantités de documents sont donc disponibles également de ce côté pour les saisies-test en phase de réalisation.

Le groupe BCU et le Ci ont approfondi le sujet des métadonnées, des interfaces de soumission et de recherche par type de documents, ainsi que des

classifications nécessaires à la description des documents sur SERVAL, en intégrant à ses réflexions les développements les plus récents en matière de standards internationaux.

1.4 Activités de projet durant la phase Concept

Le travail de projet a été très intensif durant la phase concept. En dehors des nombreuses heures passées par les collaborateurs du projet à l'élaboration des divers points du concept, les présentations et meetings suivants ont eu lieu:

5 septembre 2006	Présentation SERVAL à la CORES (BCU Riponne) (JF)
13 octobre 2006	Séance Comité de projet SERVAL (PJG, JG, HV, JF, IdK, CB, PK)
30 novembre 2006	Séance du groupe SERVAL-BCU: Métadonnées (CB, PS, OS, DM, MP, SC, SK, FR, AS, LD, PI)
18 décembre 2006	Présentation BIB et introduction à SERVAL au DPT (Département de Pharmacologie et Toxicologie, CHUV) (JG, PG, FM (Fabrice Marchon), LU)
19 décembre 2006	Séance Comité de projet SERVAL (PJG, JG, HV, JF, IdK, CB, PK)
26 janvier 2007	Journée de travail SERVAL: Poster OAI5 (JF, LU)
7 février 2007	Présentation BIB-SERVAL à M. le Vice-doyen Clémence, SSP (JG)
8 février 2007	Séance de travail pour compléter/modifier le contenu de l'enquête (IdK, JG, LU)
9 février 2007	Journée de travail SERVAL: Métadonnées (JF, LU)
12 février 2007	Présentation BIB et introduction à SERVAL à l'équipe ADIFAC, CHUV (JG, PG, LU)
13 février 2007	Séance Comité de projet SERVAL (PJG, JG, HV, JF, IdK, CB, PK)
14 février 2007	Présentation SERVAL à la DISCO (BCU Dorigny) (JF)
14 février 2007	Présentation BIB-SERVAL à Mme la Vice-doyenne Santiago, SSP (JG)
21 février 2007	Présentation BIB-SERVAL à M. Farenc, responsable informatique SSP (JG)
22 février 2007	Présentation du projet SERVAL et de notre expérience Fedora et FEZ pour des représentants de l'UNIZH, de la Bibliothèque UNIZH-Irchel et de l'ETHZ (JG, JF, PG, LU)
23 février 2007	Journée de travail SERVAL: Métadonnées, poster (JF, LU)
2 mars 2007	Journée de travail SERVAL: Identité, poster (JF, LU)
16 mars 2007	Journée de travail SERVAL: Structure d'objet, workflows (JF, LU)
19 mars 2007	Séance de lancement de l'enquête (IdK, JF, OS, PD, YB, JB, GG, MJ)
23 mars 2007	Journée de travail SERVAL: Classement, métadonnées, workflows (JF, LU)
27 mars 2007	Présentation BIB-SERVAL (JG, CB, LU, PG, JF)

30 mars 2007	Journée de travail SERVAL: Métadonnées, poster (JF, LU, CB)
3 avril 2007	Journée de travail SERVAL: Métadonnées, structure d'objets (LU, JF)
17 avril 2007	Séance Comité de projet SERVAL (PJJ, JG, HV, JF, IdK, CB, PK)
18-20 avril 2007	Conférence OAI5 au CERN: Présentation d'un poster SERVAL (LU, JF)
3 mai 2007	Présentation SERVAL et discussion FEDORA & FEZ avec des représentants de l'UNIGE et de la Bibliothèque Virtuelle de l'UNIGE (JG, JF, PG, LU)

1.5 Le concept SERVAL: Un projet d'intégration

Comparé à d'autres projets de serveurs institutionnels ou thématiques en Suisse, SERVAL se révèle être un projet à caractère d'intégration. Autant l'UNIL et le CHUV que la BCU et la BiUM partent d'un certain nombre d'acquis en matière de mise à disposition de documents scientifiques en Open Access. Les thèses lausannoises existant sous forme numérique sont mises à disposition sur Internet depuis des années par l'intermédiaire du site Cyberdocuments¹² UNIL et sur RERO DOC¹³. Les références des articles en préprint et en postprint sont saisies et mises à disposition par l'intermédiaire d'Unisciences¹⁴ et RERO DOC. La BCU et la BiUM mettent un grand nombre de documents à disposition par l'intermédiaire de diverses banques de données plus ou moins spécialisées.

Cette situation comporte de nombreux avantages :

- Un grand nombre de documents scientifiques numériques existent déjà à l'UNIL, à la BiUM et la BCU, et pourront être intégrés rapidement sur SERVAL
- Les chercheurs connaissent déjà Unisciences et l'utilisent pour créer leurs pages biographiques et leurs listes de publications
- De nombreux objets de e-learning existent sur la plateforme Moodle de l'UNIL, sur MyUNIL et sur Ariadne.
- De nombreuses photographies et vidéos numériques existent déjà dans les trois institutions partenaires et pourront enrichir SERVAL rapidement.

Cette situation de projet d'intégration implique cependant une certaine complexité :

- Harmoniser les pratiques et les outils UNIL/CHUV, BiUM et BCU utilisés actuellement

¹² <http://www2.unil.ch/cyberdocuments/> [vu le 25 avril 2007].

¹³ <http://doc.rero.ch/> [vu le 25 avril 2007].

¹⁴ Voir l'article de Axel Broquet, « Unisciences : un retour prometteur », in : i-Ci, 16,2006. en ligne sous : http://www3.unil.ch/spip/article80.html?var_recherche=uniscience [vu le 25 avril 2007].

- Harmoniser les formats des documents et les métadonnées des documents existants
- Modifier des workflows existants dans les quatre institutions
- Intégrer l'interface BIB-MyUNIL, bien acceptées par les chercheurs, au projet SERVAL

Unisciences par exemple, comporte une interface de soumission très élaborée offrant des services avancés d'importation de données fort appréciée des chercheurs. L'intégration de la partie publication d'Unisciences dans SERVAL est donc un facteur-clé de succès. Mais cette intégration demande quelques compromis au niveau des métadonnées demandées aux chercheurs (réduites au format BibTeX). La réduction des métadonnées au niveau de ce que le chercheur est habitué à saisir lui-même induit des exigences au niveau du processus de gestion: les métadonnées BIB devront être complétées par les bibliothécaires facultaires, puis par les administrateurs SERVAL afin de correspondre aux exigences décrites dans ce concept.

Au niveau des collections numériques, SERVAL est également un projet d'intégration. SERVAL proposera deux types de collections:

- La production scientifique UNIL/CHUV (serveur institutionnel)
- Les e-collections de la BCU et de la BiUM (documents scientifiques produits hors-UNIL/CHUV)
- D'autres collections, produites par les rétronumérisations, la photographie numérique BCU etc.

Toutes les collections comprendront à la fois des métadonnées renvoyant à des documents déposés sur SERVAL, et des métadonnées faisant référence à des documents se trouvant sur d'autres serveurs institutionnels ou thématiques (p.ex. pour la production scientifique d'un chercheur avant son arrivée à l'UNIL).

Pourquoi avoir différents types de collections sur SERVAL ? Parce que de grands besoins existent pour tous ces types de documents, qu'il serait extrêmement difficile dans un projet d'intégration de ne pas soutenir ces différents types et... parce que techniquement cela ne ferait aucun sens. Le produit FEDORA permet facilement de gérer plusieurs types de collections sans multiplier les serveurs.

Nous mentionnons cependant d'emblée la pluralité des collections, car elles ont des exigences profondément différentes à bien des égards. La production scientifique UNIL/CHUV ne doit jamais être confondue avec des éléments des e-collections, même lorsque ceux-ci concernent des auteurs travaillant actuellement à l'UNIL. Il sera cependant possible de les intégrer d'une manière croisée.

2 Types de documents SERVAL

2.1 Types de documents déposables

À terme, SERVAL acceptera une grande variété de documents. Dans un premier temps, l'accent sera mis sur l'import et l'acquisition de documents considérés comme prioritaires par rapport aux objectifs SERVAL. Dès les phases Réalisation et Introduction, les types de documents suivants seront ainsi systématiquement recueillis :

- a) Les thèses électroniques. Déjà disponibles sur Cyberdocuments UNIL et RERO DOC, les thèses électroniques existantes ainsi que leurs métadonnées seront récupérées au moyen d'un import. Les nouvelles thèses électroniques seront par la suite directement intégrées sur SERVAL au moyen de l'interface BIB.
- b) Les articles en préprint et postprint : dès leur entrée en fonction, les administrateurs SERVAL auront pour tâche de faire entrer toutes les métadonnées disponibles concernant les publications scientifiques UNIL/CHUV autres que les thèses, en se basant sur l'ancien catalogue et en faisant importer les données d'Unisciences. Les chercheurs des groupes-pilotes saisiront dès la phase Réalisation leurs documents sur SERVAL au moyen de l'interface BIB.
- c) Documents numériques BCU/BIUM : dès la phase Réalisation, les responsables des projets de numérisation déposeront les données générées par les programmes de numérisation sur SERVAL au moyen de l'interface FEZ.

Comme SERVAL devra être enregistré sur OpenDOAR dès qu'il sera accessible sur Internet, les politiques relatives aux types de documents et aux formats acceptés devront être déclarées publiquement.

2.2 Formats envisagés

Les formats suivants sont envisageables sur SERVAL pour les ressources mises en ligne :

Adobe (pdf), AIFF (aiff, aif, aifc), audio (au, snd), BMP, GIF, HTML (htm, html), PNG, JPEG (jpg, jpeg), LaTeX, Mathematica, Microsoft Word, MPEG (mpeg, mpg, mpe), MPEG Audio (mpa, abs, mpega), Photo CD, Photoshop (psd, pdd), Postscript (ps, eps, ai), RealAudio (ra, ram), RTF, SGML (sgm, sgml), TeX, TeX dvi, Text (txt, asc), TIFF (tiff, tif), Video Quicktime (mov, qt), WAV, WordPerfect, XML. On rajoutera à cette liste les formats courants des tableurs et des présentations: PPT et XLS.

Cependant, seul un sous-ensemble de ces formats sera effectivement autorisé au lancement de SERVAL et pour l'archivage pérenne, seul un sous-ensemble de ces formats sera défini.

Pour les documents prioritaires, nous aurons d'abord affaire à des documents MS Word et PDF (pour les thèses, les articles en préprint et postprint), ainsi qu'à des formats TIFF et JPEG pour la photographie numérique BCU, à des formats TIFF et PDF pour la numérisation BCU.

3 Métadonnées SERVAL

Les documents déposés sur SERVAL sont décrits par des métadonnées très complètes. Ces métadonnées intègrent différents processus liés aussi bien à la navigation, aux interfaces de recherche qu'à l'archivage pérenne. Il s'agit de définir, pour chaque type de document, comment celui-ci sera mis à disposition (navigation), trouvable (interface de recherche), complété par les administrateurs pour être conforme à diverses exigences (historique du document, format en cas de reformatage de masse, code PAC pour l'archivage pérenne, harvesting OAI-PMH, etc.).

Les métadonnées SERVAL se subdivisent donc en 3 catégories: les métadonnées bibliographiques, les métadonnées de conservation et les métadonnées relationnelles.

3.1 Métadonnées bibliographiques

FEDORA/FEZ possède une grande souplesse quant à la description des documents. Pour certains types de documents, il existe cependant des standards de métadonnées bibliographiques au niveau national auxquels il faut se tenir afin de garantir les workflows, en particulier avec la Bibliothèque Nationale. C'est le cas notamment des thèses.

Comme décrite au chapitre 2, une grande variété de documents pourra être déposée sur SERVAL, mais tous les types de documents n'ont pas la même priorité. Dans un premier temps, la priorité revient aux thèses et articles scientifiques. C'est donc sur ces catégories de documents que le groupe BCU s'est d'abord penché.

Une fois établie, notre politique de métadonnées sera mise en ligne, afin que SERVAL puisse être inclus dans OpenDOAR.

3.1.1 Métadonnées bibliographiques pour les documents scientifiques prioritaires

Les métadonnées bibliographiques seront saisies en premier lieu par les chercheurs eux-mêmes (pour plus de précision, voir les workflows), puis complétées ou corrigées par les bibliothécaires facultaires.

Dans l'application Unisciences, il existe depuis peu une interface BIB permettant aux chercheurs de saisir leur bibliographie. Cette interface comprend des fonctionnalités particulièrement attrayantes pour les chercheurs : saisie de leurs publications dans un format BibTeX, import rapide de leurs données bibliographiques depuis d'autres serveurs (comme PubMed). Le problème principal des serveurs institutionnels étant la difficulté d'amener les chercheurs à

déposer, l'intégration de l'interface BIB à SERVAL nous semble une excellente solution. Par l'utilisation de BIB et d'Unisciences, le chercheur dépose automatiquement son document sur SERVAL.

L'utilisation de l'interface BIB amène un certain nombre de contraintes. La première est de devoir définir en tout premier lieu le type du document, ce choix menant ensuite à l'interface de saisie paramétrée pour le type de document en question. Nos types de documents prioritaires se déclinent sur l'interface de saisie BIB de la manière suivante:

- **Thèse (doctorat)** : une thèse de doctorat.
- **Thèse (maîtrise)**: une thèse de maîtrise (mémoire de Master).
- **Article**: un article d'une revue ou d'un magazine.
- **Non publié** : un document ayant un auteur et un titre, mais non publié.
- **Rapport**: un rapport publié par une école ou une autre institution, habituellement un élément au sein d'une série
- **Livre** : un livre et son éditeur.
- **Collection (partie)**: une partie d'un livre ayant son propre titre
- **Livre (partie)**: une partie d'un livre qui peut être un chapitre, une section ou un ensemble de pages.
- **Acte de conférence**: Le compte-rendu d'une conférence.
- **Acte de conférence (partie)**: Une partie de compte-rendu d'une conférence.
- **Brochure**: un travail imprimé et relié, mais sans éditeur ou institution de financement nommé explicitement
- **Manuel**: une documentation technique
- **Autre**: à utiliser lorsque aucun autre type de document ne convient

Les sections qui suivent décrivent les champs nécessaires pour ces types de documents, en spécifiant le caractère obligatoire ou facultatif du champ et sa déclinaison dans les interfaces BIB et FEZ.

Notes :

- Pour les champs qui sont obligatoires dans BIB, leur présence peut être requise à des moments différents du processus de dépôt de document. Pour les différencier nous avons utilisé la légende suivante :
 - ¹ : requis par le standard BIBTEX et donc obligatoire pour la soumission d'une publication via BIB.
 - ² : non requis par le standard BIBTEX. Obligatoire seulement lorsque l'auteur arrive à la version définitive du document et veut le rendre visible sur le catalogue Unisciences et harvestable par les services collecteurs de métadonnées.
- **Classification** : dans BIB, seulement le premier niveau de la classification FEZ sera requis.
- **Auteurs** : dans BIB il y a un champ contenant tous les auteurs, séparés par un caractère de séparation. Dans FEZ il y aura un champ par auteur. Format du champ auteur: Nom, prénom.

- Rédacteurs : dans BIB il y a un champ contenant tous les rédacteurs, séparés par un caractère de séparation. Dans FEZ il y aura un champ par rédacteur. Format du champ rédacteur: Nom, prénom.
- Date : il n'y a pas de jours dans la date. Le format est: AAAA ou AAAA-MM
- Mots-clé : dans BIB il y a un champ contenant tous les mots-clé, séparés par un caractère de séparation. Dans FEZ il y aura un champ par mot-clé.
- Copyright : le champ Copyright de BibTeXml permet d'exprimer en texte libre l'information concernant les droits d'auteurs attachés à la publication. Ce champ est facultatif mais il aura une valeur par défaut.

Thèse doctorat

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	O	author	
directeur	I	O		
co-directeur	I	F		
langue publication	O ²	O	language	
date: -> année	O ¹	O	year	Date à laquelle la thèse est défendue.
-> mois	F	F	month	
résumé	O ²	O	abstract	
langue du résumé	I	O		
résumé langue 2	I	F		
langue 2	I	F		
résumé large public	I	F		
nombre de pages	F	F	size	Format: 230 p.
note	F	F	note	
institution	O ¹	O	school	Valeur par défaut: 'Université de Lausanne'. Cette valeur est éditable.
faculté	I	O		dropdown list. Si institution = 'Université de Lausanne' => le champ faculté contiendra par défaut la faculté de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra choisir une autre faculté de la dropdown list.
institut/unité	I	F		texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.

adresse	F	F	address	<i>les chercheurs utilisent souvent ce champ pour indiquer l'adresse du département dans laquelle la publication a été faite.</i>
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	

Thèse maîtrise

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	O	author	
directeur	I	O		
co-directeur	I	F		
langue publication	O ²	O	language	
date: -> année	O ¹	O	year	<i>Date à laquelle le mémoire est défendu.</i>
-> mois	F	F	month	
résumé	O ²	O	abstract	
langue du résumé	I	O		
résumé langue 2	I	F		
langue 2	I	F		
résumé large public	I	F		
type de mémoire	O ²	O	type	<i>dropdown list avec les types de mémoires possible.</i>
nombre de pages	F	F	size	<i>Format: 230 p.</i>
note	F	F	note	
institution	O ¹	O	school	<i>Valeur par défaut: 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O		<i>dropdown list. Si institution = 'Université de Lausanne' => le champ faculté contiendra par défaut la faculté de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra choisir une autre faculté de la dropdown list.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le</i>

				<i>spécialiste pourra modifier le contenu.</i>
adresse	F	F	address	<i>les chercheurs utilisent souvent ce champ pour indiquer l'adresse du département dans laquelle la publication a été faite.</i>
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	

Article

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	O	author	
langue publication	O ²	O	language	
date: -> année	O ¹	O	year	<i>Date d'écriture ou de publication de l'article.</i>
-> mois	F	F	month	
publié dans	O ¹	O	journal	
éditeur	I	F		<i>L'éditeur du journal dans lequel l'article a été publié.</i>
volume	F	F	volume	
numéro	F	F	number	
intervalle de pages	F	F	pages	<i>Format: 12-35 Intervalle de pages dans le journal. Ce n'est pas le nombre de pages de l'article, mais l'intervalle de pages que l'article occupe dans le journal.</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de</i>

				<i>l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>
état	O ²	O	nouveau champ	<i>Etat de publication de l'article : soumis à un journal, accepté par un journal et publié par un journal.</i>
peer-review	O ²	O	nouveau champ	<i>Flag indiquant si l'article est validée par des pairs ou pas.</i>

Document non publié

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	O	author	
langue publication	O ²	O	language	
date: -> année	O ²	O	year	<i>Date de rédaction du travail.</i>
-> mois	F	F	month	
note	O ¹	F	note	
nombre de pages	F	F	size	<i>Format: 230 p.</i>
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Rapport

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	O	author	
langue publication	O ²	O	language	
date: -> année	O ¹	O	year	<i>Date à laquelle le rapport a été finalisé.</i>
-> mois	F	F	month	
nombre de pages	F	F	size	<i>Format: 230 p.</i>
type	F	F	type	<i>Type de Rapport.</i>
isrn/strn	F	F	number	<i>ISRN: Internation Standard Technical Report Number STRN: Standard Technical Report Number</i>
note	F	F	note	
institution	O ¹	O	institution	<i>Valeur par défaut: 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O		<i>dropdown list. Si institution = 'Université de Lausanne' => le champ faculté contiendra par défaut la faculté de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra choisir une autre faculté de la dropdown list.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>
adresse	F	F (lecture)	address	<i>les chercheurs utilisent souvent ce champ pour indiquer l'adresse du département dans laquelle la publication a été faite.</i>
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	

Livre

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	Un des trois obligatoire, mais ils ne sont pas mutuellement exclusifs.	author	
rédacteurs	Mutuellement exclusifs. Un des deux est obligatoire.		Editor	
auteurs collectivité	I			
langue publication	O ²	O	language	
date: -> année	O ¹	O	year	<i>Date à laquelle le livre a été publié</i>
-> mois	F	F	month	
indicateur de la date	I	F		<i>Cet indicateur (flag) sera disponible seulement en Fez. Si nous ne connaissons pas la date de publication du e-book, la personne qui soumet la publication crochera cet indicateur pour indiquer que la date au-dessus correspond à la date de soumission dans FEZ et non la date de publication.</i>
éditeur	O ¹	O	publisher	
lieu de publication	F	O	address	
collection	F	F	series	
volume	F	F	volume	
numéro	F	F	number	
édition	F	F	edition	
e-ISBN	F	F	isbn	
nombre de pages	F	F	size	<i>Format: 230 p.</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté</i>

				<i>retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>Format: texte libre.</i>

Collection (partie)

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre de la partie	O ¹	O	title	
sous-titre	I	F		
titre du livre	O ¹	O	Booktitle	<i>Le titre du livre auquel appartient la partie référencée.</i>
auteurs	O ¹	Un des trois est obligatoire.	author	Auteur(s) de la partie référencé.
rédacteurs	F		editor	Rédacteur(s) du livre dans laquelle la partie référencée est incluse.
auteurs collectivité	I			
langue publication	O ²	O	language	
date: -> année	O	O	year ¹	<i>Date à laquelle le livre a été publié.</i>
-> mois	F	F	month	
éditeur	F	O	publisher	
lieu de publication	F	F	address	
collection	F	F	series	
volume	F	F	volume	
numéro	F	F	number	
édition	F	F	edition	
e-ISBN	F	F	isbn	
type	F	F	type	
chapitre	F	F	chapter	
intervalle de pages	F	F	pages	<i>Format: 12-35</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	

Livre (partie)

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ¹	Dans BIB	author	<i>Les auteurs ou les rédacteurs du livre.</i>

rédacteurs	Mutuellement exclusives. Un de deux obligatoire.	l'un de trois doit être remplie. Mais, les	editor	<i>Ces deux champs sont mutuellement exclusifs</i>
auteurs collectivité	I	trois aussi peuvent être remplie.		
langue publication	O ²	O	language	
date: -> année	O ²	O	year	<i>Date à laquelle le livre a été publié.</i>
-> mois	F	F	month	
chapitre	O ¹	O	chapter	
intervalle de pages	O ¹	O	pages	<i>Format: 12-35</i>
éditeur	F	F	publisher	
lieu de publication	F	F	address	
collection	F	F	series	
volume	F	F	volume	
numéro	F	F	number	
édition	F		edition	
e-ISBN	F		isbn	
type	F	F	type	
note	F		note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Acte de conférence

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
rédacteurs	O ²	1 de deux est Obligatoire.	editor	
auteurs collectivité	I			
langue publication	O ²	O	language	
date: -> année	O	O ¹	year	<i>Date à laquelle les actes de la conférence ont été publiés.</i>
-> mois	F	F	month	
collection	F	F	series	
volume	F	F	volume	
numéro	F	F	number	
organisation	F	F	organization	<i>L'organisateur de la conférence..</i>
lieu de la conférence	F	F	address	
éditeur	F	F	publisher	
nombre de pages	F	F	size	<i>Format: 230 p.</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Cette valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Acte de conférence (partie)

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
titre du compte-rendu de la conférence.	O ¹	O	booktitle	<i>Le titre du compte-rendu de la conférence à la quelle appartient la partie référencée.</i>
auteurs	O ¹	Un de trois est Obligatoire	author	
rédacteurs	F		editor	
auteurs collectivité	I			
langue publication	O ²	O	language	
date: -> année	O ¹	O	year	<i>Date à laquelle les actes de la conférence ont été publiés</i>
-> mois	F	F	month	
collection	F	F	series	
volume	F	F	volume	
numéro	F	F	number	
intervalle de pages	F	F	pages	<i>Format: 12-35</i>
organisation	F	F	organization	<i>L'organisateur de la conférence..</i>
lieu de la conférence	F	F	address	
éditeur	F	F	publisher	
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Ce valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Brochure

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ²	Un de deux est Obligatoire.	author	
auteurs collectivité	I			
langue publication	O ²	O	language	
date: -> année	O ²	O	year	<i>Date à laquelle la brochure a été finalisée.</i>
-> mois	F	F	month	
notes de publication	F	I	howpublished	<i>Remarques sur la publication, un complément ou un errata.</i>
adresse	F	F	address	
nombre de pages	F	F	size	<i>Format: 230 p.</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Ce valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Manuel

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ¹	O	title	
sous-titre	I	F		
auteurs	O ²	Un de deux est Obligatoire.	author	
auteurs collectivité	I			
langue publication	O ²	O	language	
date: -> année	O ²	O	year	<i>Date à laquelle le manuel a été finalisé.</i>
-> mois	F	F	month	
organisation	F		organization	<i>L'éditeur du rapport.</i>
adresse	F		address	
édition	F		edition	
nombre de pages	F	F	size	<i>Format: 230 p.</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Ce valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Autre

Champ	Obligatoire /Facultatif /Inexistant BIB	Obligatoire /Facultatif /Inexistant FEZ	Champ BIBTEXML	Commentaire
titre	O ²	O	title	
sous-titre	I	F		
auteurs	O ²	F	author	
langue publication	O ²	O	language	
date: -> année	O ²	O	year	<i>Date à laquelle le Document de Travail a été rédigé.</i>
-> mois	F	F	month	
notes de publication	F	I	howpublished	<i>Remarques sur la publication, un complément ou un errata.</i>
nombre de pages	F	F	size	<i>Format: 230 p.</i>
note	F	F	note	
résumé	F	F	abstract	
classification	O ²	O	category	<i>Classification FEZ.</i>
mots-clés	F	F	keywords	
copyright	F	O	copyright	
institution	I	O/F*		<i>Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut sera 'Université de Lausanne'. Ce valeur est éditable.</i>
faculté	I	O/F*		<i>Format: dropdown list. Il est obligatoire seulement si la publication fait partie de la collection UNIL. Dans ce cas, sa valeur par défaut correspondra à la faculté retrouvée dans l'annuaire LDAP.</i>
institut/unité	I	F		<i>texte libre Si institution = 'Université de Lausanne' => le champ institut/unité contiendra par défaut le nom de l'institut ou unité de la personne comme indiqué dans l'annuaire LDAP. Mais le champ sera éditable et le spécialiste pourra modifier le contenu.</i>

Pour tous les documents destinés à la publication on distinguera trois états : soumis (à un éditeur, pour le titre de périodique indiqué), accepté, publié. Tous les documents pourront également être filtrés par le critère « publication validée par des tiers ».

Dans l'état actuel de l'interface BIB, il n'est pas possible de distinguer les types de documents comme « matériel de cours », « poster » ou « PPT présenté à une conférence ». Tous ces types de documents seront donc entrés par les auteurs par l'interface BIB par le type « non publié ». En attendant que ces types

puissent également être saisis dans BIB, les bibliothécaires auront la possibilité de créer ces types de documents sur FEZ.

3.1.2 Métadonnées bibliographiques pour les autres types de documents prioritaires: MODS

Pour les autres types de documents, nous proposons de nous en tenir au standard MODS¹⁵ [Metadata Object Description Schema].

Le département "*Network Development and MARC Standards Office*" de la Bibliothèque du Congrès a développé, avec l'aide d'experts intéressés, un schéma pour un ensemble d'éléments bibliographiques qui est très utile pour la description d'objets numériques, particulièrement dans un contexte bibliothécaire. Le *Metadata Object Description Schema* (MODS), un XML Schema, est capable de soutenir autant des données existant sur MARC 21 que la création de notices de description de ressources originales. MODS contient un sous-ensemble de champs MARC et utilise des *tags* basés sur le langage courant plutôt que numériques, parfois en réunissant des éléments du format bibliographique MARC 21. MODS est exprimé en langage XML schema¹⁶ du Consortium WWW. Le standard est maintenu par le *Network Development and Marc Standards Office*¹⁷ de la Bibliothèque du Congrès.

La nouvelle version de FEZ (version 1.3) utilise le standard MODS pour décrire les métadonnées des différents types de publications. FEZ permet de configurer les éléments MODS pour chaque type de publication. À titre d'exemple, on trouvera ci-dessous des copies d'écran correspondant aux métadonnées d'un objet "audio" et d'un objet "image".

Nous pouvons tirer de MODS tous les types de documents prioritaires dont nous avons besoin, hors ceux qui sont définis dans le chap. 3.3.1. MODS permet en outre de définir soi-même des métadonnées. Il s'agit donc d'une base de travail qui n'est nullement restrictive.

Pour les personnes qui sont à l'origine des objets (les créateurs) MODS permet de définir tous les rôles dont nous avons besoin : créateur, photographe, éditeur, directeur, etc. en se basant sur les rôles définis par MARC¹⁸.

MODS est également livré avec un manuel d'utilisation complet¹⁹, donnant de nombreux exemples.

¹⁵ La page officielle MODS : <http://www.loc.gov/standards/mods/> [vu le 30 avril 2007].

¹⁶ Sur XML Schema: <http://www.w3.org/XML/Schema> [vu le 11 avril 2007].

¹⁷ Site du Network Development : <http://www.loc.gov/marc/ndmsso.html> [vu le 11 avril 2007].

¹⁸ <http://www.loc.gov/marc/relators/> [vu le 11 avril 2007].

¹⁹ Manuel d'utilisation MODS: http://dl.lib.brown.edu/documentation/MODS_Guidelines.pdf [vu le 11 avril 2007].

Exemple de métadonnées MODS pour un objet Audio

Title *	<input type="text"/>
Sub title	<input type="text"/>
Description	<div style="border: 1px solid black; height: 100px;"></div>
Author(s) 1 <i>(More Author(s) input boxes will appear as they are used up)</i>	<input type="text"/>
Recording Engineer(s) 1 <i>(More Recording Engineer(s) input boxes will appear as they are used up)</i>	<input type="text"/>
Type Of Resource	Please choose an option <input type="button" value="v"/>
Digital origin	Please choose an option <input type="button" value="v"/>
Creation date	April <input type="button" value="v"/> 05 <input type="button" value="v"/> 2007 <input type="button" value="v"/>
Recording date	April <input type="button" value="v"/> 05 <input type="button" value="v"/> 2007 <input type="button" value="v"/>
Recording place	<input type="text"/>
Digitisation date	April <input type="button" value="v"/> 05 <input type="button" value="v"/> 2007 <input type="button" value="v"/>
Edition notes	<input type="text"/>
Publisher	<input type="text"/>
Language	<input type="text"/>
Format	<input type="text"/>
Duration	<input type="text"/>
Classification	<div style="border: 1px solid gray; padding: 5px;"> <input type="button" value="v"/> </div> <input type="button" value="Add Subject"/> <input type="button" value="Remove Subject"/>
Keyword(s) 1 <i>(More Keyword(s) input boxes will appear as they are used up)</i>	<input type="text"/>
Rights	<input type="text"/>
Additional Notes	<div style="border: 1px solid black; height: 100px;"></div>

Fig. 1 Métadonnées MODS pour objet audio

3.2 Métadonnées de conservation

Les documents déposés sur SERVAL sont appelés à être conservés plus ou moins longtemps (voir chapitre 5 « PAC » et chapitre 6 « e-Archiving »). Les décisions politiques, les processus relatifs à l'intégration, l'archivage, l'accès et la préservation des documents sont souvent tributaires des données relatives au format des fichiers. Ces données doivent donc tenir compte d'un maximum de processus (workflows). SERVAL sera d'autant plus efficace s'il est capable d'autonomiser ces divers processus.

Si une petite partie de ces renseignements pourrait être saisie par l'auteur du document, la majeure partie des métadonnées de conservation sont fournies par le *header* des fichiers. Les métadonnées de conservation varient suivant les formats de documents, et à l'intérieur d'un même format, suivant la configuration des appareils qui les génèrent (p. ex. dans le cas des photographies, des scans, des vidéos etc.). Étant donné que ces métadonnées sont représentables en XML, ces variations ne posent pas de problèmes dans FEDORA.

L'acquisition des métadonnées de conservation doit de préférence se faire automatiquement, par extraction du *header* des fichiers. Nous proposons d'utiliser pour ce processus le logiciel JHOVE, qui est déjà intégré dans FEZ, et que FEDORA intégrera au cours de l'année 2007.

JHOVE est le *JSTOR/Harvard Object Validation Environment*, un projet commun de JSTOR et de la Harvard University Library²⁰. JHOVE fournit des fonctions permettant l'identification du format d'un fichier, la validation du format ou encore la caractérisation d'objets numériques. L'ensemble des caractéristiques rapportées par JHOVE correspond au concept de *representation information* introduit par le modèle de référence Open Archival Information System (OAIS)²¹²² [ISO/IEC 14721]^{23, 24}.

Le minimum d'information rapporté par JHOVE se compose du *pathname* ou URI du fichier, de la dernière date de modification, la taille en byte, le format, la version du format, le MIME type, les profils du format et de manière optionnelle les checksums (CRC32, MD5, SHA-1). Ce minimum est celui qui est rapporté

²⁰ Site JHOVE: <http://hul.harvard.edu/jhove/index.html> [vu le 12 avril 2007].

²¹ Rapport du Consultative Committee for space data systems: <http://public.ccsds.org/publications/archive/650x0b1.pdf> [vu le 12 avril 2007].

²² Toutes les activités concernant OAIS sur le site de la NASA: <http://nost.gsfc.nasa.gov/isoas/>

²³ Norme ISO: <http://www.iso.org/iso/fr/CatalogueDetailPage.CatalogueDetail?CSNUMBER=24683&ICS1=49&ICS2=140&ICS3=>

²⁴ Voir sur le site de la BNF, Catherine Lupovici, La gestion et l'entretien des données numériques (2006): <http://www.bnf.fr/pages/infopro/journeespro/pdf/entetiens06/Lupovici.pdf>

notamment lors de dépôt de fichier en format fermé (MS Word, MS Excel, MS PPT). Pour les formats ouverts, les métadonnées rapportées sont nettement plus riches. Ces métadonnées supplémentaires, dépendantes du type de média, sont consistantes avec le standard NISO Z39.87²⁵ Data Dictionary pour l'image fixe et l'esquisse AES pour le standard de métadonnées audio numérique.

Pour les tâches d'identification, validation et caractérisation, JHOVE utilise une architecture extensible de *plug-ins*. La version initiale inclut des modules pour le flux de byte arbitraire, ASCII, UTF-8, texte codé, GIF, JPEG2000, JPEG, TIFF, AIFF audio, WAVE audio, PDF, HTML et XML.

Pour les documents saisis par l'interface FEZ, les métadonnées de conservation seront donc générées directement à la saisie. Pour les documents arrivant par l'interface BIB, les métadonnées de conservation seront extraites par JHOVE au moment de leur indexation sur FEZ.

3.3 Métadonnées relationnelles

Les *Métadonnées Relationnelles* d'un objet contiennent l'information relative aux liens que cet objet possède avec d'autres ressources.

L'architecture FEDORA permet de définir des liens entre les objets (les ressources) stockés dans le dépôt. Ces liens sont qualifiés, c'est-à-dire qu'ils décrivent le type de relation établi entre deux objets.

Chaque relation est composée par les trois éléments suivants :

<Objet Sujet><Propriété de la Relation><Objet Cible>

Objet Sujet : c'est l'objet propriétaire de la relation, c'est-à-dire l'objet qui contient les *Métadonnées Relationnelles* qui décrivent la relation.

Propriété de la Relation : c'est la qualification du type de la relation. Fedora propose une Ontologie de Relations (*Fedora Relationship Ontology*²⁶) qui définit une série de types de relations de base. Cette ontologie peut être raffinée et/ou étendue selon les besoins spécifiques de chaque dépôt.

Objet Cible : c'est l'objet lié à l'objet sujet.

Par exemple :

<document X><est membre de><collection Y>
<chapitre X><fait partie de><livre Y>

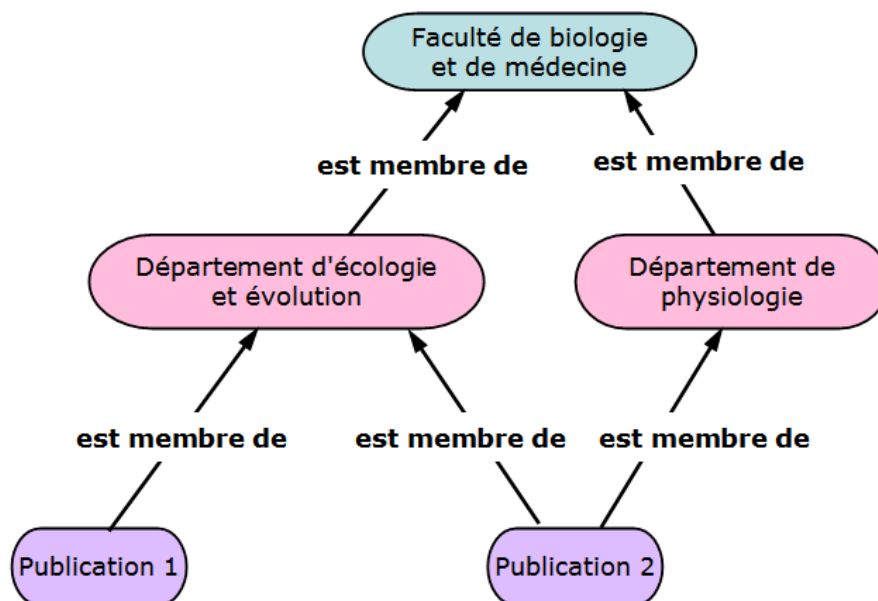
²⁵ Standard NISO Z 39.87: <http://www.niso.org/standards/resources/Z39-87-2006.pdf?CFID=52484813&CFTOKEN=18089179> [vu le 12 avril 2007].

²⁶ Voir sous : <http://www.fedora.info/definitions/1/0/fedora-relsext-ontology.rdfs>

Cette souplesse et cette richesse à l'heure d'exprimer des relations permettent d'utiliser les *Métadonnées Relationnelles* pour définir des relations structurelles ainsi que des relations sémantiques.

a) Relations structurelles :

Permettent la création d'une structure hiérarchique au sein du dépôt, comme illustré par la figure suivante :

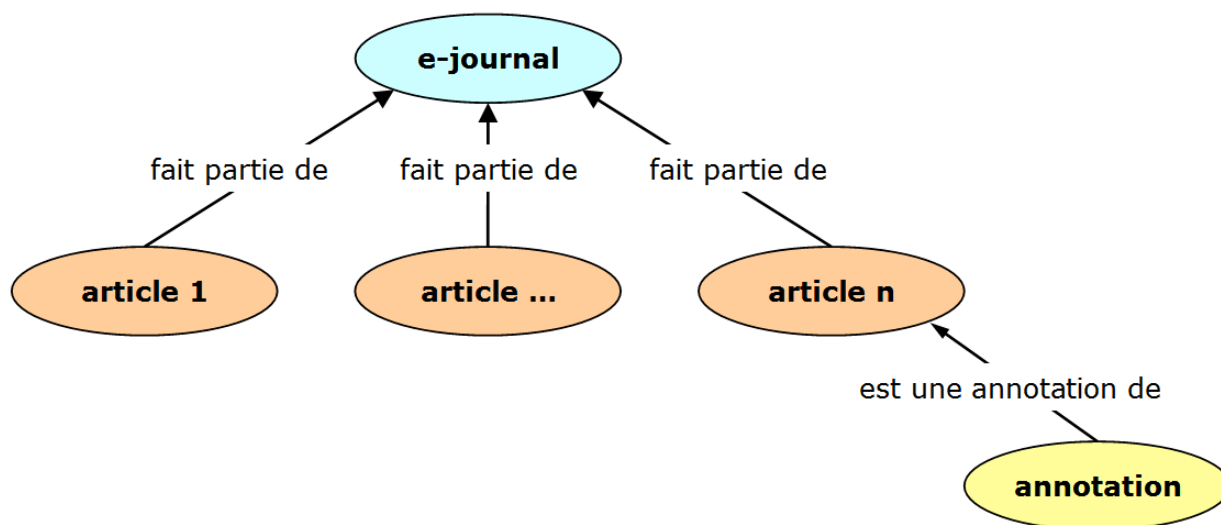


La possibilité d'organiser des publications dans des collections est très utile pour la gestion de ces objets (par exemple au niveau de l'héritage de droits d'accès) ainsi que pour faciliter la navigation et la recherche de ces objets.

b) Relations sémantiques :

Permettent la définition des relations bibliographiques entre les objets, telles que définies par les *Functional Requirements for Bibliographic Records* (FRBR)²⁷. La figure suivante montre un exemple de relations entre différents objets qui forment un e-journal :

27 Voir sous : <http://www.ifla.org/VII/s13/frbr/frbr.pdf>



La possibilité de créer un réseau d'objets où les objets sont reliés d'après des informations contextuelles est très utile. Par exemple : les citations ou les annotations collaboratives.

Comme montré dans les exemples précédents, il est possible de définir plusieurs relations et/ou types de relations pour un même objet. Ainsi, une ressource peut appartenir à plusieurs collections. Il est aussi possible de définir plusieurs relations entre deux mêmes objets.

Les *Métadonnées Relationnelles* sont stockées au niveau de l'Objet Sujet (l'objet qui porte la relation), à l'intérieur d'un *datastream* particulier nommé « RELS-EXT » (voir chapitre 13.2 pour plus de détails).

Les Métadonnées Relationnelles sont encodées en format XML en utilisant le standard *Resource Description Framework* (RDF)²⁸. Voici un exemple :

```

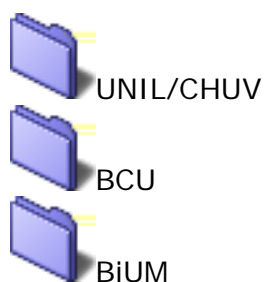
<rdf:RDF>
  <rdf:description rdf:about="info:fedora/servaldev:3855">
    <rel:isMemberOf rdf:resource="info:fedora/servaldev:3821"/>
  </rdf:description>
</rdf:RDF>
  
```

28 Voir sous : <http://www.w3.org/TR/rdf-primer/>

4 Collections sur SERVAL

Les collections constituent le premier moyen de structuration des contenus sur SERVAL. Fedora/FEZ permet actuellement de créer des arborescences de dossier à n'importe quelle profondeur. Tout en exploitant cette possibilité, on veillera à garder ces arborescences les plus simples possibles.

Comme les accès administratifs peuvent être liés à chaque dossier, il apparaît logique qu'au premier niveau, on trouve une subdivision pour chaque partenaire (*community*, voir chap. 9.3 fig, 10).



Si dans le futur d'autres partenaires collaboreraient sur SERVAL, ils obtiendraient leurs propres dossiers principaux. Chaque institution administre sa propre collection. L'administrateur d'une collection n'a un accès complet que sur sa collection, non sur les autres. Seul l'administrateur technique SERVAL au Ci a un accès complet.

En dehors de la collection UNIL/CHUV regroupant les résultats de la recherche, chaque partenaire est libre d'organiser sa propre collection comme il l'entend.

4.1 Collection scientifique UNIL/CHUV

Pour l'UNIL, l'intégration à SERVAL de la partie Publications de Unisciences requiert cependant une adéquation de la structure des collections. Comme un chercheur fait partie d'une unité de recherche et est authentifié comme tel par le LDAP UNIL, ses publications doivent se retrouver dans un dossier de cette unité. La collection UNIL est donc organisée d'après la même structure qu'Unisciences :

Unité	Groupe de Recherche
Faculté de théologie et de sciences des religions	
Section de théologie	
Section des sciences des religions	
Institut romand des sciences bibliques (IRSB)	
Départements interfacultaires	
Département interfacultaire d'histoire et de sciences des religions (DIHSR)	
Département interfacultaire d'éthique (DIE)	
Observatoire des religions en Suisse (ORS)	
Faculté de droit et des sciences criminelles	
Centre de droit comparé, de droit européen et de	

législations étrangères (CDCE)	
Institut de recherches sur le droit de la responsabilité civile et assurances (IRAL)	
Centre d'économie politique Walras-Pareto (CWP)	
Centre de droit public	Droit international public
	Droit public suisse
Centre de droit privé	Branches de culture
	Matières de droit privé
École des sciences criminelles - Police scientifique (IPS)	
Écoles des sciences criminelles - Criminologie et droit pénal (ICDP)	ICDP Internef
Faculté des lettres	
Section de philosophie	Épistémologie & philosophie des sciences
	Philosophie antique & médiévale
	Philosophie moderne et contemporaine
	Philosophie systématique
Section d'histoire	Histoire médiévale
Section de français	Français médiéval
	Linguistique française
Section d'italien	
Section d'espagnol	Président de la section
Section d'allemand	
Section d'anglais	Anglais médiéval
Section de langues et civilisations slaves	
Section de langues et civilisations orientales	
Section de linguistique	
Section d'histoire de l'art	
Section d'histoire et d'esthétique du cinéma	
Section d'informatique et méthodes mathématiques (IMM)	
Institut Benjamin Constant (IBC)	
Centre de recherches sur les lettres romandes (CRLR)	
Institut d'archéologie et des sciences de l'Antiquité (IASA)	Histoire ancienne
Institut de linguistique et des sciences du langage (ILSL)	
École de français langue étrangère (FLE)	
Cours de vacances	
Faculté des SSP	
Institut de mathématiques appliquées (IMA)	
Observatoire Science, Politique, Société (OSPS)	
Institut d'études politiques et internationales (IEPI)	
Institut d'anthropologie et de sociologie (IAS)	
Institut de sociologie des communications de masse (ISCM)	
Institut des sciences sociales et pédagogiques (ISSP)	
Institut interdisciplinaire d'étude des trajectoires biographiques (ITB)	
Institut de psychologie (IP)	
Institut des sciences du sport et de l'éducation physique (ISSEP)	
Institut d'histoire économique et sociale (IHES)	
Faculté des HEC	
Département de management (DEMA)	

Unité d'enseignement et de recherche en tourisme (UERT)	
Unité de recherche en comptabilité, contrôle et finance (URCCF)	
Département d'économétrie et d'économie politique (DEEP)	
Institut des systèmes d'information (ISI)	
Institut de banque et finance (IBF)	
Institut de macroéconomie appliquée (Créa)	
Institut de sciences actuarielles (ISA)	
Institut de recherche en management (IRM)	
Institut d'économie et management de la santé (IEMS)	
Faculté de biologie et médecine	
Centre de neurosciences psychiatriques (CNP)	
Institut Ludwig de recherche sur le cancer, Division de Lausanne	
Institut suisse de recherche expérimentale sur le cancer (ISREC)	
Département de biologie cellulaire et morphologie (DBCM)	
Département de biochimie (DB)	
Département de pharmacologie et de toxicologie (DPT)	Groupe Abriel
	Groupe Broillet
	Groupe Cotecchia
	Groupe Diviani
	Groupe Felley-Bosco
	Groupe Firsov
	Groupe Geering
	Groupe Horisberger
	Groupe Hummler
	Groupe Kellenberger
	Groupe Rossier
	Groupe Schild
Groupe Staub	
Département de physiologie (DP)	
Centre de microscopie électronique (CME)	
Institut de biotechnologie	
Département d'écologie et d'évolution (DEE)	Groupe Bernasconi
	Groupe Chapuisat
	Groupe Cherix
	Groupe Christe
	Groupe Fumagalli
	Groupe Galland
	Groupe Gigord
	Groupe Goudet
	Groupe Guisan
	Groupe Hausser
	Groupe Hirzel
	Groupe Keller
	Groupe Neet
	Groupe Perrin
Groupe Robinson-Rechavi	
Groupe Roulin	

	Groupe Salamin
	Groupe Salzburger
	Groupe Sanders
	Groupe Vogel
	Groupe Wedekind
Département de biologie moléculaire végétale (DBMV)	
Laboratoire d'analyse ultrastructurale (LAU)	
Département de microbiologie fondamentale (DMF)	
Centre intégratif de génomique (CIG)	
Institut universitaire d'histoire de la médecine et de la santé publique (IUHMSP)	
Institut universitaire de médecine légale (IUML)	Laboratoire suisse d'analyse du dopage (LAD)
Institut de médecine sociale et préventive (IMSP)	
Institut universitaire romand de santé au travail (IST)	
Institut de radiophysique appliquée (IRA)	
Policlinique médicale universitaire (PMU)	
Service d'ophtalmologie	
Centre pluridisciplinaire d'oncologie clinique (CePO)	
Division autonome de médecine préventive hospitalière	
Service des soins intensifs de chirurgie	
Service de médecine interne A	
Service de médecine interne B	
Service de cardiologie	
Service d'endocrinologie, diabétologie et métabolisme	
Service de gastro-entérologie et hépatologie	
Service d'hématologie	
Service d'angiologie	
Service d'immunologie et d'allergie (IAL)	
Service des maladies infectieuses	
Division de néphrologie	
Division de neuropsychologie	
Division de pharmacologie et toxicologie cliniques	
Division de physiopathologie clinique	
Service de pneumologie	
Division des soins intensifs de médecine	
Service de dermatologie et vénéréologie	
Service de neurologie	
Service de rhumatologie, médecine physique et réhabilitation	
Service d'anesthésiologie	
Service de chirurgie cardiovasculaire	
Service de chirurgie viscérale	
Service de chirurgie plastique, reconstructive et esthétique	
Service de neurochirurgie	
Service d'orthopédie et de traumatologie de l'appareil moteur	
Service d'urologie	
Département de gynécologie-obstétrique et génétique	
Service de génétique médicale	
Service de chirurgie pédiatrique	

Département de pédiatrie	
Service de radiodiagnostic et de radiologie interventionnelle	
Service de radiooncologie	
Service de médecine nucléaire	
Laboratoire central de chimie clinique	
Institut de microbiologie (IMU)	
Institut de pathologie (IPA)	
Service de psychiatrie générale (PGE)	
Service de psychiatrie de liaison (PLI)	
Service de psychiatrie de l'âge avancé (SUPAA)	
Service universitaire de psychiatrie de l'enfant et de l'adolescent (SUPEA)	
Faculté des géosciences et de l'environnement	
Institut de géologie et paléontologie (IGP)	
Institut de géophysique (IG)	
Institut de minéralogie et géochimie (IMG)	
Institut de géomatique et d'analyse du risque (IGAR)	
Institut de géographie (IGUL)	
Institut de politiques territoriales et d'environnement humain (IPTEH)	
Autres institutions	
Centre du droit de l'entreprise (CEDIDAC)	
Institut suisse pour l'étude de l'art (ISEA)	
Centre de traduction littéraire (CTL)	
Institut de hautes études en administration publique (IDHEAP)	
Institut suisse de bioinformatique	
Musée de géologie	

4.2 Collection de la BCU

Les bibliothèques partenaires sont libres de structurer leur collection comme elles le désirent. Il sera utile de faire quelques expériences durant la phase Réalisation. L'ensemble de la collection BCU est supervisée par l'administrateur SERVAL BCU.

Pour la BCU, les collections suivantes sont envisagées:



e-collection



Rétronumérisation d'imprimés



Rétronumérisation Manuscrits



Photographie

La e-collection contiendra des métadonnées et parfois des documents récoltés (ou achetés) par les bibliothécaires universitaires pour la collection de la BCU. De plus en plus souvent, les éditeurs commerciaux proposent l'achat de livres électroniques. Ces documents peuvent être téléchargés sur un serveur local et pourront être introduits dans une collection e-book SERVAL²⁹. La e-collection sera probablement subdivisée en domaines académiques tels qu'ils sont représentés au sein de la BCU, afin de permettre de générer des listes de ressources numériques dans le site Web BCU pour chaque domaine. Ce dossier est géré en central par l'administrateur SERVAL BCU. Les sous-dossiers sont gérés par les bibliothécaires universitaires.

La numérisation Google contiendra les PDF des livres de la collection imprimée BCU numérisés par le biais des projets de numérisation Google et 4DigitalBooks de la BCU. Ce dossier sera géré par Silvio Corsini, responsable de la numérisation.

La numérisation des manuscrits contiendra les fichiers relatifs à la collection des manuscrits. Ce dossier sera géré par Danielle Mincio, responsable de la section des manuscrits.

La photographie contiendra la collection de photographies numériques de la BCU. Ce dossier sera géré par Laurent Dubois, photographe de la BCU.

D'autres dossiers pourront être créés, p.ex. pour des programmes de numérisation de la Riponne (Valdensia, musique etc.), la vidéo, le son etc.

4.3 Collections de la BiUM et du Cemcav

Comme la BCU, la BiUM pourra structurer une collection comme en fonction de ces besoins et des documents à déposer.

²⁹ Pour le PAC, il faudra examiner quel est leur statut. C'est bien sûr l'éditeur qui est en charge de l'archivage pérenne du document mais la BCU devra s'en préoccuper à un autre titre. Il s'agit d'un document qui a été acheté et fait partie de la collection BCU au même titre qu'un ouvrage imprimé. A ce titre, on doit prendre en considération son archivage.

5 La classification SERVAL

Une classification est un élément indispensable à un serveur académique, pour différentes raisons: c'est un moyen de retrouver des documents d'un même domaine de recherche. Toute classification implique une catégorisation et une hiérarchisation des connaissances. C'est une démarche nécessairement arbitraire, toujours discutable. Une classification sommaire est vite sans intérêt pour l'utilisateur, une classification trop élaborée inutilisable pour l'auteur sensé attribuer un ou plusieurs éléments de classification à son document.

Dans les projets européens présentés récemment, l'idée d'un classement possédant une arborescence d'une profondeur de trois niveaux revient souvent (notamment dans les projets allemands). Cette profondeur est encore simple à comprendre, tout en permettant un degré de finesse suffisamment poussé pour être intéressant. Une telle classification de profondeur 3 existe déjà dans FEZ. Il s'agit de l'*Australian Standard Research Classification* (voir annexe 3), qui a l'avantage de générer un arbre assez équilibré (contrairement à la CDU ou d'autres classifications purement bibliothécaires). Nous proposons donc de l'utiliser.

Notons qu'il n'est pas indispensable d'utiliser les 3 niveaux de classification: un chercheur peut n'indiquer que le niveau le plus général. Du point de vue du workflow, il est indispensable que le chercheur rentre lui-même cette information, car il n'est pas toujours possible à l'administrateur facultaire de la compléter. C'est donc un champ obligatoire à remplir dans BIB. Un document peut posséder plusieurs éléments de classification. L'administrateur peut compléter les données rentrées par le chercheur par de nouvelles entrées, mais ne devrait pas modifier celles qui ont été créées directement par le chercheur (sinon ce dernier ne retrouvera pas son document dans l'arborescence qu'il a choisie lui-même). Précisons encore que l'auteur aura la possibilité dans un autre champ d'ajouter des mots-clés pour préciser son domaine de recherche. La classification est obligatoire, mais les mots clés sont facultatifs.

Cette classification offre l'avantage de couvrir tous les domaines académiques. Elle peut donc être appliquée autant à la collection scientifique UNIL qu'aux e-collections de la BCU et de la BiUM (qui peut couvrir des domaines qui ne sont pas ou plus enseignés à l'UNIL). Elle a l'inconvénient de ne pas être de culture française, et encore moins suisse. Après discussion, le comité de projet conclut cependant qu'il vaut mieux garder la classification la plus standardisée possible. Des modifications ne sont envisagées que dans des cas très précis (p. ex. pour le Droit Suisse). Afin d'éviter l'affichage d'éléments de classification vides, on peut envisager de n'afficher que ceux qui contiennent des documents.

Le choix de cette classification doit encore être confirmé ou infirmé par le comité de pilotage, car la CUS est engagée depuis peu dans des projets qui utilisent d'autres classifications. Une harmonisation est envisageable.

6 Droits d'auteur : Gestion et conseil

6.1 Situation juridique actuelle pour les publications en cours

Il existe dans le monde deux systèmes de droits d'auteur assez différents : le système européen de droit d'auteur, et le système étatsunien du copyright.

La situation en Europe, a été formulée très récemment de la manière suivante par la Commission des Communautés Européennes:

"Au moment de publier leurs articles dans des revues, les auteurs transfèrent normalement leurs droits aux éditeurs. Les éditeurs assurent ensuite la diffusion de ces articles afin d'obtenir un retour sur leurs investissements. Les chercheurs, les organismes de financement et les bibliothèques ont suggéré que les pratiques contractuelles actuelles puissent avoir un impact négatif sur l'accès et la diffusion et qu'elles méritent une réflexion".

...

Dans le cas des données brutes, la question des droits de propriété intellectuelle (DPI) est différente. Alors que les données brutes ne sont pas protégées en tant que telles par les DPI, la directive européenne 96/9/EC sur la protection juridique des bases de données protège les efforts dédiés à l'organisation des données. Dans ce cadre, des préoccupations ont été exprimées sur l'impact de la protection sui generis pour les bases de données non originales sur l'accessibilité des données de la recherche scientifique.

Si des données personnelles sont traitées ou diffusées dans ce contexte, l'accès à ces données et leur utilisation doivent respecter les règles sur la protection des données personnelles détaillées dans les directives de l'UE 95/46/EC et 2002/58/EC."

[Communication de la Commission au Conseil, au Parlement Européen et au Comité Economique et Social Européen COM(2007)56, février 2007]

6.1.1 La LDA suisse

Le droit Suisse est proche du droit européen en la matière et est actuellement sujet à révision. Un avis d'expertise a été demandé fin 2006 par le Consortium des bibliothèques universitaires suisses à M. Willi Egloff, avocat et docteur en

droit à Berne. Ce document n'est malheureusement pas disponible en libre accès, mais peut être obtenu auprès de l'auteur³⁰.

En Suisse, toute création de l'esprit qui a un caractère individuel est a priori une œuvre pouvant être protégée (art. 2 LDA), indépendamment de sa valeur ou de sa qualité esthétique, et indépendamment de la forme que revêt la publication (imprimée ou électronique). La LDA protège également les « projets », qu'il s'agisse d'une esquisse provisoire, d'un premier jet ou d'une version définitive. Un préprint est donc tout aussi protégé par le droit d'auteur qu'une version publiée.

Selon le droit suisse, seule la personne physique qui a créé l'œuvre peut être auteur (jamais une institution). Si plusieurs personnes ont contribué à créer l'œuvre, le droit d'auteur leur appartient en commun. Ce droit est cessible.

En cas de cession, le droit d'auteur est transféré à une autre partie de manière plus ou moins intégrale. En effet, au moment de la publication, l'auteur abandonne une partie de ses droits à l'éditeur selon les termes définis par le contrat d'édition. L'auteur garde cependant toujours les prérogatives d'ordre moral liées à sa qualité d'auteur (droit de reconnaissance de la qualité d'auteur, protection de l'intégrité de l'œuvre, art. 9 et art. 11, 2^e al. LDA).

La loi prévoit diverses licences légales pour des utilisations d'œuvres bien définies (art. 19 LDA) : l'usage privé, à des fins pédagogiques, et la reproduction d'exemplaires à des fins d'information interne ou de documentation. Cette disposition s'applique également aux bibliothèques, qui ont donc le droit de reproduire des exemplaires d'œuvres à ces fins, car elles payent pour cela une redevance à la société de gestion ProLitteris³¹. Les bibliothèques peuvent également faire des exemplaires pour l'usage privé de tiers, si ceux-ci détiennent eux-mêmes le droit de reproduction en vertu de l'art 19 LDA (c.-à-d. si ces tiers utilisent le document dans un dessein privé, pédagogique ou de documentation interne). Ces reproductions n'ont jamais le droit d'être à but commercial et ne doivent également en aucun cas avoir le caractère d'un produit de substitution de même valeur que l'exemplaire proposé sur le marché. De fait, d'après l'art. 19 LDA, les bibliothèques sont donc en droit de numériser des œuvres protégées par le droit d'auteur. Elles sont également en droit d'archiver ces copies numériques « pour assurer la conservation de l'œuvre » (art. 24 LDA). Mais si les bibliothèques souhaitent mettre des contenus à disposition sur Internet, elles ont besoin d'une autorisation des ayants droits sur les contenus.

³⁰ Willi Egloff, « *Questions de droit d'auteur en rapport avec les services en ligne des bibliothèques universitaires et avec le libre accès* ». Avis d'expertise élaboré sur mandat du Consortium des bibliothèques universitaires suisses, Zurich, 13 octobre 2006.

³¹ Ces questions sont actuellement en discussion également au niveau européen. Voir « European Digital Library Initiative, High Level Expert Group – Copyright Subgroup, *Report on Digital Preservation, Orphan Works and Out-of-Print Works, Selected Implementation Issues* (approved on 18.4.2007). Ce rapport contient notamment un modèle de contrat de licence de numérisation de livres épuisés.

Les ayants droits sont les auteurs eux-mêmes dans les cas des préprint, les éditeurs dans le cas de la version de l'article telle qu'elle est parue dans une revue, l'auteur ayant remis ses droits à l'éditeur par un contrat de cessation, le contrat d'édition. Ces contrats peuvent différer beaucoup d'un éditeur à l'autre, ou même d'un titre à un autre chez un même éditeur.

Les auteurs ne pouvant céder que les droits qu'ils possèdent encore, les conditions de ces contrats peuvent bien entendu interférer avec le droit de mettre un texte dans sa forme définitive (publiée) à disposition du grand public sur SERVAL. La cessation des droits sur la version publiée n'implique cependant pas celles sur les versions préprint (sauf convention explicite le stipulant). Les chercheurs sont donc parfaitement habilités à mettre à disposition les préprints de leurs travaux sur SERVAL.

Pour les versions postprint, de nombreux éditeurs autorisent la mise à disposition des articles sur le serveur institutionnel ainsi que sur le site privé de l'auteur, l'autorisation pouvant être assortie de l'obligation de mentionner le titre dans lequel l'article a été publié et d'y renvoyer à l'aide d'un lien. Si une telle autorisation n'existe pas, la mise à disposition du postprint sur SERVAL n'est pas permise. Ce type de clause peut également contenir un embargo (6 mois, 12 mois, 24 mois, etc.).

On peut envisager que les droits d'auteur (ou partie des droits) soient transférés à une institution par le règlement d'études ou d'examens, les conditions d'engagement, etc. Cela fait sens dans le cas des règlements de bachelor, de master, de doctorat et d'habilitation, car dans ce cas les étudiants doivent de toute façon remettre à la BCU ou à l'UNIL un certain nombre d'exemplaires de leurs travaux. Une telle obligation de remise doit cependant avoir une base légale, donc être inscrite soit dans la loi d'organisation de l'UNIL, soit dans les règlements d'études ou d'examens ou dans les conditions d'engagement.

6.1.2 Creative Commons

Une autre approche, présentée récemment aux auteurs du domaine des Lettres au colloque de l'ASSHS déjà cité³², est celle de « Creative Commons³³ ». Comme elle se base sur le système étatsunien du Copyright, elle est d'une pratique assez différente du droit suisse et européen, car elle vise l'exercice de droits individuels. Ces droits peuvent cependant être exercés tout en respectant le droit suisse et il est donc possible aux auteurs UNIL de munir leurs documents de licences « Creative Commons », dont voici les principes.

³² Les présentations, notamment celle de Michel Jaccard sur des aspects juridiques, sont disponibles sur le site de l'ASSHS : <http://www.sagw.ch/frz/index.asp?pag=Publikationen>, sous « liste des publications », puis « Open Access – du principe à la mise en pratique ». [vu le 25 avril 2007].

³³ Pour la version suisse de Creative Commons, voir le site : <http://www.creativecommons.ch/> [vu le 29 avril 2007.]

Pour protéger une œuvre sur la base d'une licence « Creative Commons », il faut le mentionner expressément pour chaque œuvre, ce qui peut être fait au moyen de boutons constitués par la combinaison d'une ou plusieurs des icônes suivantes :



Attribution : Vous laissez les autres copier, distribuer, montrer et utiliser votre œuvre – et les œuvres dérivées basées sur la vôtre – mais seulement en vous citant de la manière requise.



Non commercial. Vous laissez les autres copier, distribuer, montrer et utiliser votre œuvre – et les œuvres dérivées basées sur la vôtre – mais uniquement dans des buts non commerciaux.



Pas d'œuvres dérivées. Vous laissez les autres copier, distribuer, montrer et utiliser uniquement des copies conformes de votre œuvre. Il n'est pas permis d'en créer des œuvres dérivées.



Partager dans les mêmes conditions. Vous permettez à d'autres de distribuer des œuvres dérivées de la vôtre, mais uniquement sous des conditions de licence identiques aux vôtres.

Les six licences de base de Creative Commons sont :



Cette licence est la plus restrictive, la licence « publicité gratuite ». Elle permet aux autres de télécharger l'œuvre et de la partager avec des tiers, pour autant qu'ils vous citent correctement et mentionnent correctement votre lien ; mais aucune modification de l'œuvre n'est permise, ni aucune utilisation commerciale.



Cette licence permet aux autres de modifier votre œuvre, d'en faire des travaux dérivés, pour autant que ces œuvres ne sont pas commerciales et que vous êtes cités correctement en tant qu'auteur de l'œuvre de base et que les œuvres dérivées soient soumises aux mêmes conditions de licence que la vôtre. Toute nouvelle œuvre basée sur la vôtre aura les mêmes conditions de licence, donc sera non-commerciale par nature.



Cette licence permet aux autres de modifier votre œuvre et d'en faire des œuvres dérivées non-commerciales. Bien que ces œuvres doivent vous citer correctement et être non-commerciales, elles ne doivent pas obligatoirement comprendre des termes de licence identiques.



Cette licence permet de redistribuer, commercialement ou non-commercialement, pour autant que l'œuvre reste inchangée dans son ensemble et que vous êtes cité correctement.



Cette licence permet aux autres de modifier l'œuvre et produire des œuvres dérivées, même à des fins commerciales, pour autant que vous êtes cité correctement et que la nouvelle œuvre est soumise aux mêmes conditions de licence. Cette licence est souvent comparée aux licences des logiciels Open Source. Toutes les nouvelles œuvres basées sur la vôtre seront soumises à la même licence, donc tous les dérivés pourront également être commercialisés.



Cette licence permet aux autres de distribuer et de modifier l'œuvre, d'en produire des œuvres dérivées, même commerciales, pour autant que vous soyez crédité correctement comme étant l'auteur de la création originale. C'est la licence qui accorde le plus de droits aux autres.

Rappelons enfin que si ces problèmes de droits d'auteur ne sont pas tous solubles définitivement à l'heure actuelle, ils ne sont pas non plus propres à SERVAL. Toutes les modalités de diffusion de l'information sont aujourd'hui renégociées entre éditeurs commerciaux et partisans de l'Open Access. L'essentiel est de suivre cette problématique au plus près afin d'ajuster au fur et à mesure les conditions de dépôt dans SERVAL.

6.2 Gestion des droits d'auteur dans SERVAL

Un premier point important dans la gestion des droits d'auteur est d'obtenir de l'auteur le droit d'archiver et de mettre à disposition (sur Intranet ou Internet) son document. Pour obtenir ce droit, l'auteur doit au minimum être instruit des suites de son action de déposer (c.-à-d. certifier avoir pris connaissance des conditions de dépôt) et donner son accord au dépôt en cliquant sur un bouton ("je veux publier ce document sur SERVAL").

L'interface FEZ contient un message en anglais que nous nous proposons de traduire comme suit en français:

Pour les œuvres déposées par leur propre auteur: en auto-archivant cette collection de fichiers, je donne le droit à SERVAL de la stocker et de la rendre accessible en ligne gratuitement et publiquement de manière permanente. Je déclare que ce matériel est ma propre propriété intellectuelle. Je comprend que SERVAL n'assume aucune responsabilité s'il y a violation de droits d'auteur par

la distribution de ces fichiers. (Tous les auteurs sont vivement encouragés à affirmer clairement leur droits d'auteur sur la page titre de leur ouvrage).

Pour des ouvrages déposés par quelqu'un d'autre que l'auteur: je déclare par la présente que la collection de fichier que j'archive dans SERVAL est du domaine public. Si tel n'est pas le cas, je porte l'entière responsabilité pour toute lésion du droit d'auteur que la distribution de ces fichiers pourrait engendrer.

En cochant cette case et en soumettant ce formulaire, vous acceptez les termes de ce contrat.

Selon les discussions qui ont eu lieu le 1^{er} mars 2007 à Berne lors du colloque de l'ASSHS³⁴ "*Open Access - Du principe à la mise en pratique*"³⁵, cette manière de procéder laisse encore un certain flou juridique. Il n'est pas entièrement certain que les éditeurs ne pourraient en aucun cas se retourner contre les propriétaires du serveur, mais jusqu'à ce jour, il n'y a du moins pas d'exemples d'une telle démarche.

Normalement, la version déposée en tant que postprint correspond à la dernière version de l'auteur, c.-à-d. celle qu'il a envoyé à l'éditeur pour impression. La grande majorité des éditeurs n'autorise pas le dépôt sur serveur institutionnel du PDF de l'article dans sa version éditée.

6.3 Information et conseil pour les auteurs

Reste que très souvent, les auteurs sont fort mal renseignés sur les conditions juridiques dans lesquelles ils publient, parfois même lorsqu'ils le font depuis longtemps. Comme cette ignorance pourrait être à tort un obstacle pour déposer un document dans la partie Internet de SERVAL, il nous semble approprié d'offrir aux auteurs des opportunités de s'informer sur leurs droits. Cette information sera dispensée par deux canaux différents :

- Par un lien sur la base de donnée SHERPA
- Par un lien sur les coordonnées du conseiller en publication.

6.3.1 SHERPA/RoMEO

Le projet SHERPA/RoMEO (Right Metadata for Open archiving) est achevé. Il était financé par le JISC (2002/3) pour investiguer les problèmes juridiques liés au dépôt sur des serveurs institutionnels en Angleterre. SHERPA/RoMEO propose

³⁴ Site de l'ASSHS : <http://www.sagw.ch/frz/index.asp> [vu le 25 avril 2007].

³⁵ Les présentations, notamment celle de Michel Jaccard sur des aspects juridiques, sont disponibles sur le site de l'ASSHS : <http://www.sagw.ch/frz/index.asp?pag=Publikationen>, sous « liste des publications », puis « Open Access – du principe à la mise en pratique ». [vu le 25 avril 2007].

actuellement une banque de données avec une interface de recherche en ligne listant les conditions des éditeurs face à l'auto-archivage.

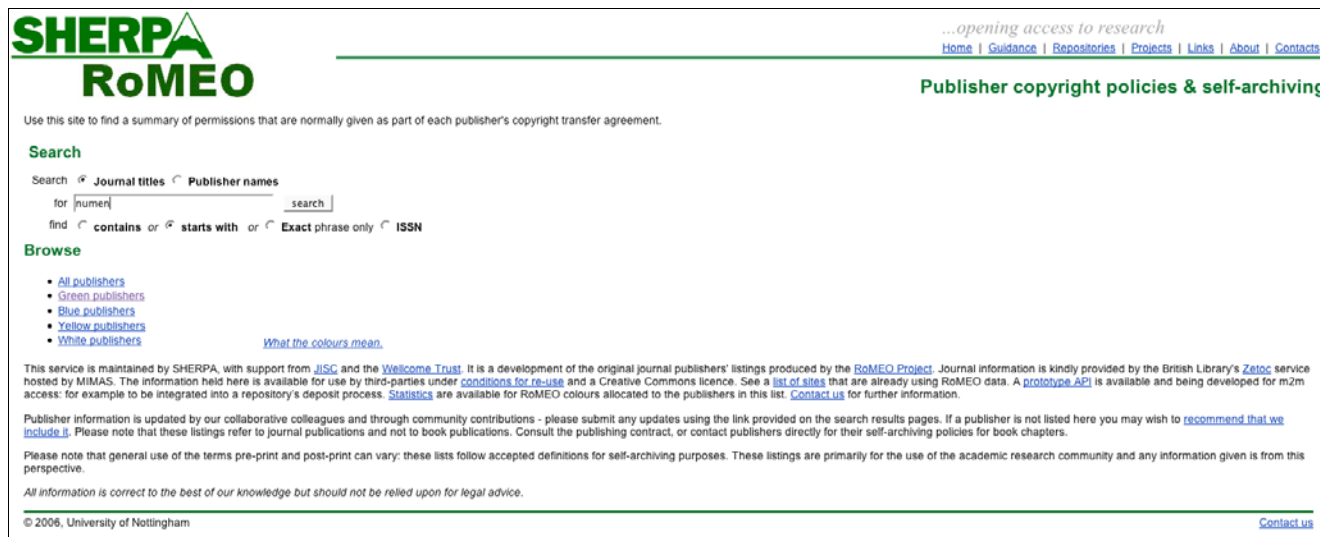


Fig. 3 L'interface de recherche de SHERPA/RoMEO

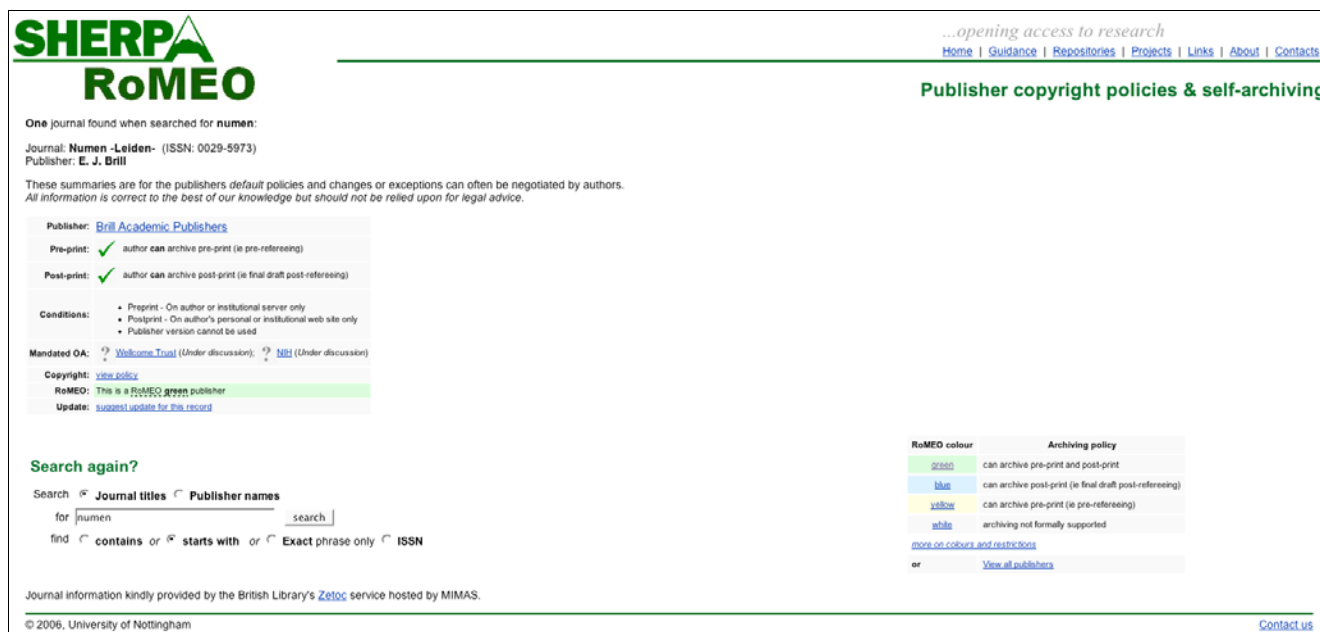


Fig. 4 Résultat sur une recherche dans SHERPA/RoMEO. Les éditeurs sont classés dans quatre catégories : les verts qui permettent autant l'archivage du préprint que du postprint, les bleus qui permettent d'archiver la dernière version peer-reviewed (mais pas le postprint formaté), les jaunes qui ne permettent l'archivage que du pré-print avant correction par le comité de lecture, et les blancs, qui n'autorisent pas formellement l'archivage.

Une équipe allemande a présenté une interface SHERPA/RoMEO en allemand. Ils nous ont proposé, si nous le désirons, de mettre le code à disposition pour créer facilement une interface en français. Des discussions sont en cours avec l'EPFL

pour examiner la possibilité de compléter cette interface en y ajoutant des revues locales non contenues dans SHERPA.

6.3.2 Conseil en droit d'auteur

En ce qui concerne le conseil en droit d'auteur, l'EPFL possède un conseiller en la personne de M. Fred Fenter. M. Fenter a travaillé dans l'édition et est actuellement en charge des éditions scientifiques de l'EPFL (PPUR). Comme le recours à ce service de conseil n'occupera pas une personne à un pourcentage élevé, nous proposons d'avoir recours au service de M. Fenter dans un ordre d'env. 0.05 ETP et enjoignons l'UNIL et l'EPFL de passer un accord dans ce sens dès début 2008.

En outre, le site CODUL³⁶ oriente d'ores et déjà le chercheur sur les sources d'information concernant les modalités de la diffusion en Open Access.

6.4 Situation juridique pour les documents rétronumérisés

Les projets de numérisation de la BCU et de la BiUM posent également des questions de droits d'auteur. En Suisse, une œuvre est protégée jusqu'à 70 ans après le décès de l'auteur. Toute numérisation dans un but autre que celui de créer une copie de sécurité est donc sujette à une autorisation par l'auteur lui-même ou par ses héritiers. 70 ans après la mort de l'auteur, l'ouvrage passe dans le domaine public et peut donc être numérisé et mis à disposition sur Internet sans autre.

Comme on l'a vu dans l'introduction de ce chapitre, dans le cas des livres, c'est en principe l'éditeur qui possède les droits, ce qui complique la question. Comme on l'a vu aussi, la bibliothèque a le droit de numériser pour son usage privé, et pour l'usage privé de tiers, ainsi que pour des raisons d'archivage.

On peut donc tirer comme enseignement qu'un ouvrage qui est protégé par le droit d'auteur peut être numérisé et mis à disposition sur Intranet, mais que, sans accord du détenteur des droits, il ne pourra pas être mis à disposition sur Internet.

Tous les ouvrages libres de droit (dont les auteurs sont décédés depuis plus de 70 ans) peuvent cependant être numérisés et mis à disposition sur Internet.

Dans la pratique, une même œuvre pouvant être rééditée, il est indispensable que les responsables des projets de numérisation soient bien au courant des droits d'auteur et choisissent les ouvrages à numériser en conséquence. En cas de numérisation d'un ouvrage sous droit d'auteur et en l'absence de contrat avec le détenteur de droit, le fichier peut être déposé sur SERVAL muni d'un embargo (indéfini, ou 70 ans) et ne sera visible sur Internet qu'après la fin de l'embargo.

³⁶ Site CODUL : <http://www.unil.ch/codul/page32008.html>

7 PAC SERVAL

Le PAC (code *Preservation And Conservation*) des collections imprimées BCU contient 5 niveaux de conservation. Pour SERVAL, nous sommes donc partis de la base d'un code PAC qui contiendrait également 5 niveaux. Le groupe PACRERO a récemment travaillé sur l'établissement d'un code PAC pour les documents électroniques. Comme nous aimerions collaborer avec RERO pour assurer la base du e-archiving SERVAL, il est utile de respecter le même schéma. Appliqué à SERVAL, on obtient le PAC suivant :

NIVEAU	TYPE DE DOCUMENT	TYPE DE PUBLICATION	ACTION	PASSAGE
0	Né électronique	Documents figés articles membres de l'UNIL (post print) thèses mémoires de licence ou de maîtrise documents officiels de l'UNIL inventaires des fonds manuscrits BCU inventaires des fonds manuscrits BCU bases de données fermées BCU (mss, etc.) publications officielles de la BCU photographie originale (atelier BCU + unimedia)	document de base à conserver à long terme avec toutes ses fonctions d'origine dans un dépôt électronique à long terme (Repository)	Peut être éliminé si le document est remplacé par un nouveau ou transféré dans un nouveau logiciel et fait foi comme original pour son auteur
			ce document ne doit être utilisé pour l'accès en ligne	
			copie du document pour accès en ligne en niveau 3	
		Documents évolutifs Bases de données ouvertes UNIL et BCU (manuscrits, passe partout, fmp base unil)	Rythme d'envoi d'état à définir pour chaque base	

POUR TOUS LES DOCUMENTS NÉS NON ÉLECTRONIQUES, C'EST LE DOCUMENT ORIGINAL SUR SON SUPPORT ORIGINAL (PAPIER, PARCHEMIN, FILM, ETC.) QUI EST CONSERVÉ POUR L'ÉTERNITÉ

1	Numérisation coûteuse faite sur la base d'un document analogique	Numérisation faite de manière manuelle pour des documents originaux fragiles et précieux (manuscrits, livres anciens,	document de base à conserver à long terme dans dépôt électronique type Repository	Peut être éliminé si une nouvelle numérisation est faite avec des techniques plus modernes
			ce document ne doit	

	photographies anciennes, documents sonores anciens) ayant nécessité un investissement important en temps de travail ou un investissement financier	pas être utilisé pour l'accès en ligne	
		copie du document pour accès en ligne en niveau 3	

PAS DE NIVEAU 2 POUR LE MOMENT. CORRESPOND À UN STATUT INTERMÉDIAIRE. DÉPEND DE L'ÉVOLUTION FUTURE DES DOCUMENTS ÉLECTRONIQUES

3	Copie des numérisations coûteuses	Copie accessible en ligne des documents de niveau 0 et 1	S'assurer d'avoir toujours une copie à jour et conforme à l'original avec toutes les fonctions d'origine	Evolue en parallèle avec le document conservé pour l'éternité
	Numérisation standard	Numérisation industrielle des documents imprimés conservés sur papier	uniquement garantir l'accès	
	Document "en travail"	Document qui est en permanence modifié par des ajouts (type liste de discussion, travaux en groupe, séminaires, cours) préprint (définition RERO)	uniquement garantir l'accès	Le travail terminé et validé par les instances académiques peut devenir un niveau 0
	Documents volatiles	Guide du lecteur BCU Formation des usagers BCU en ligne et tous les utilitaires d'accès à la BCU Guide de l'étudiant, programme des cours, informations administratives, etc... et tous les utilitaires de l'UNIL Rapports de recherche post master	uniquement garantir l'accès	Une version de ces documents utilitaires peut être archivée à long terme après leur période d'utilité si les institutons concernées le souhaite.
	Documents validés et édités	post print au sens de RERO	uniquement garantir l'accès est archivé sur papier ou par l'éditeur	

4	<p>Le niveau 4 étant le niveau d'élimination par défaut, il est absurde de l'attribuer à un document électronique. Soit le document n'entre pas dans le Serval, soit s'il est éliminé, il est remplacé par une version actualisée.</p> <p>La version éliminée disparaît, n'a plus de métadonnée donc l'introduction d'un niveau 4 est inutile.</p>
---	--

Nous devons mettre en ligne notre politique PAC pour pouvoir être inclus dans OpenDOAR.

Même si SERVAL ne possède pas, resp. n'est actuellement pas intégré dans une infrastructure de e-archiving, il ne faudra pas hésiter à donner un niveau PAC 0 aux documents et collections que l'on perdrait définitivement en cas de perte sur SERVAL. Le niveau PAC 0 est à comprendre comme une mise en garde, un appel à porter une attention particulière à une certaine partie de la collection. Les documents recevant un niveau PAC 0 devront en principe recevoir des URN, lorsque cela sera possible.

Le niveau 1 est un niveau qui peut, entre autres, faire apparaître l'équilibre des coûts. On traite différemment le résultat d'une numérisation très coûteuse que celui d'une numérisation bon marché. Il y a un équilibre à trouver entre les coûts d'archivage de l'original analogique et les coûts d'archivage du numérique, l'original devant toujours être sauvegardé de manière prioritaire.

Le niveau 3 concerne ce que l'on peut perdre, ce qui est facilement remplaçable. Les niveaux 2 et 4 ne sont pas utilisés à l'heure actuelle.

8 SERVAL et e-archiving

8.1 Etat actuel des efforts de e-archiving en Suisse

Durant la phase Concept, des contacts ont été pris au sujet du e-archiving avec des partenaires potentiels, notamment:

- avec le directeur des Archives cantonales vaudoises, qui a déjà adressé plusieurs argumentaires au Chancelier et à son chef de département pour les sensibiliser à la nécessité de prévoir une archive numérique pour le canton ; le directeur des Archives salue tout rapprochement potentiel entre son institution et la BCU dans ce projet, et est favorable à une démarche conjointe
- avec le chef de l'informatique cantonale, qui a confirmé avoir reçu les demandes des Archives, mais n'avoir ni les moyens ni les compétences pour répondre dans l'immédiat à ces besoins nouveaux
- avec la Cheffe du SERAC, pour sensibiliser le Département à cette nécessité, dans le cadre des démarches en cours pour mettre en place un centre d'archivage à long terme pour l'imprimé et le numérique (pour la BCU, pour le canton ou pour RERO)

Pour la BCU, l'archivage pérenne est une préoccupation importante et fondamentale pour ne pas perdre à terme l'investissement consenti dans les projets de rétronumérisation (projets autour des publications vaudoises, de la *Gazette de Lausanne* et de la *Nouvelle Revue de Lausanne*, ou en collaboration avec Google ou Assy). Même si, dans l'immédiat, c'est la mise en ligne de ces ressources de manière large qui importe, via Google Books, SERVAL et RERO DOC.

Les archives cantonales possèdent une coordination, KOST³⁷. Après un premier projet d'archivage de données de gestion (KOSTProbe) en 2006, KOST prévoit pour 2007 de s'essayer à deux formes de projet, en vue de l'élaboration d'une méthode:

- Un projet de méthode/système se préoccupe, en collaboration avec trois archives, de l'archivage de documents provenant de systèmes de GED: AUGev.
- Un projet de *policy* se propose de réaliser un catalogue des formats d'archives: KaD.
- Un projet de planification élabore les bases pour l'achat d'un système d'archives.

Comme on peut le voir, KOST se préoccupe surtout de l'archivage des documents administratifs.

³⁷ Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen, en ligne sous : <http://www.kost-ceco.ch/> [vu le 30 avril 2007].

La BN réfléchit également à la question ; l'idée est de mettre à disposition des espaces d'archivage pour les cantons. C'est le domaine académique qui reste le moins bien couvert, les chercheurs de nos universités ne publiant pas forcément dans une revue suisse (archivés à la BN dans le cadre du projet e-Helvetica). Ne sont pas couverts par l'archivage pérenne de la BN: les données scientifiques de laboratoire, les articles et publications des scientifiques publiés par des éditeurs étrangers, les mémoires, thèses de master et autres formes de travaux de fin d'études, les préprints, le résultat des numérisations, la littérature grise.

Le projet de E-Archiving du Consortium national des bibliothèques s'occupe en priorité des ressources électroniques commerciales (licences électroniques) pour lesquelles il est possible de sauvegarder localement (en Suisse) certaines données. Le projet ne couvre donc pas non plus ce type de documents.

8.2 Etat actuel du projet URN de la BN

Un groupe de travail national travaille actuellement à la mise au point d'une politique nationale des URN³⁸. Un URN (Uniform Resource Name) est un identificateur univoque pour les ressources dans Internet. Il est utilisé comme référence stable à des URL (Uniform Resource Locator ou « lien » dans Internet). La Bibliothèque nationale suisse (BN) attribue des URN sur la base du NBN³⁹ (National Bibliography Number).

Le principe des identificateurs durables et univoques (persistent identifiers) comme les URN est qu'ils pointent toujours sur l'URL le plus actuel ou qu'ils renvoient à l'URL original. Si ce dernier n'est plus accessible, un renvoi peut être fait à une copie d'archive. La durabilité d'un persistent identifier n'est garantie que s'il est entretenu, par exemple par des maisons d'édition, des producteurs ou des bibliothèques. Derrière chaque persistent identifier, il doit y avoir une institution qui veille à ce que les ressources Internet elles-mêmes ainsi que les liens sur ces ressources soient tenus à jour.

Seuls les documents qui seront archivés de manière pérenne recevront donc des URN et seules des institutions archivant de manière pérenne pourront en donner.

Le groupe de travail URN n'a pas encore rendu son rapport final. Plusieurs variantes ont été proposées par le groupe de travail; c'est la variante mixte qui a été choisie, c.-à-d. l'attribution d'URN en collaboration entre la BN et d'autres institutions. A l'heure actuelle, on ne sait pas encore quels seront exactement les critères pour être considéré comme institution procédant à un archivage pérenne.

³⁸ Les informations suivantes sont tirées du dernier communiqué du groupe de travail URN.

³⁹ Pour URN :NBN, voir RFC 3188, sous : <http://www.faqs.org/rfcs/rfc3188.html> [vu le 29 avril 2007].

Les URN concernent uniquement les documents devant être archivés de manière pérenne, mais on peut imaginer la mise en place d'un système semblable pour les identifiants des autres documents.

8.3 Objectifs SERVAL en matière de e-archiving

SERVAL n'est pas conçu comme une archive numérique pérenne au sens de la norme OAIS⁴⁰, car un tel projet dépasse de loin le cadre financier alloué actuellement au projet. SERVAL a cependant déjà permis aux membres du Comité de projet de se familiariser avec les exigences du e-archiving, dans le but de participer activement à une telle discussion au niveau cantonal, régional ou national.

L'excellent rapport d'OCLC du CRL⁴¹ formule de la manière suivante ce qui fait un *trustworthy digital repository*:

« In determining trustworthiness, one must look at the entire system in which the digital information is managed, including the organization running the repository, its governance, organizational structure and staffing ; policies and procedures ; financial fitness and sustainability ; the contracts, licenses, and liabilities under which it must operate ; and trusted inheritors of data, as applicable. Additionally, the digital object management practices, technological infrastructure, and data security in place must be reasonable and adequate to fulfill the mission and commitments of the repository. »

A ce niveau, SERVAL est un réel progrès en vue de l'archivage pérenne, car il crée une infrastructure centrale pour le dépôt des documents numériques. Le rassemblement des documents et métadonnées permet d'éviter les pertes dues au morcellement des collections sur des serveurs, disques durs et disques optiques éparpillés dans les institutions. A l'heure actuelle, les données sont souvent perdues car lorsque l'on veut les utiliser, on ne peut plus lire le support ou le format dans lesquels elles ont été stockées. L'administration centrale des documents permettra aussi de quantifier le nombre de documents selon les formats ou l'ancienneté et de prendre des premières mesures de conservation.

Dans un premier temps, on mettra l'accent, pour le e-archiving, sur les publications scientifiques UNIL/CHUV, étant donné que c'est cette partie de la production scientifique institutionnelle qui n'est pas couverte par l'archivage pérenne de la BN. Une copie d'archive des articles natif-électroniques UNIL pourraient recevoir un code PAC 0, et devraient idéalement être stockés sur un système d'archivage pérenne à créer. Durant la phase Réalisation, nous proposons de passer en revue l'Audit Checklist d'OCLC et du CRL, ce qui donnera des impulsions utiles pour la suite des réflexions sur l'archivage pérenne.

⁴⁰ Norme OAIS voir le site de la NASA : <http://nost.gsfc.nasa.gov/isoas/> [vu le 29 avril 2007].

⁴¹ OCLC and CRL, *Trustworthy Repositories Audit & Certification : Criteria and Checklist*, version 1.0, février 2007. En ligne sous : www.crl.edu/PDF/trac.pdf [vu le 29 avril 2007].

9 Fonctionnalités SERVAL

Dans ce chapitre, nous allons décrire comment les principales fonctionnalités de SERVAL seront implémentées et présentées à ses utilisateurs. Ces fonctionnalités sont les suivantes :

- Saisie des publications
- Validation des publications
- Recherche des publications
- *Metadata Harvesting*

L'archivage fait partie aussi des fonctionnalités offertes par SERVAL, mais ce sujet est traité dans un chapitre dédié. Pour plus de détails, voir chapitre 8.

9.1 Saisie des publications

L'une des mesures de la réussite d'un serveur institutionnel est le nombre de ressources numériques qu'il contient. Il est donc essentiel de fournir des outils adaptés aux utilisateurs pour les encourager à soumettre leurs publications.

Dans le projet SERVAL, nous proposons deux interfaces de soumission : BIB et FEZ; BIB pour la saisie des publications scientifiques par leurs auteurs et FEZ pour la saisie de publications scientifiques et e-collections spécialisées par les institutions partenaires.

Pourquoi proposer deux interfaces distinctes ? Parce qu'elles sont destinées à des utilisateurs aux besoins et compétences distincts. BIB est destinée aux auteurs et doit rester le plus simple possible et proche des pratiques de ceux-ci. FEZ est plutôt destinée aux professionnels bibliothécaires et aux administrateurs facultaires qui compléteront les données saisies par les auteurs dans un processus de validation ou soumettront des documents (collections) spécialisés.

9.1.1 Saisie de publications via BIB

L'application BIB a été développée par le DEE (*Département d'Ecologie et Evolution de l'UNIL*) comme réponse à un besoin interne. Elle est basée sur le standard de métadonnées BIBTEXML⁴², spécialement adapté aux publications scientifiques et très répandu dans l'échange de bibliographies sur Internet.

À la demande du DEE, le Ci a repris en charge cette application et l'a intégrée dans le portail MyUnil. Avec la refonte du catalogue Unisciences, elle est devenue l'unique interface de saisie des publications scientifiques pour les auteurs, les données saisies étant disséminées aussi bien dans le catalogue Unisciences que dans les sites Web des unités, et, à terme, dans le dépôt institutionnel SERVAL, comme nous le proposons ici.

⁴² Voir sous : <http://bibtexml.sourceforge.net/index.html>

La soumission d'une publication via BIB consiste à introduire les métadonnées relatives et à éventuellement y joindre son contenu. Les types de documents et métadonnées associés supportés par BIB correspondent au format BIBTEXML, ce qui permet l'importation depuis une base externe (par exemple EndNote) des références déjà publiées dans un autre dépôt (par exemple PubMed). Pour le document lui-même, les auteurs peuvent télécharger directement leur version sur le serveur (pour autant qu'ils en aient le droit) ou le référencer via une URL, si la publication est déjà disponible sur un autre serveur.

My UNIL Université de Lausanne
Faculté des sciences sociales et politiques
Bienvenue Ecrssp Portal_t

Aide | Actualiser | Votre avis | Contact

Recherche
Personnes

Mon Bureau | Actualités | Cours | Dossier administratif | **Publications** | Outils | Campus | Page actualisée le 26 avril 2007 16:44:50

Publications

bib dossier en cours d'édition Ecrssp Portal_t

article Un article d'un journal ou d'un magazine.
BibtexML @article
propriétaire Ecrssp Portal_t

champs obligatoires *

auteur *

titre *

journal *

année *

volume

numéro

pages

mois

note

résumé

mots-clé

URL

document attaché

Note: les champs "url" et "document attaché" sont mutuellement exclusifs.

création
Placez le curseur de votre souris sur l'intitulé du champ pour en obtenir une description.
La syntaxe des auteurs est particulièrement importante pour garantir le tri et l'affichage corrects des entrées.
documents
Pour attacher un document (par exemple un pdf) à cette entrée, choisissez un document de votre disque. Il sera envoyé et stocké sur un serveur de l'Université. Le lien vers ce fichier apparaîtra dans le champ "url" la prochaine fois que vous éditez cette entrée. Pour effacer ce fichier, il vous suffira de vider ou modifier ce champ.
Vous pouvez aussi directement saisir une url, par exemple si le pdf est déjà librement disponible ailleurs ou si vous ne pouvez l'envoyer pour des questions de droits.

Fig. 5 – BIB : saisie d'une publication

Importation

BibTeX

Fig. 6 - BIB: barre d'importation

Notre proposition d'utiliser BIB comme interface de soumission de publications scientifiques pour les auteurs est basée sur les arguments suivants :

1. **Unicité de l'outil de saisie.** BIB est déjà utilisé par les auteurs pour saisir leurs publications affichées dans Unisciences et les sites Web.
2. **Intégration dans MyUNIL.** BIB est déjà intégré dans le portail MyUNIL.
3. **Basé sur un standard.** Les types de documents ainsi que leurs métadonnées sont basés sur le standard BIBTEML, couramment utilisé dans la communauté scientifique.
4. **Simplicité et souplesse de l'interface.** L'interface de BIB est très simple, légère et facile à utiliser. Ce facteur est déterminant pour inciter les auteurs à soumettre leur production.
5. **Outil configurable.** La maîtrise de l'outil permet de le configurer de telle sorte que la saisie des métadonnées soit la plus légère possible (ensemble obligatoire minimal, remplissage automatisé lorsque possible).
6. **Outils d'importation intégrés.** BIB contient déjà des outils pour importer des références de publications depuis une base externe.
7. **Basé sur la structure Unisciences.** La structure des unités et des équipes de recherches réutilise celle mise en place dans Unisciences.
8. **Techniquement, intégrable avec Fedora.** BIB est développé en JAVA, donc intégrable de manière consistante à Fedora.
9. **Bonne expérience du DEE.** Le DEE utilise BIB à satisfaction depuis plus d'une année.
10. **Bon feedback.** Lors des diverses présentations BIB-SERVAL que nous avons effectuées ces derniers mois, l'accueil a été très favorable et les membres des différents départements concernés se sont montrés très enthousiastes face à l'utilisation de cette interface.

Actuellement, les auteurs peuvent à tout moment modifier les métadonnées ou le contenu d'une publication soumise. Cette possibilité devra être restreinte à l'étape finale du workflow de soumission.

9.1.2 Saisie de publications via FEZ

FEZ est une application Web, conçue et implémentée par l'Université de Queensland (Australie) pour combler le manque d'interfaces graphiques pour gérer un dépôt Fedora. Les principales caractéristiques de FEZ et les motivations de son choix sont décrites dans le chapitre 11.

Parmi ses nombreuses fonctionnalités, FEZ permet la soumission des documents et leurs métadonnées associées vers un dépôt Fedora. Les types de documents et leurs métadonnées sont entièrement configurables par l'administrateur du

dépôt. Sont compris, par défaut, les standards de métadonnées Dublin Core (DC)⁴³ et MODS⁴⁴.

Dans le cadre du projet SERVAL, nous proposons l'utilisation de FEZ comme interface de saisie d'une partie de la production scientifique (par exemple celle dont les auteurs ont quitté l'UNIL), pour compléter et valider les saisies faites par les auteurs et pour la saisie de documents multimédia (par exemple numérisation des livres ou d'enregistrements musicaux) par les institutions partenaires (BCU/BiUM) ainsi que les administrateurs facultaires.

The screenshot shows a web form titled "Create New Image 1.0 Record". The form is organized into several sections:

- Member of Collections:** A dropdown menu with options: Bib Journal Articles, Images, Période 1748-1750, Publications IUHM, Publications IUMSP, Règlements et Directives, Thèses lausannoises.
- Title:** A text input field.
- Creator(s) 1:** A text input field with a note: "(More Creator(s) input boxes will appear as they are used up)".
- Research Fields, Courses and Disciplines:** A vertical scrollable list with "Add Subject" and "Remove Subject" buttons.
- Keyword(s) 1:** A text input field with a note: "(More Keyword(s) input boxes will appear as they are used up)".
- Description:** A large text area.
- Publisher:** A text input field.
- Contributor 1:** A text input field with a note: "(More Contributor input boxes will appear as they are used up)".
- Date:** A date picker showing April 26, 2007.
- Type, Format, Source, Language, Relation, Coverage, Rights:** Each has a corresponding text input field.
- Succeeds:** A text input field with a note: "if this object succeeds another object then enter the existing object PID here".
- Refereed?:** A checkbox.
- Copyright Agreement *:** A checkbox with a long text block explaining the terms of deposit.
- Additional Notes:** A text area.
- Add Links:** A section with "Description for Link 1" and "URL 1" input fields.
- Upload Files:** A section with "Description for File Upload 1" and "File Upload 1" input fields, and a "Browse..." button.

At the bottom, there are buttons for "Publish", "Submit for Approval", "Done", and "Reset Form". A note at the bottom left says "* Required fields".

Fig. 7 - FEZ: saisie d'une image

⁴³ Voir *Core Metadata Initiative* : <http://dublincore.org/>

⁴⁴ Voir le site Web officiel de MODS : <http://www.loc.gov/standards/mods/>

Cette interface est configurable; elle sera donc adaptée aux besoins réels et traduite en français.

9.2 Validation des publications

Toute publication soumise à SERVAL sera validée avant d'être visible sur Internet et avant que ses métadonnées soient *harvestables*.

Ces validations interviendront à différents moments du workflow (voir chapitre 9) et seront effectuées au moyen des deux interfaces BIB et FEZ.

9.2.1 Validation via BIB

Lorsqu'un chercheur saisit une publication, avant que cette publication soit visible dans les pages Web de son institut, l'administrateur de sa faculté, s'il existe, pourra valider la publication. Cette fonctionnalité de BIB n'est pas encore implémentée, mais le sera dans le cadre du projet SERVAL.

9.2.2 Validation via FEZ

Après validation éventuelle par la faculté, mais avant transfert dans la partie visible sur Internet du dépôt institutionnel, une publication devra être validée par un spécialiste du domaine qui aura accès aux métadonnées saisies par l'auteur et pourra les corriger et les enrichir au besoin. Cette validation sera faite au travers de l'interface FEZ.

9.3 Interfaces de recherche

Si précédemment nous avons mentionné l'importance d'offrir de bons outils pour la saisie du contenu dans le serveur institutionnel, il en va de même pour la recherche de son contenu. Outre les possibilités offertes par les moteurs de recherches classiques style Google, nous proposons deux interfaces : d'une part FEZ, pour les recherches effectuées par les institutions partenaires et d'autre part une nouvelle application Web dédiée exclusivement à la recherche et ouverte à tout public sur Internet.

9.3.1 FEZ en tant qu'interface de recherche

FEZ contient de manière native une interface de recherche composée des éléments suivants :

- **Recherche simple.** FEZ permet d'effectuer des recherches sur les champs indexés de sa base de données, comme le titre, la description ou encore les mots-clés d'un document. L'administrateur peut configurer, pour chaque type de document, quels champs seront indexés.



Fig. 8 - FEZ: barre de recherche simple

- **Recherche avancée.** FEZ permet d'affiner les recherches par champ de description du document, comme le titre, l'auteur, etc. Les champs où ce type de recherche est permis sont également configurables par l'administrateur.

Fig. 9 - FEZ: interface de recherche avancée

- **Navigation dans le dépôt.** Finalement, FEZ fournit aussi la possibilité de naviguer dans le dépôt en fonction de sa hiérarchie (les collections), de la classification, des dates de soumission ou des auteurs. Il offre aussi une vue sur les nouveautés introduites dans le dépôt.

Browse

[by Communities & Collections](#) | [by Subject](#) | [by Year](#) | [by Author](#) | [by Latest additions](#)

Fig. 10 - FEZ: critères de navigation

9.3.2 Nouvelle application Web pour la recherche SERVAL

L'interface FEZ étant réservée aux membres des institutions partenaires, BCU, BiUM et les administrateurs facultaires, nous proposons le développement d'une nouvelle application Web qui sera exclusivement dédié à la recherche et qui sera ouverte à tout public sur Internet.

Les spécifications fonctionnelles et techniques de cette application ne sont pas encore définies, mais l'idée est de fournir des fonctionnalités similaires, mais simplifiées, à celles de FEZ telles recherche simple, recherche avancée et navigation. Par ailleurs, une authentification permettra d'enrichir le résultat des recherches par des entrées disponibles en intranet seulement ou sous embargo. Cette authentification sera dans un premier temps réalisée via l'annuaire LDAP de l'UNIL, une extension au système AAI- Shibboleth étant envisageable.

9.4 Metadata Harvesting

À part les interfaces de recherche décrites dans la section précédente, il existe un autre moyen pour améliorer la visibilité des ressources contenues dans un dépôt d'archives ouvertes: il s'agit du *metadata harvesting* (ou moissonnage de métadonnées en français). Cette fonctionnalité consiste à exposer les métadonnées des documents déposés dans notre dépôt pour qu'ils soient collectés par des fournisseurs de services externes. Ces fournisseurs de services peuvent être locaux (comme RERO⁴⁵ ou la Bibliothèque Nationale) ou bien internationaux (comme OAISTER⁴⁶).

Pour faciliter la description et la diffusion des métadonnées, un protocole spécialisé, le OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*)⁴⁷, a été élaboré en 1999 dans le cadre des travaux de l'OAI (Open Archives Initiative)⁴⁸. Ce protocole définit les conditions de transfert de métadonnées d'une archive ouverte, produites par un fournisseur de données, vers le serveur d'un fournisseur de services. Après une phase d'expérimentation, le protocole a atteint une stabilité depuis juin 2002, avec sa version 2, considérée comme la spécification définitive. Actuellement, l'usage de ce protocole est devenu une référence en ce qui concerne la diffusion des métadonnées.

Dans notre cas, le produit FEDORA offre un service qui implémente le protocole OAI-PMH. Nous pourrions exposer nos métadonnées à travers ce service (voir chapitre 11 pour plus de détails).

En ce qui concerne le contenu que nous allons exposer, nous proposons de diffuser les métadonnées de ressources satisfaisant les conditions suivantes :

⁴⁵ Réseau de bibliothèques de Suisse occidentale. Voir : <http://www.rero.ch/>

⁴⁶ Voir : <http://www.oaister.org/>

⁴⁷ Voir : <http://www.openarchives.org/pmh/>

⁴⁸ Voir : <http://www.openarchives.org>

- La ressource numérique est dans son état final et n'est donc plus modifiable.
- La ressource numérique a été produite par l'une des institutions partenaires de ce projet (UNIL, BCU ou CHUV/BiUM) ou elle est le résultat de la numérisation de collections BCU ou BiUM libres de droits.

10 Workflows de dépôt de documents et de métadonnées

10.1 Workflow pour le dépôt de la production scientifique par les auteurs

Ce workflow est initié par les utilisateurs finaux dans leur rôle de producteur d'information. Il concerne l'un des types de documents prioritaires dans le projet SERVAL: les articles en travail ou déjà publiés par les chercheurs UNIL/CHUV.

Dans le schéma ci-dessous, les différentes cases numérotées représentent le même objet FEDORA dans un état particulier correspondant au processus de workflow, les flèches entre ces cases représentent les actions entreprises par les divers acteurs menant au changement d'état de l'objet.

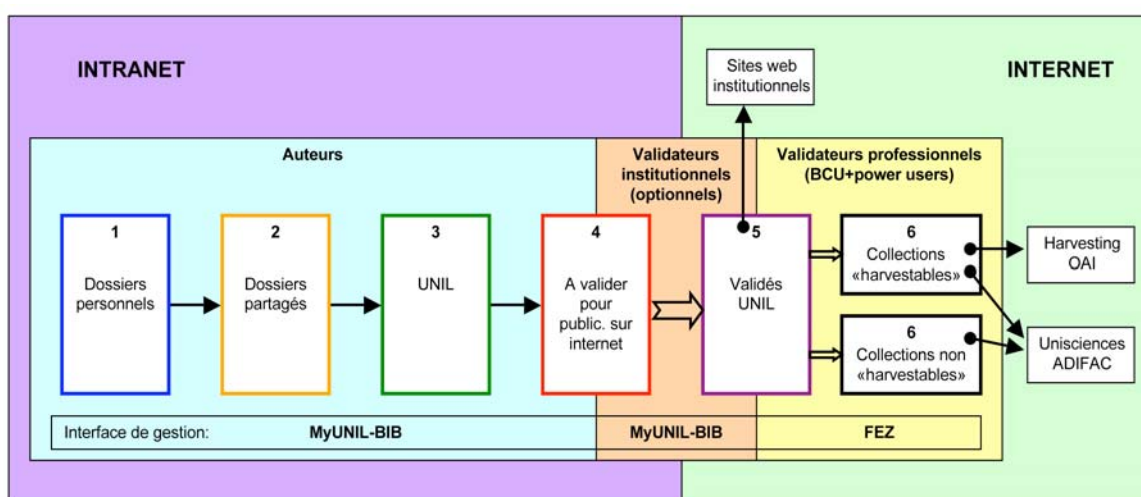


Fig.11 Workflow: dépôt de la production scientifique par les auteurs

1. L'auteur crée son document en format courant sur son dossier personnel dans MyUNIL en saisissant les métadonnées dans BIB (format BibTeX). Ce document n'est d'abord accessible que par lui seul ou par ses coauteurs. Dans ce second cas, l'accès est généré dans BIB pour les co-auteurs (authentifiés UNIL) rendant l'objet visible dans leur dossier personnel. C'est le même objet FEDORA qui est visible chez ces co-auteurs.
2. Étape optionnelle: l'auteur décide de partager son document avec un groupe de non-auteurs (authentifiés UNIL toujours), comme l'équipe de recherche par exemple, qui peut alors le voir et le modifier.
3. L'auteur décide d'ouvrir l'accès à son document à la communauté UNIL. Le document est dès lors visible pour tous les membres UNIL correctement authentifiés, mais modifiable uniquement par l'auteur et les co-auteurs.
4. L'auteur décide de rendre visible son document sur Internet. Il donne son accord dans l'interface BIB en cochant en particulier la case sur les droits

d'auteur (voir chapitre 6). À ce stade, l'auteur doit, s'il ne l'a pas déjà fait, fournir un certain nombre de métadonnées obligatoires, notamment remplir un champ concernant l'embargo (les métadonnées seront visibles, mais le contenu peut être soumis à un embargo plus ou moins long, jusqu'à indéterminé). L'objet est alors en attente de validation (optionnelle) par l'institut/faculté. À cette étape, le document est considéré comme terminé.

5. Si une validation interne (institut/faculté) existe, le document proposé par l'auteur au point précédent est validé et devient visible sur les sites Web institutionnels. L'objet est alors en attente de validation par les administrateurs SERVAL à la BCU/BiUM.

6. Après avoir complété les métadonnées via l'interface FEZ, les administrateurs valident la publication pour SERVAL. À ce moment, l'objet FEDORA est indexé dans FEZ. Il fait alors partie d'une des collections SERVAL. Certaines collections (production scientifique UNIL, résultats de la numérisation BCU libre de droits) sont harvestables, d'autres non (production scientifique pré-UNIL, e-collection de la BCU et de la BiUM ne provenant pas de la numérisation). Dès cet instant, le document est visible dans les interfaces de présentation SERVAL, Unisciences et ADIFAC, selon la nature du document (un document provenant de la numérisation BCU n'apparaîtra pas dans Unisciences; un document de recherche de l'UNIL apparaîtra dans Unisciences et un document de recherche du CHUV apparaîtra dans ADIFAC).

Le workflow est un peu différent pour les thèses. D'une part, les thèses exigent des métadonnées plus complexes que les autres types de documents. D'autre part, la mise en ligne dans SERVAL ne peut se faire qu'après le dépôt de la version imprimée à la BCU. Or dans certains domaines académiques, plusieurs années peuvent s'écouler entre la soutenance de la thèse et sa parution sous forme imprimée. Dans un premier temps, le thésard pourra donc saisir les métadonnées sommaires de sa thèse dans BIB, afin de pouvoir éditer rapidement ces métadonnées sur son site personnel. Puis, au moment du dépôt de la thèse à la BCU (version imprimée), il remplira un formulaire pour compléter ces métadonnées et les rendre compatibles avec les exigences nationales (l'archivage pérenne est assuré par la BN).

10.2 Workflow pour le dépôt de la production UNIL, scientifique ou non-scientifique, par les institutions partenaires

Ce workflow est initié par les utilisateurs institutionnels dans leur rôle de producteurs d'information. Il concerne d'une part des actions de saisie de métadonnées avec import ou non des documents numérisés. C'est le cas notamment si une faculté décide d'investir dans une saisie complète des métadonnées de tous les mémoires rédigés au sein de la faculté, ou lorsque le photographe de la BCU dépose ses nouvelles productions sur SERVAL.

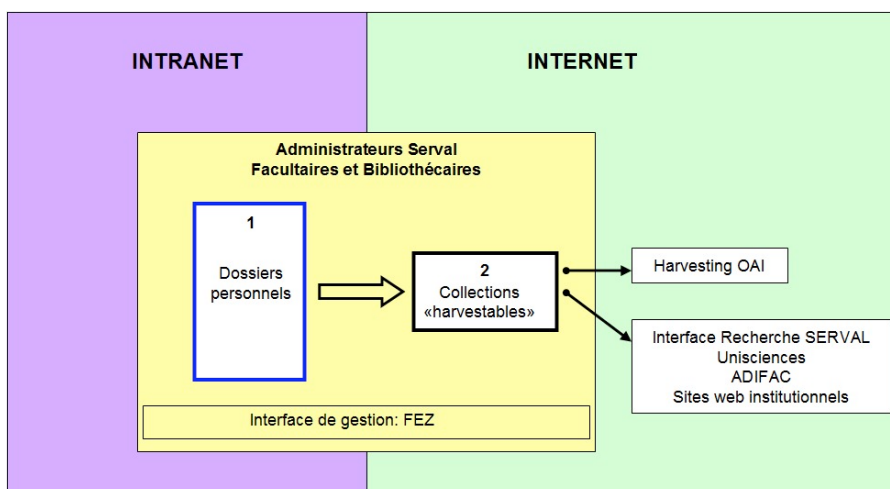


Fig. 12 - Workflow: dépôt de la production UNIL par les institutions partenaires

1. L'administrateur SERVAL (facultaire ou bibliothécaire) saisit les métadonnées d'un document qu'il a dans son dossier personnel dans l'interface FEZ dédiée au type de document qu'il doit saisir. Si document numérique il y a, il joint ce document aux métadonnées, certifie le respect des droits d'auteurs et soumet le document à la publication définitive. L'objet FEDORA est directement indexé dans FEZ et fait alors partie d'une des collections SERVAL. Il n'est visible que pour des utilisateurs authentifiés à l'UNIL.

2. L'administrateur SERVAL décide de rendre le document visible sur Internet. Dès cet instant, le document est visible dans les interfaces de présentation SERVAL, Unisciences et ADIFAC, sites Web institutionnels, selon la nature du document (un document provenant de l'atelier de photographie BCU n'apparaîtra pas dans Unisciences; un document de recherche de l'UNIL apparaîtra dans Unisciences et un document de recherche du CHUV apparaîtra dans ADIFAC). La production propre étant en principe harvestable, ces collections seront donc soumises au harvesting par OAI-PMH.

10.3 Workflow pour la saisie de e-collections non UNIL par les partenaires

Ce workflow est en principe initié par les bibliothécaires administrateurs SERVAL pour saisir les métadonnées et éventuellement importer des documents qui ne proviennent pas de production propre (scientifique ou non) des institutions partenaires. Il concerne par exemple le résultat de petits programmes de numérisation BCU ou BiUM pour lesquels on n'aura pas prévu de procédure d'import en masse ou la saisie de références sur les documents scientifiques intéressants se trouvant sur d'autres serveurs OA par les bibliothécaires.

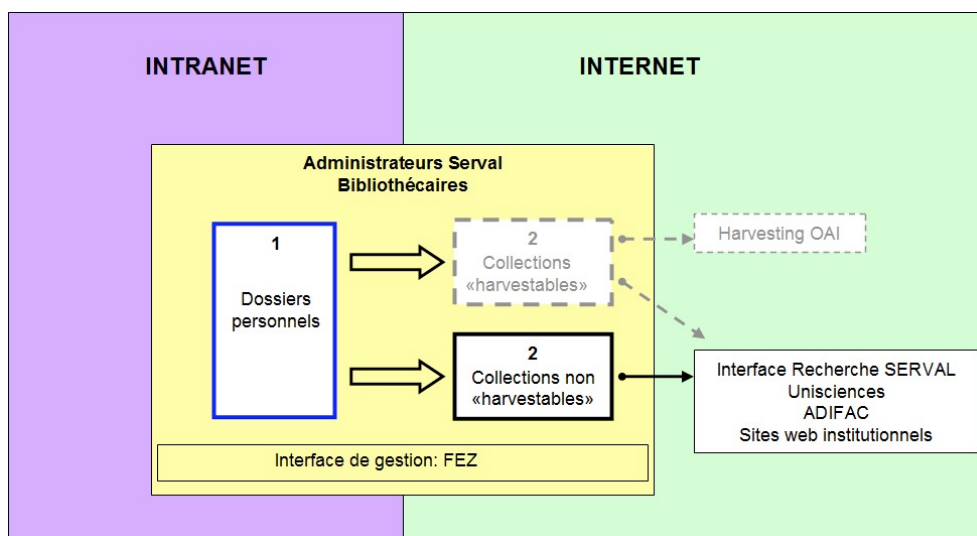


Fig. 13 - Workflow: saisie de e-collections non UNIL par les partenaires

1. L'administrateur SERVAL (facultaire ou bibliothécaire) saisit les métadonnées d'un document qu'il a dans son dossier personnel dans l'interface FEZ dédiée au type de document qu'il doit saisir. Si document numérique il y a, il joint ce document aux métadonnées, certifie le respect des droits d'auteurs et soumet le document à la publication définitive. L'objet FEDORA est directement indexé dans FEZ et fait alors partie d'une des collections SERVAL. Il n'est visible que pour des utilisateurs authentifiés à l'UNIL.

2. L'administrateur SERVAL décide de rendre le document visible sur Internet. Dès cet instant, le document est visible dans les interfaces de présentation SERVAL, Unisciences et ADIFAC, site Web institutionnel, selon la nature du document (un livre numérisé BCU n'apparaîtra pas dans Unisciences; un article non-UNIL saisi par un bibliothécaire n'apparaîtra pas dans Unisciences, mais sur le site Web BCU, sur une liste de ressources par discipline). Comme il ne s'agit pas d'une production propre, ces collections ne seront donc pas soumises au harvesting par OAI-PMH, sauf s'il s'agit d'un résultat de numérisation des collections BCU ou BiUM libres de droits.

10.4 Workflow d'importation

Là où existent de grandes masses de métadonnées et de documents numériques de même nature, on procédera à l'importation en masse. L'importation se déroulera essentiellement en trois étapes:

- Exportation des données du système source si possible en format XML.
- Transformation des données XML en FOXML (Fedora Object XML) ou METS au moyen de XSLT⁴⁹.
- Insertion des objets FOXML ou METS⁵⁰ dans SERVAL.

⁴⁹ XSLT sur le site du W3C: <http://www.w3.org/TR/xslt> [vu le 13 avril 2007].

⁵⁰ Sur le standard METS, voir le site: <http://www.loc.gov/standards/mets/> [vu le 13 avril 2007].

Les données d'Alexandrie (BiUM) tout comme celles de RERO DOC sont exportables en format XML. Il devrait donc y avoir relativement peu de cas pour lesquels cette procédure ne serait pas recevable. Pour ces cas, d'autres processus d'importation devront être développés.

11 Utilisateurs, authentification et sécurité

Dans le rapport d'analyse préliminaire, les publics cibles de SERVAL ont été déterminés de la manière suivante :

- Les administrateurs du dépôt institutionnel (sur le serveur)
- Les utilisateurs finaux dans leur rôle de lecteurs et chercheurs d'information
- Les utilisateurs finaux dans leur rôle de producteurs d'information
- Les utilisateurs institutionnels dans leur rôle de chercheurs d'information
- Les utilisateurs institutionnels dans leur rôle de producteurs d'information
- Les agrégateurs de contenus
- Les méta-utilisateurs

11.1 *Les administrateurs du dépôt institutionnel*

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Administrateur technique (Ci)	- gestion technique du dépôt. - Gestion des logiciels intégrés dans le système. - Création des scripts pour l'importation massive des ressources numériques.	Divers outils d'administration.	LDAP-UNIL FEDORA FEZ	Accès en lecture et écriture à tous les éléments du système SERVAL.
Administrateurs bibliothécaires (BCU, BiUM)	- Validation et correction des métadonnées (classification comprise).	FEZ	LDAP-UNIL FEZ	Accès en lecture et écriture aux ressources numériques qui doivent être validées.
Administrateurs facultaires (UNIL, BCU, CHUV)	- Aide aux auteurs pour l'utilisation de l'interface de soumission - Validation des publications au sein de l'unité.	BIB	LDAP-UNIL	Accès en lecture et écriture aux documents soumis au sein de leurs unités et qui n'ont pas encore été validés.

	- Validation et correction des métadonnées (classification comprise).	FEZ	LDAP-UNIL FEZ	Accès en lecture et écriture aux ressources numériques qui doivent être validées.
--	---	-----	------------------	---

11.2 Les utilisateurs finaux dans leur rôle de lecteurs et chercheurs d'information

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Membre de l'UNIL	- Chercher des ressources stockées dans SERVAL.	Interface de Recherche SERVAL	LDAP-UNIL	Accès en lecture à toutes les ressources numériques du dépôt qui sont exposées dans les zones Intranet et Internet. Le contenu d'une publication sous embargo est aussi accessible en lecture.
Internaute	- Chercher des ressources stockées dans SERVAL.	Interface de Recherche SERVAL	Aucune	Accès en lecture à toutes les ressources numériques du dépôt qui sont exposés dans la zone Internet. Le contenu des publications qui sont sous embargo n'est pas accessible.

11.3 Les utilisateurs finaux dans leur rôle de producteurs d'information

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Membre (chercheur, professeur) de l'UNIL ou du CHUV qui est auteur d'une publication scientifique	- Déposer un nouveau document dans SERVAL. - Importer dans SERVAL les références des documents déjà déposés dans d'autres dépôts.	BIB	LDAP-UNIL	Accès en lecture et écriture aux documents dont l'utilisateur est un des auteurs.

11.4 *Les utilisateurs institutionnels dans leur rôle de chercheurs d'information*

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Institutions partenaires du projet SERVAL	- Alimenter automatiquement les systèmes ou les pages Web qui affichent les ressources numériques stockées dans SERVAL.	Sites Web institutionnels Unisciences Adifac	Non authentifié	Accès en lecture à toutes les ressources numériques du dépôt qui sont exposés dans la zone Internet. Le contenu des publications qui sont sous embargo n'est pas accessible.

11.5 *Les utilisateurs institutionnels dans leur rôle de producteurs d'information*

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Administrateurs bibliothécaires (BCU, BiUM) Administrateurs facultaires (UNIL, BCU, CHUV)	- Déposer une nouvelle ressource numérique sur SERVAL.	FEZ	LDAP-UNIL FEZ	Droit de créer et valider une ressource numérique.

11.6 *Les agrégateurs de contenus*

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Fournisseurs de services qui collectent des métadonnées, comme (OAISTER, RERO, BN, etc.)	- Exposer les métadonnées des ressources numériques susceptibles d'être collectés par des fournisseurs de services externes.	OAI-PMH	Non	Accès en lecture aux métadonnées des ressources numériques définies en tant que <i>harvestables</i> ou <i>collectables</i> .

11.7 Les méta-utilisateurs

Utilisateur	Actions	Interface	Authentification	Accès et sécurité
Les directions de chaque institut partenaire (UNIL, BCU, CHUV/BiUM) et les organismes subventionnant la recherche.	- Analyser les résultats de la recherche.	Une interface dédiée à cette analyse n'existe pas, mais pourrait être développée. Des extractions seront réalisées par les administrateurs.	LDAP-UNIL	Accès en lecture à l'ensemble des objets stockés dans le dépôt.

12 Architecture

12.1 *Choix technologiques*

En août 2006, le comité de pilotage du projet SERVAL a décidé de poursuivre les tests préliminaires avec le produit FEDORA⁵¹. L'idée adoptée était de tester FEDORA pour s'assurer des possibilités réelles du produit. Depuis octobre 2006, nous avons installé et testé les versions 2.1.1 et 2.2 de FEDORA, ainsi que les versions 1.2 et 1.3 de l'interface FEZ (système de gestion du contenu et du workflow d'un dépôt numérique basé sur FEDORA).

L'architecture et la conception du produit FEDORA, les tests et analyses que nous avons effectués en interne ainsi que les expériences et études externes sur le produit⁵² ont fait que nos attentes par rapport au produit FEDORA se sont confirmées.

12.1.1 Principales caractéristiques de l'architecture FEDORA

*FEDORA (Flexible Extensible Digital Object and Repository Architecture)*⁵³ est un système fondé sur une architecture d'objets numériques extensibles et flexibles. Ce système, conçu conjointement par l'Université de Virginia et l'Université de Cornell, est actuellement financé par la *Andrew W. Mellon Foundation*. Différents groupes de travail participent au développement des nouvelles fonctionnalités et à l'amélioration de l'existant. Ces groupes de travail sont supervisés par un comité d'architecture. Il est important de signaler que, autour de ces groupes, existe une forte communauté de personnes qui échangent leurs connaissances par rapport au produit et qui développent aussi de nouveaux services qui peuvent après être intégrés dans le système principal. FEDORA est écrit en Java et il est disponible en Open Source, sous les termes de la licence ECL (*Educational community License 1.0*)⁵⁴.

Dans ce chapitre, nous allons voir les principaux aspects de l'architecture FEDORA qui ont déterminé notre choix pour ce produit. Ces aspects sont les suivants :

- Architecture orientée service (*FEDORA Service Framework*)
- *Fedora Core Service Architecture* (conforme au OAIS)
- Politique de conservation
- Implémentation du protocole OAI-PMH
- Politique de sécurité configurable et ajustable

⁵¹ Voir le Protocole de la séance SERVAL du 4 août 2006 :

http://www.bscw.unil.ch/bscw/bscw.cgi/d110138/2006_08_04ProtocoleSERVAL.doc

⁵² Voir l'étude accompli en Nouvelle Zélande *Technical Evaluation of selected Open Source Repository Solutions on behalf of CPIT* :

<https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf>

⁵³ Fedora main page : <http://www.Fedora.info>

⁵⁴ Voir sous : <http://www.opensource.org/licenses/ecl1.php>

La structure des objets, qui peuvent être hétérogènes et complexes, est présentée dans un chapitre propre.

Architecture orientée service (Fedora Service Framework)

Le schéma suivant montre la base de l'architecture orientée service de FEDORA.

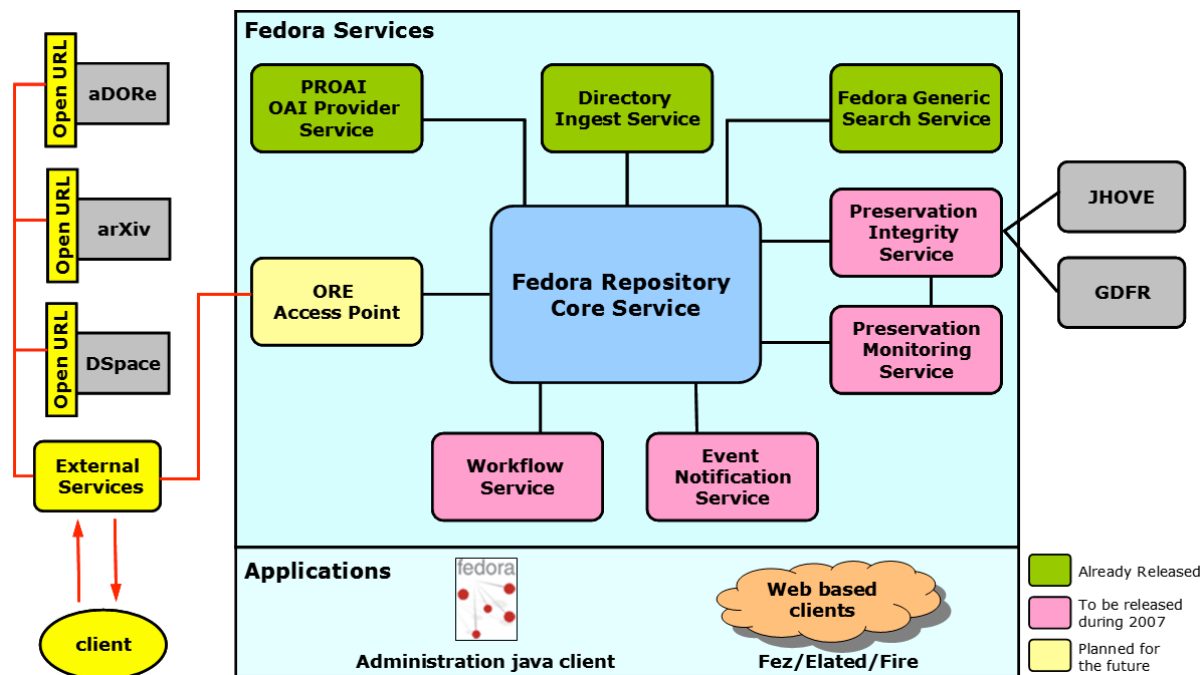


Fig. 14 - FEDORA Service Framework

Au centre de ce framework figure le *FEDORA Repository Core Service*. Ce service (expliqué plus en détail dans la section suivante) est en charge d'offrir les fonctionnalités de base du dépôt, c'est-à-dire la soumission et la validation des objets, le stockage et l'indexation des objets soumis, l'accès aux objets déposés, leur recherche, la gestion du dépôt, la sécurité et le contrôle des objets du dépôt, etc.

Autour de ce service central gravitent les autres services qui l'utilisent pour offrir eux-mêmes des services plus spécialisés. Ces services sont déployés sous forme d'applications Web indépendantes. Les services les plus importants sont les suivants :

- *FEDORA OAI Provider Service* : implémente le protocole OAI-PMH.
- *Directory Ingest Service* : offre des fonctionnalités spécialisées pour la soumission des objets.
- *FEDORA Generic Search* : permet d'effectuer des recherches au sein du contenu des objets. Notre interface de recherche sera basée sur ce service.
- *Preservation Integrity Service* et *Preservation Monitoring Services* : services spécialisés dans la conservation des objets.
- *Event Notification Service* : en charge de la notification des événements.

- *Workflow Service* : service spécialisé dans la gestion de workflow.
- *ORE Access Point* : implémente le standard OAI-ORE (Open Archives Initiative – Object Reuse and Exchange). Le but de ce standard est de définir les bases pour échanger des objets entre différents systèmes (dépôts).

Cette architecture modulaire a deux avantages principaux : d'une part elle permet de maintenir les fonctionnalités basiques du dépôt de la manière la plus stable possible, car les nouvelles fonctionnalités sont intégrées dans de nouveaux services sans aucun impact dans le *FEDORA Core Service* ; d'autre part, elle facilite l'intégration de nouveaux services de façon indépendante et modulaire.

FEDORA Core Service Architecture (conforme au OAIS)

Le schéma suivant décrit l'architecture du *FEDORA Core Service*, qui est conforme au modèle de référence OAIS⁵⁵.

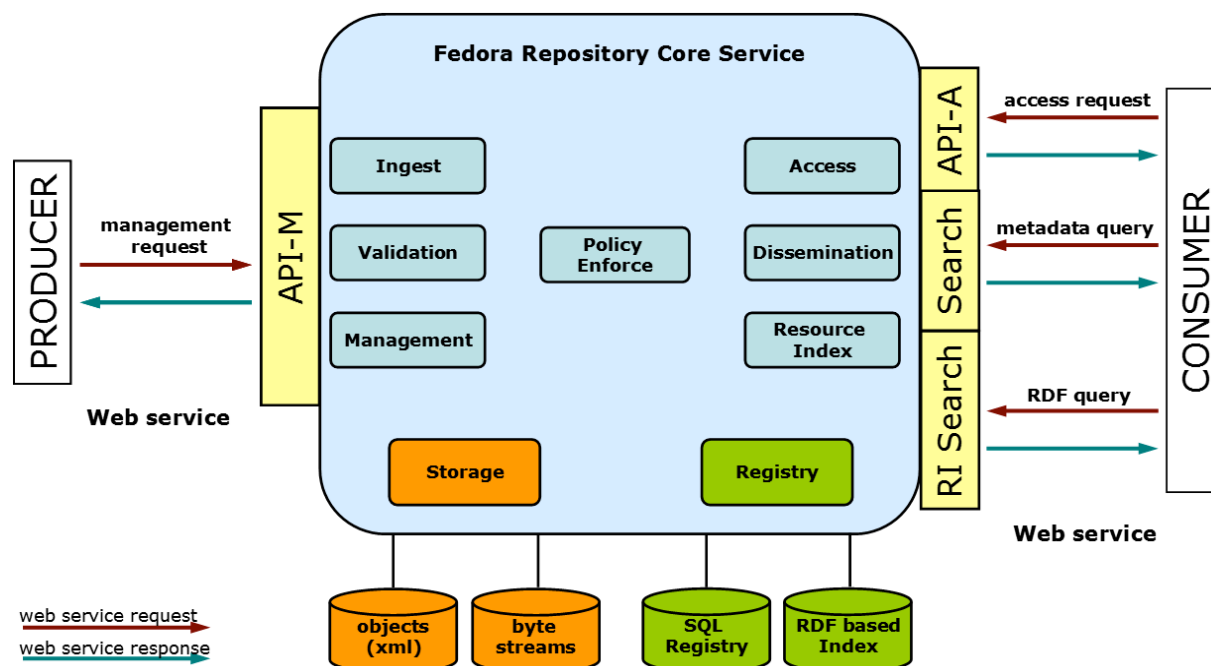


Fig. 15 - FEDORA Repository Core Service

Comme montré dans le schéma ci-dessus, le FEDORA Core Service est composé des modules suivants :

- *Ingest* : insère les objets dans le dépôt.
- *Validation* : valide les objets insérés dans le dépôt.
- *Management* : gère les objets insérés dans le dépôt.
- *Policy* : gère le contrôle d'accès sur les objets insérés dans le dépôt.
- *Access* : gère l'accès aux objets insérés dans le dépôt.

⁵⁵ Open Archival Information System : <http://ssdoo.gsfc.nasa.gov/nost/isoas/>

- Dissemination : fournit du contenu dynamique à partir du contenu statique des objets.
- Resource Index : garde les relations entre les objets et fournit un accès aux objets en fonction de leurs relations.
- Storage : stocke les objets: d'une part, la structure de l'objet en format XML, d'autre part, le contenu de l'objet (par exemple un fichier PDF représentant un article) en forme de byte-stream.
- Registry : indexe l'information relative aux objets, d'une part dans une base de données SQL pour les champs indexés par FEDORA (comme le titre, par exemple), d'autre part dans une base au standard *Resource Description Framework (RDF)* pour les relations entre les objets.

Les fonctionnalités fournies par ces modules sont accessibles via des Webservices (protocoles SOAP⁵⁶ et REST⁵⁷) exposés par des interfaces FEDORA (API-A, API-M, interface de recherche, interface de recherche sur le Resource Index).

Note : techniquement, le *FEDORA Core Service* est déployé en tant qu'application Web indépendante. Cette architecture s'adapte très bien à l'environnement informatique de l'UNIL et aux compétences du Ci. Il facilite l'intégration avec des produits déjà en production, notamment avec BIB (l'application utilisée par les chercheurs pour soumettre leurs publications).

Politique de conservation

Actuellement, un groupe de travail au sein de la communauté FEDORA travaille sur la spécification et le développement d'une définition générale des services de conservation au sein du framework de services FEDORA. Comme signalé dans le schéma du framework, les services de conservation spécifiques à FEDORA seront intégrés dans le système au cours de cette année 2007.

Pour l'instant, en attendant ces nouveaux services, la politique de conservation de FEDORA est basée sur les points suivants :

- La structure des objets numériques est stockée en format XML, définie par le schéma XML *FEDORA Object XML* (FOXML)⁵⁸.
- L'insertion, l'importation et l'exportation des objets numériques sont aussi basées sur le format XML. Actuellement, les schémas FOXML et METS⁵⁹ sont supportés.
- Rétablissement du dépôt : il est possible de rétablir le contenu du dépôt à partir des structures XML sauvegardées des objets déposés.
- Toutes les modifications faites sur un objet numérique sont enregistrées (*Audit Trail*).

⁵⁶ Simple Object Access Protocol : <http://www.w3.org/TR/soap/>

⁵⁷ Representational State Transfer

⁵⁸ FOXML schéma : <http://www.Fedora.info/definitions/1/0/foxml1-0.xsd>

⁵⁹ Metadata Encoding & Transmission Standard : <http://www.loc.gov/standards/mets/>

- Il est possible de garder différentes versions du contenu des objets digitaux.
- FEDORA Journaling module : ce module optionnel, fourni avec la dernière version de FEDORA (version 2.2), s'occupe de la création de fichiers qui enregistrent toutes les actions qui ont changé l'état du dépôt. Ces enregistrements peuvent être utilisés pour rétablir le dépôt à l'état plus récent ou bien pour créer un dépôt miroir.

Implémentation du protocole OAI-PMH

Comme signalé dans le schéma qui décrit l'architecture FEDORA le service *FEDORA OAI Provider Service* implémente le protocole OAI-PMH⁶⁰. Ce service est basé sur l'application PROAI⁶¹, une application Java qui implémente la version 2.0 du protocole OAI-PMH indépendamment du type de dépôt.

Les caractéristiques les plus significatives de ce service sont les suivantes :

- Il est déployé en forme d'application Web indépendante du service de base FEDORA.
- Il supporte n'importe quel format de métadonnées, il n'est pas limité au Dublin Core.
- Les métadonnées peuvent être disponibles de façon statique ou dynamique au sein de l'objet (statiques : métadonnées stockées dans la structure de l'objet; dynamiques : métadonnées calculées dynamiquement à partir du contenu statique des objets, via le disséminateur de l'objet).
- Le contenu qui est exposé par OAI-PMH est caché, rendant possible des temps de réponse réduits ainsi que la disponibilité des données via OAI-PMH même lors d'un arrêt temporaire de FEDORA (service de base de FEDORA).

Note : FEDORA fournit une autre implémentation du protocole OAI-PMH au sein du FEDORA Core Service, le *Simple OAI-PMH Provider*. Cette implémentation n'est pas adaptée à nos besoins car elle est trop restrictive et simpliste. C'est pour cela que nous ne le détaillons pas dans ce rapport.

Politique de sécurité configurable et ajustable

En ce qui concerne la politique de sécurité de FEDORA, elle peut être configurée à différents niveaux.

Premièrement, nous pouvons définir une sécurité de base en ce qui concerne les accès au dépôt. Pour les deux interfaces d'accès, API-A (lecture des objets) et API-M (modification des objets), il est possible de définir si une authentification des utilisateurs est requise ou pas et si un certificat SSL est requis ou pas.

⁶⁰ Open Archives Initiative Protocol for Metadata Harvesting : <http://www.openarchives.org/OAI/openarchivesprotocol.html>

⁶¹ Disponible sous : <http://proai.sourceforge.net/>

En ce qui concerne l'authentification des utilisateurs, actuellement FEDORA permet de le faire via un annuaire LDAP ou via un fichier de configuration propre.

Dernièrement, FEDORA permet la création de règles d'accès personnalisés en se basant sur le langage XACML⁶². Une règle peut être définie au niveau du dépôt, donc applicable pour tous les objets y déposés, ou spécifique à un objet, donc applicable seulement pour cet objet.

12.1.2 Principales caractéristiques de l'interface FEZ

FEZ⁶³ est un système de gestion du contenu et du workflow d'un dépôt numérique basé sur FEDORA. Il a été conçu et implémenté par l'Université de Queensland (Australie) pour combler le manque d'interfaces graphiques pour gérer un dépôt FEDORA. La première version est sortie en 2005 et, avec la version 1.3 sortie au début de l'année, l'application est arrivée à un état stable. Actuellement, une équipe de développeurs continue à implémenter de nouvelles fonctionnalités. FEZ est disponible en Open Source.

Pour les utilisateurs avertis (professionnels bibliothécaires, administrateurs facultaires), nous proposons l'utilisation de FEZ comme interface de soumission, validation et administration du dépôt FEDORA pour les raisons suivantes :

1. Il s'agit d'une solution déjà implémentée et disponible en Open Source.
2. Elle fournit les fonctionnalités dont nous avons besoin pour soumettre, valider et rechercher des ressources numériques dans un dépôt FEDORA.
3. Elle fait preuve d'une grande souplesse : les workflows d'utilisation ainsi que les formats de métadonnées sont configurables.
4. L'installation par défaut inclut l'implémentation des standards DC et MODS pour décrire les ressources numériques.
5. La création des métadonnées de conservation est déjà intégrée dans le système. Lors du dépôt d'une ressource numérique, FEZ génère ses métadonnées de conservation à partir du contenu du fichier téléchargé en utilisant JHOVE.
6. La création des métadonnées relationnelles est aussi intégrée dans le système. Lors de la saisie d'un document au sein d'une collection, les métadonnées relationnelles de l'objet déposé contiendront une relation avec l'objet qui le contient.
7. L'authentification des utilisateurs peut être faite en utilisant des attributs stockés dans FEZ ou provenant d'un annuaire LDAP ou de Shibboleth.
8. Elle est implémentée en utilisant des technologies connues par le CI : php, mysql, Webservices, apache, Java.

⁶² eXtensible Access Control Markup Language : http://docs.oasis-open.org/xacml/access_control-xacml-2_0-core-spec-cd-04.pdf

⁶³ FEZ wiki : http://dev-repo.library.uq.edu.au/wiki/index.php/Main_Page

9. Diverses universités et instituts utilisent déjà FEZ dans leurs sites de production⁶⁴.
10. La communauté des utilisateurs devient de plus en plus grande. Les mailing lists (celle des utilisateurs et celle de développeurs) sont très actives.

12.2 Vue d'ensemble de l'architecture SERVAL

Comme mentionné dans le premier chapitre de ce rapport, SERVAL est surtout un projet d'intégration. Le schéma suivant montre la vue d'ensemble du système prévu, qui fait ressortir le rôle d'intégrateur du dépôt institutionnel (FEDORA) entre les systèmes existant actuellement à l'UNIL, BCU et CHUV.

⁶⁴ Université de Queensland : <http://espace.library.uq.edu.au/>
National Distance Education University (UNED) : <http://e-spacio.uned.es/fez/index.php>
Materials Digital Library Pathway : <http://matdl.org/fez/index.php>

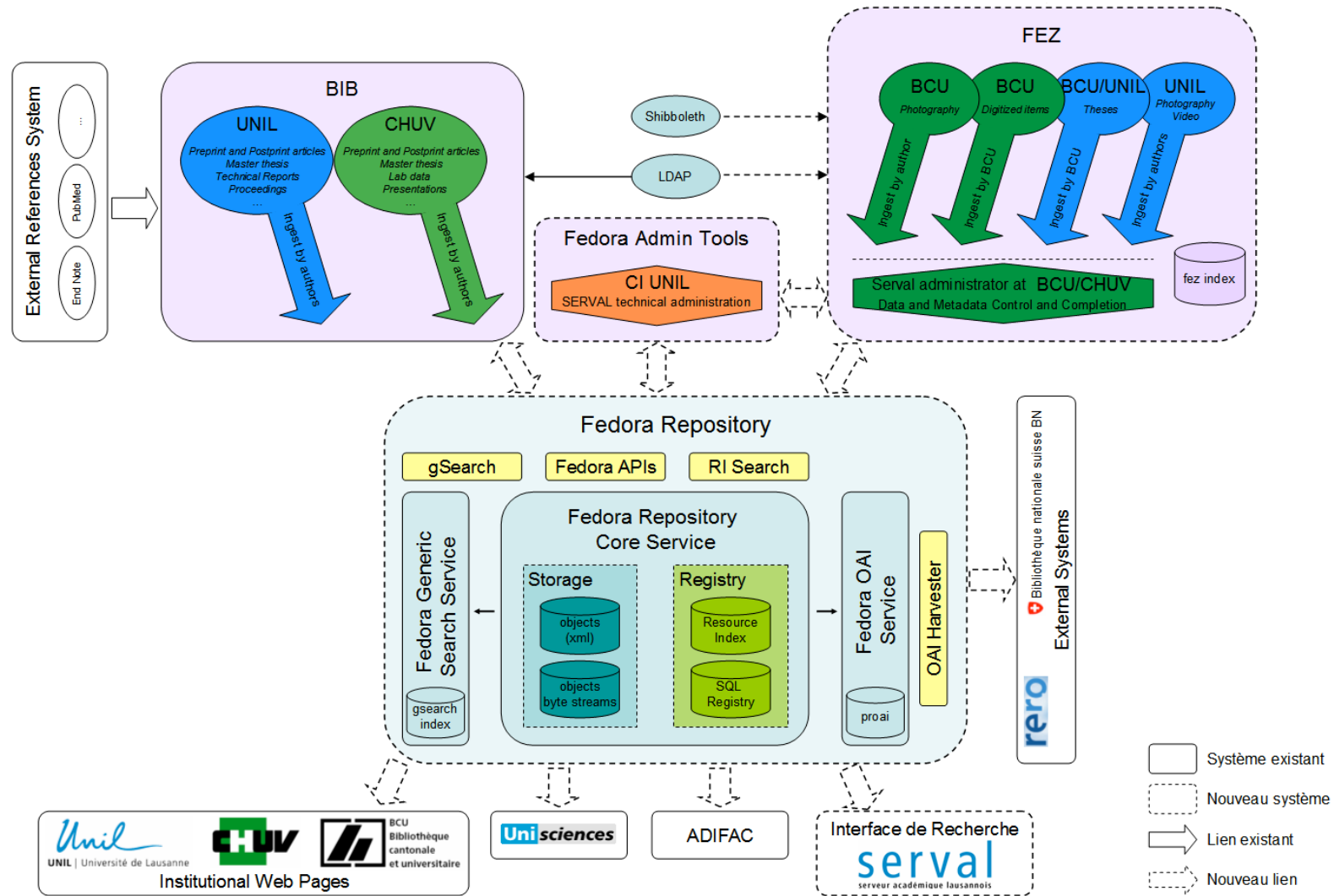


Fig. 16 - Vue d'ensemble du système SERVAL

12.3 Outils d'administration

Aussi bien FEDORA que FEZ fournissent, de manière native, des outils d'administration du dépôt. Selon les besoins, des outils additionnels pourront être développés en interne, comme pour l'importation massive d'objets depuis des systèmes externes (ex. thèses de RERO DOC, références d'Alexandrie).

12.3.1 FEDORA Admin Tools

L'outil d'administration fourni par FEDORA (*FEDORA Administration Tool*) est une application Java permettant: importer et exporter des objets ; ingérer, modifier et effacer des objets ; chercher des objets ; etc. La figure suivante montre l'interface graphique de cette application.

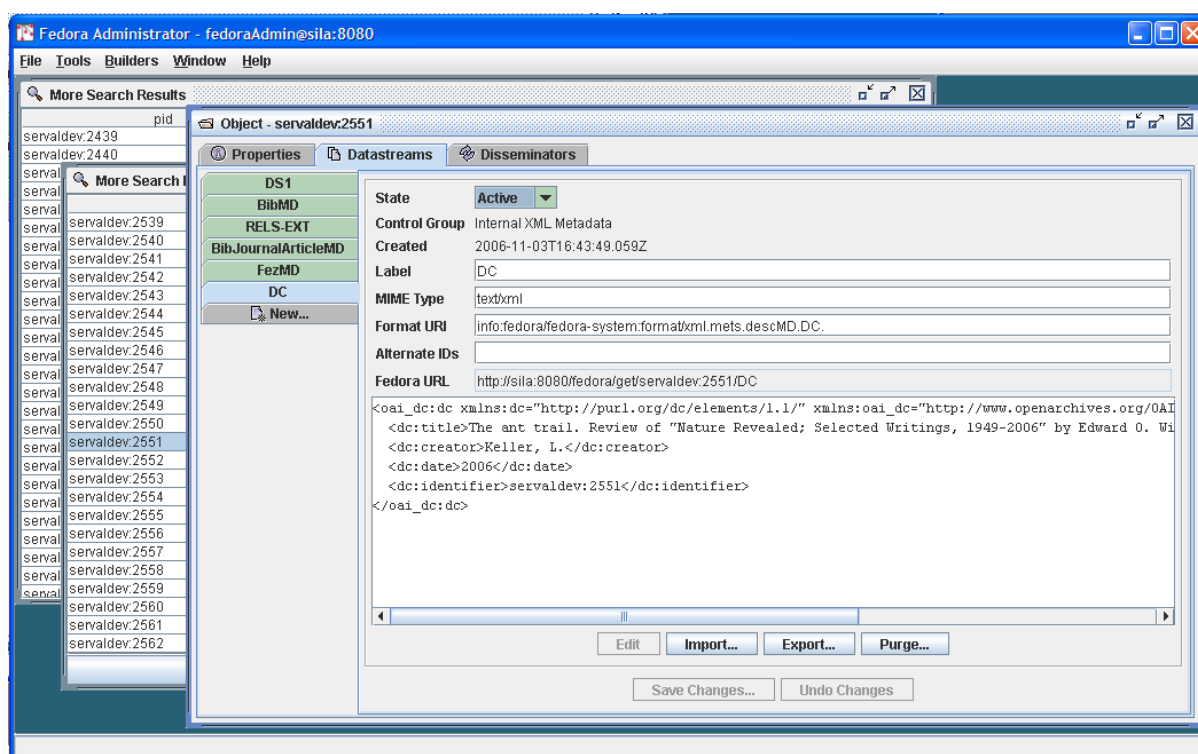


Fig. 17 – Interface d'administration FEDORA

12.3.2 FEZ Admin Tools

L'application FEZ contient une section, accessible depuis l'onglet "Administration" réservé aux administrateurs du dépôt, comme montrée dans la figure suivante où l'on voit les fonctionnalités disponibles.

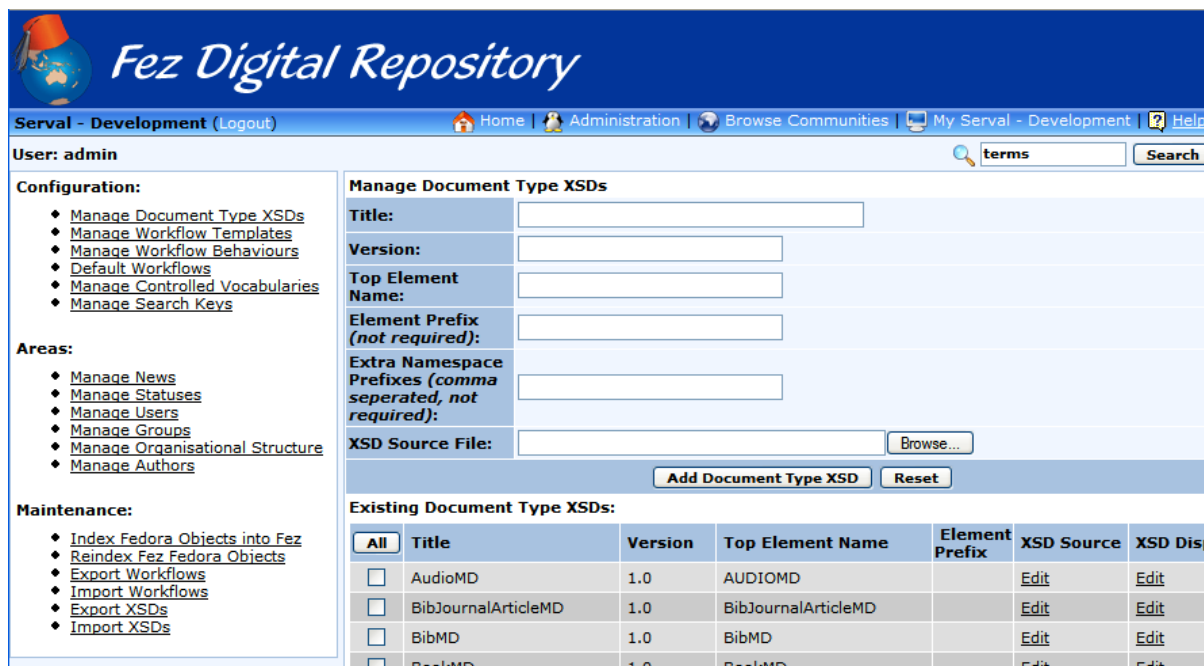


Fig. 18 – Interface d’administration FEZ

12.4 Interactions entre les différents composants du système et évolution

12.4.1 BIB <—> Dépôt institutionnel FEDORA

Comme on le voit dans la figure suivante, actuellement les informations relatives aux publications saisies via BIB sont stockées dans une base XML (eXist) pour les métadonnées et sur un serveur de fichiers (DocUnil) pour le contenu.

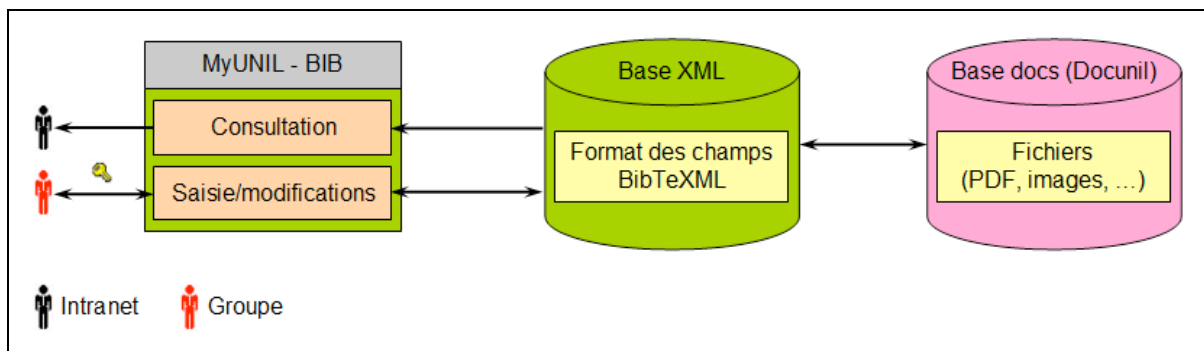


Fig. 19 - BIB: architecture actuelle

Dans le cadre du projet SERVAL, notre proposition consiste à remplacer la base XML et le serveur de fichiers par le dépôt institutionnel FEDORA, comme montré dans la figure suivante :

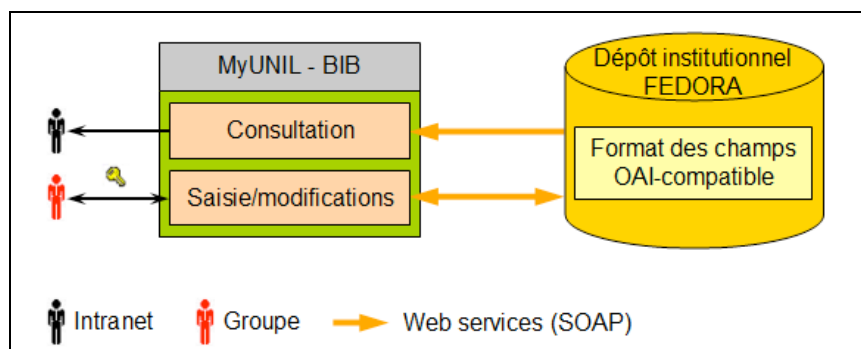


Fig. 20 - BIB: architecture prévue pour SERVAL

La communication entre l'application BIB et le dépôt FEDORA se fera au moyen de Webservices.

12.4.2 FEZ <--> Dépôt institutionnel FEDORA

La communication entre FEZ et FEDORA est aussi faite au moyen de webservices. Cette communication est déjà implémentée et intégrée dans FEZ.

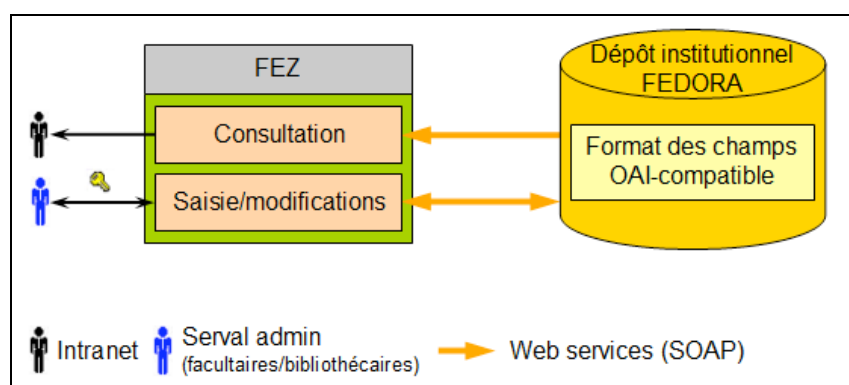


Fig. 21 - FEZ: intégration avec FEDORA

12.4.3 Dépôt institutionnel FEDORA <--> Unisciences

Depuis la dernière version d'Unisciences, les informations relatives aux publications scientifiques proviennent de la base de publications saisies depuis BIB.

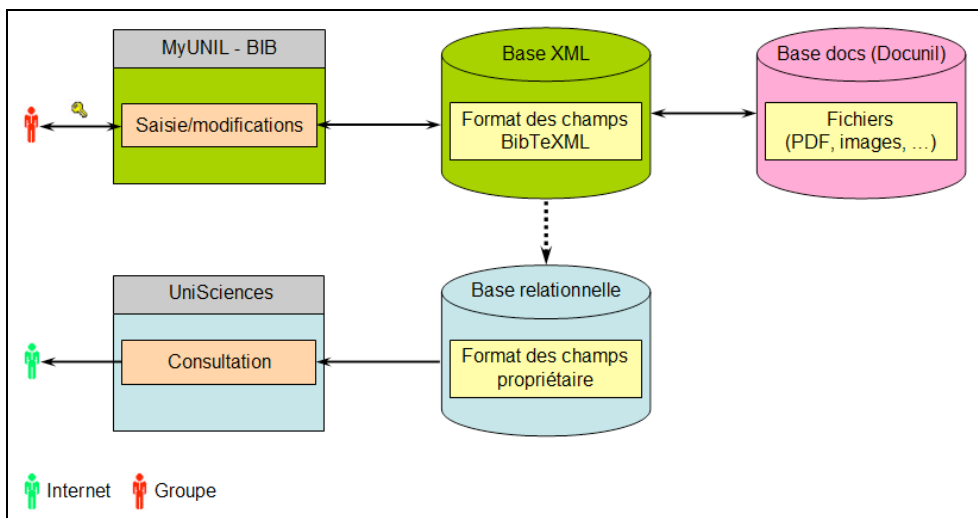


Fig. 22 - Unisciences: architecture actuelle

Avec le remplacement de la base XML et de Docunil pour le stockage des publications par le dépôt FEDORA, cette architecture évoluera vers celle montrée à la figure suivante:

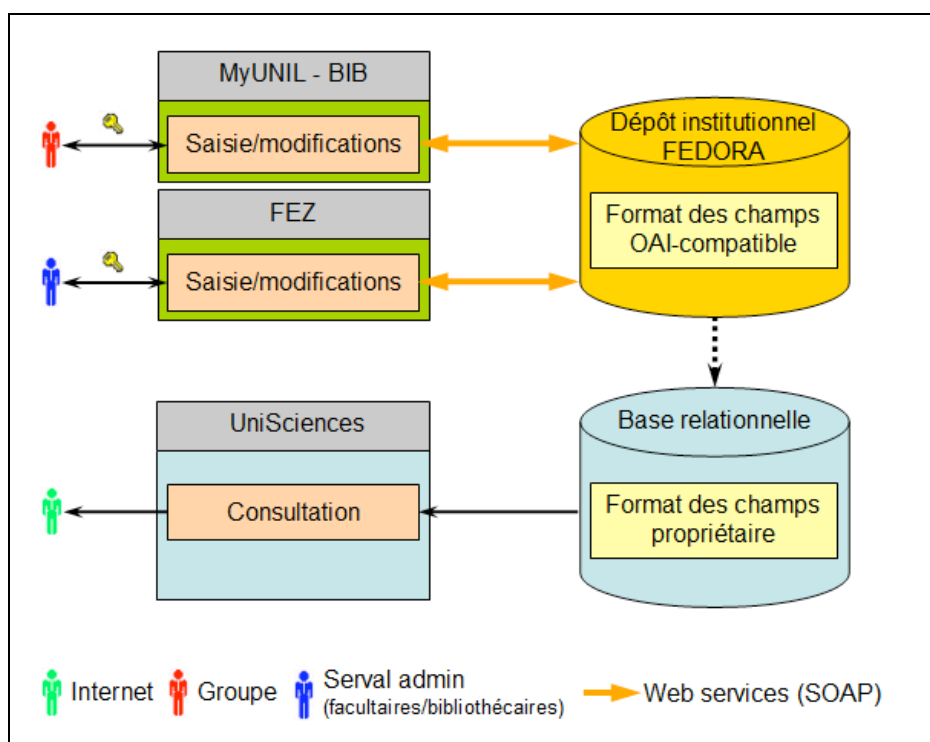


Fig. 23 - Unisciences: architecture prévue pour SERVAL

12.4.4 Dépôt institutionnel FEDORA <--> ADIFAC

Tout comme Unisciences, les publications insérées dans ADIFAC seront saisies via l'interface BIB et reportées depuis les bases de sauvegardes de celles-ci dans

la base ADIFAC. Avec l'évolution vers le dépôt institutionnel FEDORA, on retrouvera donc le même schéma d'architecture que pour Unisciences.

12.4.5 Dépôt institutionnel FEDORA <--> Sites Web institutionnels

Actuellement, les publications saisies via BIB sont automatiquement affichables sur les sites Web institutionnels utilisant l'outil de gestion Jahia.

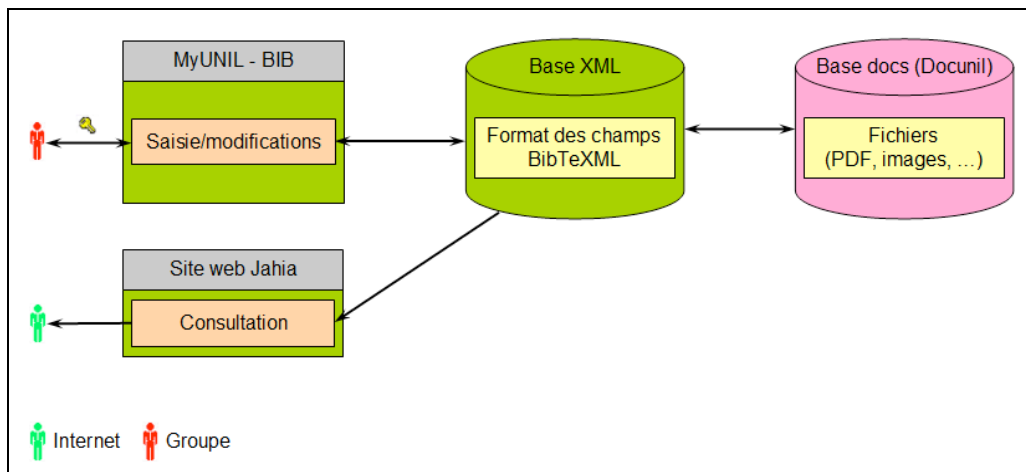


Fig. 24 - Pages institutionnelles: architecture actuelle

Avec le remplacement de la base XML et de Docunil pour le stockage des publications par le dépôt FEDORA, cette architecture évoluera vers celle montrée à la figure suivante:

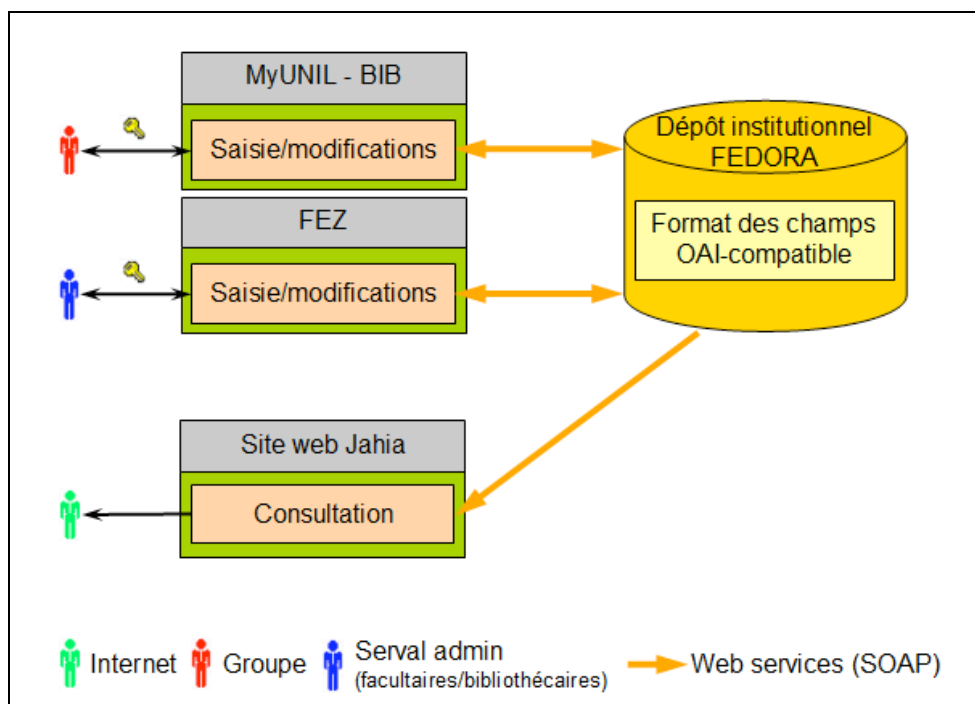


Fig. 25 - Pages institutionnelles: architecture prévue pour SERVAL

Pour les sites gérés avec un autre outil que Jahia, comme le sont ceux du CHUV, un développement est à prévoir par cette institution.

12.4.6 Dépôt institutionnel FEDORA <--> Interface de recherche SERVAL

Comme nous l'avons indiqué auparavant, nous n'avons pas encore spécifié les détails techniques de la nouvelle application de recherche qui sera développée. Cependant, ses interactions avec le dépôt se feront de la même manière que les autres composants, comme montré dans le schéma suivant :

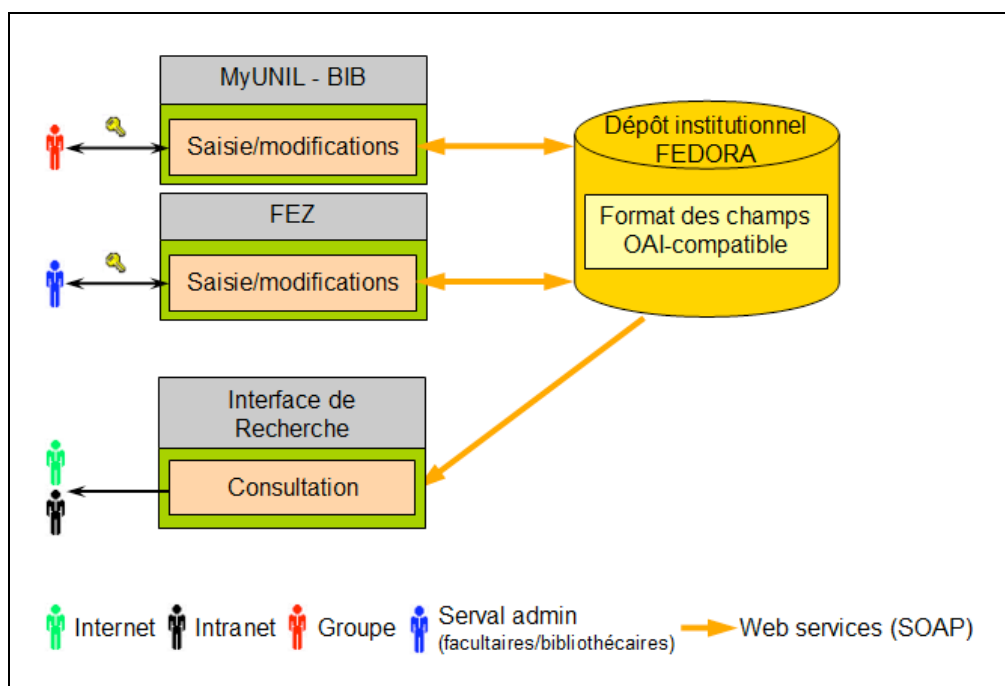


Fig. 26 - Interface de recherche: architecture prévue pour SERVAL

12.5 Données exposées par le dépôt

Dans le processus OAI de collecte automatique de données sur les dépôts institutionnels, ceux-ci décident des données exposées. FEDORA permet de configurer quelles métadonnées le seront ainsi que leur format d'exposition.

13 Modèle d'objets

Dans ce chapitre nous allons décrire le modèle d'objets pour le projet SERVAL. Nous allons commencer par décrire la structure d'objets FEDORA et ensuite nous décrirons comment celle-ci sera implémentée dans le cas SERVAL.

13.1 Structure d'un objet FEDORA

Un objet FEDORA possède la structure décrite à la figure suivante :

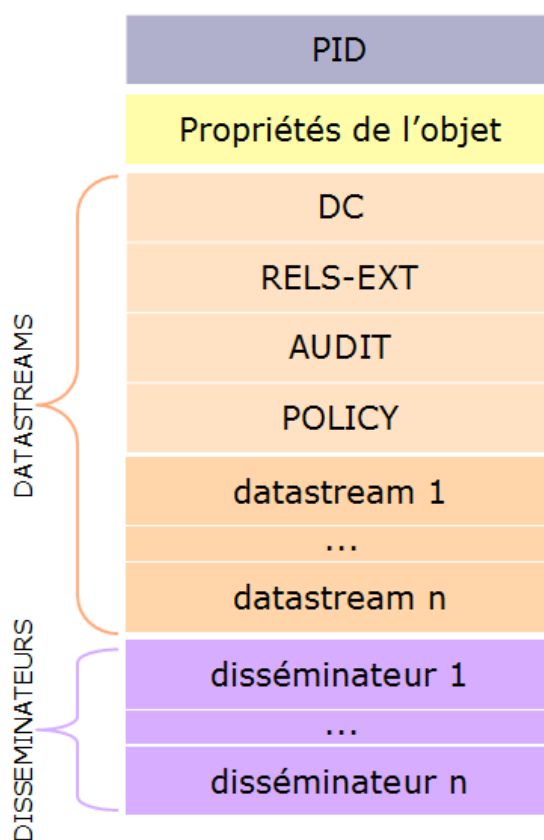


Fig. 27 - Structure d'un objet FEDORA

PID (Persistent Identifier) : il s'agit de l'identifiant unique de l'objet dans la base de donnée FEDORA. Il est obligatoire.

Propriétés de l'objet : il s'agit des propriétés système requises pour gérer les objets à l'intérieur du dépôt. Les propriétés les plus importantes sont l'état de l'objet au sein du dépôt (actif, inactif ou détruit), la date de création de l'objet et la date de la dernière modification sur l'objet. Ces propriétés sont aussi obligatoires.

Datastreams ou Flux de données : il s'agit des données qui représentent le contenu de l'objet. Le nombre de *datastreams* par objet est illimité.

Les caractéristiques principales d'un *datastream* sont les suivantes :

- Le format des données contenu dans un *datastream* est libre, mais doit correspondre à un MIME type⁶⁵ valide.
- Un *datastream* peut contenir des données représentant le contenu de l'objet (par exemple un fichier pdf contenant un article) ou bien des métadonnées décrivant l'objet (par exemple son titre, son abstract, etc.).
- Le contenu d'un *datastream* peut être stocké dans le dépôt FEDORA ou consister en une référence vers un contenu stocké dans une autre dépôt et accessible via une URL.
- Un *datastream* est versionable, ce qui veut dire qu'on peut garder différentes versions de son contenu.

Les *datastreams* suivants sont réservés par le système FEDORA :

- DC : conserve les métadonnées de l'objet en format Dublin Core. Il est obligatoire. Si ce *datastream* n'est pas fourni lors de la création de l'objet, FEDORA le crée automatiquement.
- AUDIT : conserve la liste de toutes les actions faites sur l'objet. Ce *datastream*, obligatoire, est créé et maintenu automatiquement par FEDORA.
- RELS-EXT : conserve les relations de l'objet propriétaire de ce *datastream* avec des autres objets du dépôt. Le contenu de ce *datastream* est codé en format XML en utilisant le standard *Resource Description Framework* (RDF). Ce *datastream* est optionnel.
- POLICY : conserve les règles d'accès spécifiques à l'objet. Son contenu est codé en XML en se basant sur le langage XACML. Il est aussi optionnel.

Disséminateurs : les disséminateurs d'un objet permettent d'associer l'objet FEDORA à un service externe qui peut transformer le contenu statique de l'objet (contenu dans les *datastreams*) pour le rendre dynamique. Par exemple, si nous avons un objet qui garde une image en format tiff, on peut associer un disséminateur à cet objet qui permettra de diffuser cette image originale en divers formats (jpeg, png, etc.), en diverses dimensions, diverses gammes de couleurs, etc.

13.2 Structure d'un objet SERVAL

Comme mentionné au chapitre 4, le contenu du dépôt SERVAL sera organisé dans des collections. Ces collections seront aussi stockées en tant qu'objets FEDORA dans le dépôt. Nous aurons donc, deux types d'objets :

- a) Les objets de contenu, représentant les ressources numériques contenues à l'intérieur des collections.
- b) Les objets "containers" représentant les collections.

⁶⁵ MIME Type : Multipurpose Internet Mail Extensions Type. Pour plus d'information : http://fr.wikipedia.org/wiki/Multipurpose_Internet_Mail_Extensions

13.2.1 Structure d'un objet de contenu

Un objet représentant une ressource numérique déposée dans SERVAL doit stocker différents types d'informations, au minimum le contenu de l'objet (un fichier pdf, une image, une vidéo, etc.), ses métadonnées (bibliographiques, relationnelles et de conservation) et des données d'administration de l'objet (son état dans le workflow, etc.).

Comme montré dans la figure suivante, un *datastream* spécifique sera dédié au stockage de chaque type d'information.

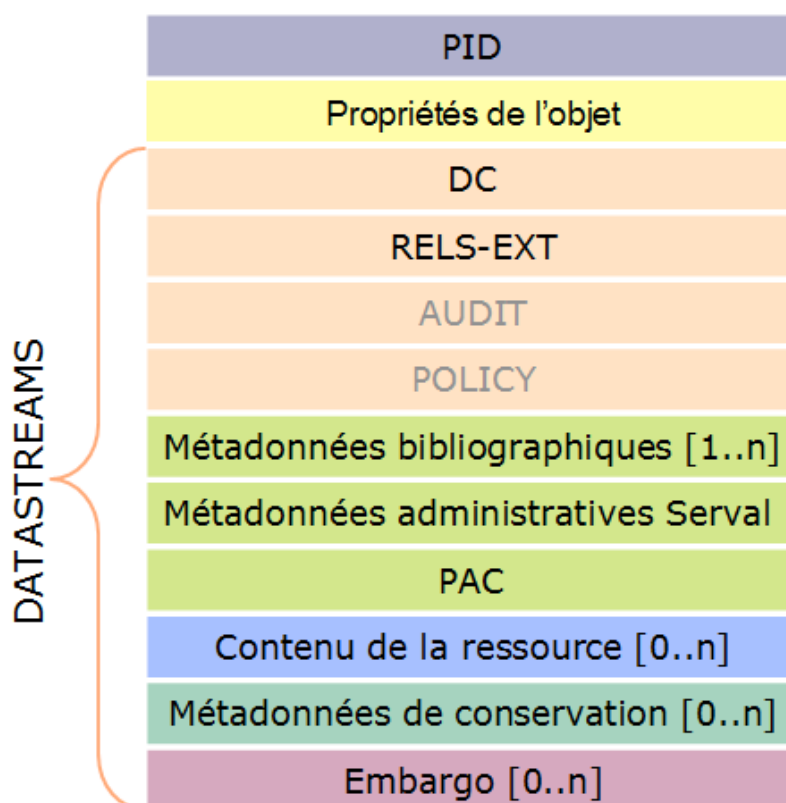


Fig. 28 - Structure du contenu numérique SERVAL

Ces *datastreams* spécifiques sont décrits ci-après.

Métadonnées relationnelles (RELS-EXT)

Cardinalité: 1

Format: XML

Ce *datastream* contiendra les relations de cet objet avec d'autres objets stockés dans le dépôt.

Métadonnées bibliographiques

Cardinalité: 1..n

Format: XML

Ces *datastreams* contiendront les métadonnées bibliographiques de l'objet. Comme mentionné auparavant, nous allons utiliser deux standards différents : BIBTEXML et MODS. Cependant, des attributs additionnels pourront être définis dans des *datastreams* supplémentaires.

Métadonnées administratives SERVAL

Cardinalité: 1

Format: XML

Ce *datastream* contiendra les attributs relatifs à l'administration de l'objet dans le dépôt SERVAL. Les éléments principaux sont :

- **Type de publication** : Cet élément contiendra le type de publication représenté par l'objet.
- **Responsable de la publication** : Cet élément contiendra les informations relatives à la personne qui a soumis la publication :

Son nom et prénom

Son adresse email

Le nom de l'application utilisé pour soumettre la publication : BIB ou FEZ

Étant donnée qu'une personne doit être authentifiée dans le système pour pouvoir saisir une publication, ces données pourront être automatiquement gérées par le système.

- **État de l'objet** : Cet élément gardera l'état courant de l'objet, sa valeur correspondant aux états définis dans le workflow.
- **Champ de l'objet** : L'information contenue dans cet élément permettra de savoir si cet objet fait partie de la production interne SERVAL (UNIL, BCU, CHUV/BiUM) ou externe.

Cette valeur doit être fournie par les personnes qui soumettent le document dans le dépôt. Dans le cas d'une soumission via BIB, cette valeur sera demandée lors du passage dans la zone Internet.

- **Harvestabilité** : Cet élément contiendra une valeur booléenne indiquant si l'objet est harvestable ou non.

Pour les documents soumis via BIB, cette valeur sera automatiquement déduite par le système (en fonction du champ de l'objet et de son contenu).

Pour les documents soumis via FEZ, cette valeur devra être fournie par la personne qui les soumet.

- **Référence du système externe** : Cet élément sera utilisé lors de l'importation massive des objets. L'idée est de garder dans cet élément l'identifiant de l'objet dans le système externe d'où il a été importé. Par exemple, lors de l'importation des données depuis Virtua, il s'agira de l'identifiant "Virtua".

PAC

Cardinalité: 1

Format: XML

Ce *datastream* contiendra les informations relatives au PAC (Preservation And Conservation). Par défaut, lorsqu'un objet est soumis à SERVAL, une valeur de PAC lui sera assignée en fonction du type de publication. Cependant, les spécialistes bibliothécaires pourront modifier cette valeur lors de la validation de la publication.

Contenu de la ressource

Cardinalité: 0..n

Format: MIME Type

Le contenu de la ressource peut être un fichier téléchargé au dépôt ou bien un URL vers un contenu dans un autre serveur.

Métadonnées de conservation

Cardinalité: 0..n

Format: XML

Pour chaque fichier qui a été téléchargé dans le dépôt, le système créera automatiquement ses métadonnées de conservation en utilisant JHOVE.

Embargo

Cardinalité: 0..n

Format: XML

Pour chaque *datastream* de contenu, l'auteur devra spécifier s'il y a ou pas et combien de temps un embargo sur le document téléchargé.

13.2.2 Structure d'un objet container

La structure d'un objet qui représente une collection est plus simple que la structure d'un objet représentant une ressource numérique. D'une part, il n'y a pas du contenu stocké au niveau de ces objets: le contenu est stocké dans les objets qui appartiennent à la collection, mais non dans l'objet qui représente la collection. D'autre part, étant donné que la hiérarchie des collections est statique, il ne faut pas gérer des changements d'état pour ces objets. Il peut y avoir des changements ponctuels, mais il s'agira principalement d'ajustements de structure faites par l'administrateur du dépôt. Sa structure est illustrée dans la figure suivante :

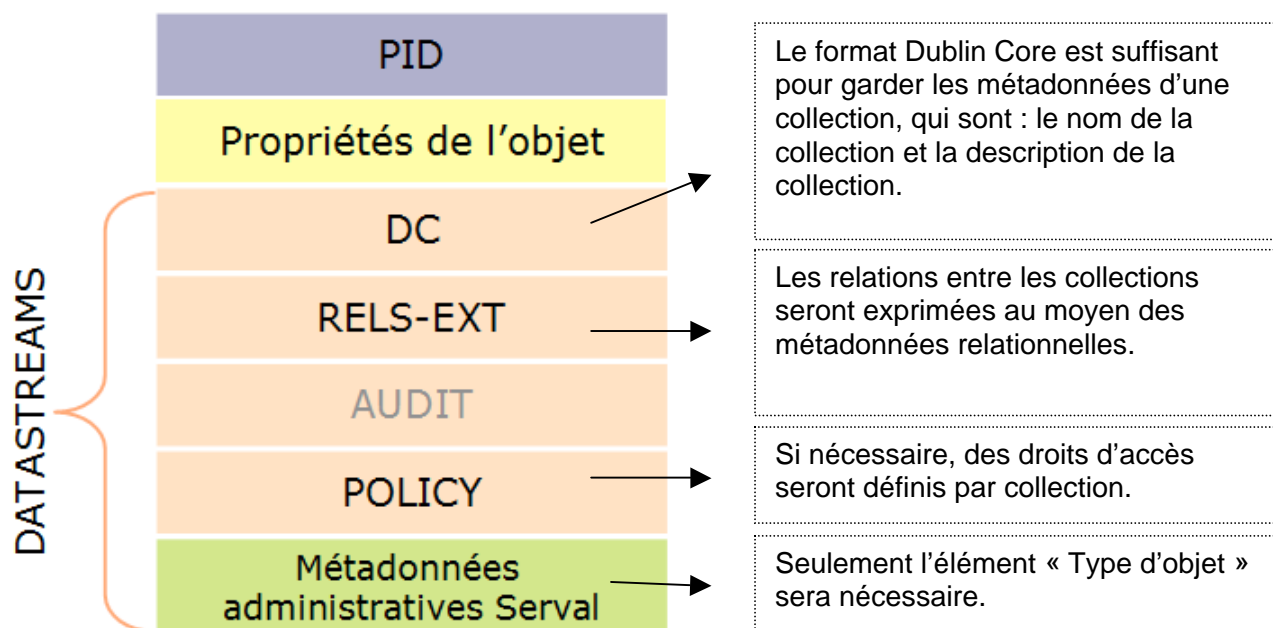


Fig. 29 - Structure d'une collection SERVAL

14 Communication du projet

14.1 *Identité SERVAL*

Unicom a élaboré une identité graphique pour le projet SERVAL, identité compatible avec le graphisme de l'UNIL, intégrant les logos des partenaires et déclinable sur les différents supports nécessaires pour les présentations, publications, sites Web etc. SERVAL. Le présent rapport utilise déjà une partie de cette ligne graphique. Le logo SERVAL est créé sur la base d'une police *Frutiger*, l'écriture pour les textes est le *Verdana*. Quelques exemples :



Fig 30 - Le logo SERVAL proposé par UNICOM, tiré du concept graphique UNIL



Fig. 31 -Modèle pour PPT et posters SERVAL

Le logo figurera sur le site Web SERVAL, ainsi que dans l'en-tête des interfaces FEZ. Unicom nous livre également les modèles pour PPT et Word pour les rapports, communications et présentations. L'identité graphique SERVAL sera utilisée pour toute communication du projet et restera en place sur les diverses interfaces une fois le projet terminé.

14.2 *Communication en phase Réalisation*

14.2.1 Communication interne UNIL/BiUM/BCU

En phase de réalisation, seuls les groupes-pilotes et les personnes participant au développement du serveur ont accès à SERVAL. Une fois le concept accepté, une présentation complète doit en être faite à la Direction UNIL, dans le but de bien communiquer l'ampleur et les limites du projet actuel, les possibilités qui sont couvertes par le projet dès 2008 et quelles possibilités s'ouvriraient en cas de continuation du projet dans une deuxième étape. La direction UNIL, en tant que méta-utilisateur de SERVAL – un des publics cible – doit être parfaitement renseigné sur les fonctionnalités auxquelles elle aura accès et quels

renseignements SERVAL peut fournir et dans quels délais. Le concept devra également être communiqué aux responsables du projet ADIFAC au CHUV par une présentation en profondeur.

De même, le concept doit être présenté dans les deux bibliothèques partenaires aux bibliothécaires universitaires, resp. aux bibliothécaires chargés du contact avec les chercheurs. La BCU veillera à informer tous les sites (Dorigny, BDSE, Riponne), et prendra contact avec les responsables des bibliothèques du campus de Dorigny, comme la bibliothèque de Biologie, des Sciences de la Terre etc. La BiUM veillera à informer les bibliothèques biomédicales du site CHUV.

Une fois la direction UNIL informée, des présentations du concept auront lieu dans les séances de Facultés – dans toutes les Facultés UNIL. Comme à ce stade les groupes-pilotes sont déjà actifs, certaines facultés auront déjà une vision plus concrète que d'autres (discours à adapter). Ces présentations devraient être faites par des responsables du projet et les bibliothécaires responsables du domaine académique correspondant à la faculté.

Au niveau des groupes de chercheurs, la communication se limite durant la phase de réalisation aux groupes-pilotes. Un premier pas dans cette direction a été entrepris au moyen de l'enquête. Les résultats de l'enquête sont tout d'abord rassemblés par les interviewers, puis mis en commun avec les autres bibliothécaires facultaires afin d'évaluer les similitudes et les différences entre les quatre groupes-pilotes. Les interviewers pourront ensuite donner un feedback au groupe de chercheurs et instaurer un discours régulier sur leurs expériences avec SERVAL.

Dès l'acceptation du concept, ce dernier sera également présenté dans son entier aux groupes-pilotes. Il est important que ces testeurs soient bien au courant des tenants et des aboutissants du projet, afin de faciliter la communication ultérieure et d'assurer la qualité de leur feedback. Pour la durée de la phase Réalisation, il sera utile de fixer avec chaque groupe des objectifs de saisie de données ainsi que d'analyse des processus (vérification des workflows).

Durant la phase de réalisation, les informations doivent pouvoir circuler très rapidement entre les différents partenaires. Le principal canal de communication entre les développeurs et les administrateurs sera la liste « serval_geeks ». Une liste « serval_users » pourra être créée pour les besoins de communication avec les membres des groupes-pilotes et de manière plus générale avec les chercheurs utilisant SERVAL.

À la BCU et à la BiUM, les premiers administrateurs SERVAL seront formés durant la phase Réalisation. La BCU engagera un(e) administrateur/trice SERVAL durant cette phase. Le responsable de l'atelier de photographie et les responsables des projets de numérisation recevront des accès sur SERVAL, procéderont à des saisies, vérifieront le bien-fondé des workflows et les affineront. À la BiUM, un administrateur SERVAL CHUV sera nommé pour la partie CHUV/BiUM.

Un autre volet de communication est l'intégration de SERVAL dans les programmes de formation donnés par les bibliothécaires. Le projet sera donc présenté aux responsables de formation de la BCU et de la BiUM, ainsi qu'aux ingénieurs pédagogiques responsables du côté UNIL de la mise en place de la ICT Driving License (si ce dernier voit le jour). SERVAL sera intégré dans les formations autant du point de vue de la recherche de documents (formation de base) que du dépôt de document (formation pour les étudiants avancés).

Durant la phase Réalisation, les avancées du projet sont également communiquées par des articles dans *i-Ci* (Ci) et *Les sentiers de l'info* (BCU), ainsi que par des présentations lors de la COBIB et de la CODUL.

14.2.2 Communication externe

Le site Web SERVAL sera non seulement actualisé mais également restructuré en fonction du Concept. Il contiendra de nouveaux thèmes et s'orientera déjà vers les questions liées à l'utilisation pratique de SERVAL.

Comme auparavant, le rapport de Concept sera envoyé aux professionnels intéressés, tout particulièrement nos voisins de l'arc lémanique (en particulier l'UNIGE) et les responsables de projets semblables et/ou complémentaires à la Confédération (BN, Consortium des Bibliothèques Suisses, projet e-Archiving, RERO).

Le projet sera brièvement présenté aux conférences traitant de sujets technologiques semblables (p.ex. OAI5 au CERN).

Durant la phase Réalisation, on présentera le projet SERVAL à RERO et à la BN, afin de discuter des modalités de harvesting, de l'avancement de la question des URN, de la collaboration pour l'archivage pérenne (quels documents et métadonnées sont dupliqués sur quels serveurs).

Durant la phase Réalisation, une discussion devra avoir lieu avec les autorités politiques vaudoises concernant la question du e-Archiving du canton de Vaud.

14.3 Communication en phase Introduction

14.3.1 Communication interne

Dès la phase Introduction, tous les membres de la communauté UNIL ont potentiellement accès à SERVAL. Le grand public a également accès aux interfaces de recherche. Une fois le concept réalisé, une présentation du produit fini sera faite à la Direction UNIL et aux responsables du projet ADIFAC au CHUV.

Du côté des bibliothèques, l'utilisation de SERVAL sera d'abord expliquée à tous les bibliothécaires universitaires, resp. aux bibliothécaires responsables de contacts avec les chercheurs. La BCU se chargera de contacter toutes les

bibliothèques sur le campus de Dorigny, la BiUM se chargera de contacter toutes les bibliothèques du site CHUV. Tous les bibliothécaires universitaires doivent devenir des super users SERVAL.

L'enquête pourra être élargie à l'ensemble de l'UNIL au moyen d'un formulaire en ligne. Les résultats de l'enquête pourront être publiés sur le site SERVAL ainsi que dans UNISCOPE, ce qui permettra de rendre la communauté UNIL attentive à la mise en œuvre de SERVAL.

Une présentation standard du produit SERVAL sera mise au point par le Ci et la BCU, présentation qui pourra être reprise par les bibliothécaires universitaires.

Ensuite, SERVAL sera présenté à tous les groupes de chercheurs UNIL/CHUV. Ces présentations nécessiteront un grand effort, d'où la nécessité de bien former tous les bibliothécaires universitaires auparavant, afin qu'ils soient capables de remplir cette tâche.

Comme tous les chercheurs auront accès à SERVAL, une helpline devra être mise en place et les administrateurs SERVAL devront être prêt à entrer en action pour soutenir les chercheurs dans le dépôt des documents. La liste « serval_geeks » restera réservée aux développeurs/administrateurs pour discuter des problèmes particuliers.

Dès la rentrée 2008, la formation concernant SERVAL sera intégrée aux programmes de formation donnés par les bibliothécaires, ainsi qu'au concept de la ICT Driving License UNIL/BCU. SERVAL sera intégré dans les formations autant du point de vue de la recherche de documents (formation de base) que du dépôt de document (formation pour les étudiants avancés).

Durant la phase Introduction, les avancées du projet seront également communiquées par des articles dans i-Ci (Ci) et *Les sentiers de l'info* (BCU), ainsi que par des présentations lors de la COBIB et de la CODUL. L'introduction de SERVAL pourra également faire l'objet de grandes présentations dans UNISCOPE, l'Auditoire, etc.

14.3.2 Communication externe

Le site Web SERVAL deviendra la page d'accès au serveur SERVAL (Web application). La navigation sera donc complètement refaite pour proposer des informations générales quant au serveur académique, ce que l'on peut y trouver, les politiques de dépôt, de format, d'archivage, PAC etc.

Comme auparavant, le rapport d'Introduction sera envoyé aux professionnels intéressés, tout particulièrement nos voisins de l'arc lémanique (en particulier l'UNIGE) et les responsables de projets semblables et/ou complémentaires à la Confédération (BN, Consortium des Bibliothèques Suisses, projet e-Archiving, RERO).

Un effort soutenu sera fait pour faire connaître le serveur sur Internet : mise au point du harvesting par RERO DOC, la BN, Google ; présentation de SERVAL sur OpenDOAR, ROAR et d'autres sites qui sont utiles pour nous positionner de manière optimale sur l'Internet.

Le projet sera présenté aux conférences traitant de sujets technologiques semblables. Nous espérons que FEDORA connaîtra un développement positif en Suisse et en Europe. Des contacts ont été pris pour mettre en place une conférence européenne des usagers de FEDORA.

Durant la phase Introduction, la question du e-Archiving du canton de Vaud devra être suffisamment avancée pour proposer un début de solution, et des contrats signés pour les coopérations avec RERO DOC et la BN.

14.4 *Communication en phase Finalisation*

En phase finalisation, tous les membres de la communauté UNIL ont accès à SERVAL et les chercheurs et les bibliothécaires savent parfaitement utiliser le produit. Le grand public a accès aux interfaces de recherche.

La phase Finalisation est le moment de faire le bilan du projet et d'en tirer les enseignements. C'est un moment plutôt introspectif pour les personnes qui ont été sollicitées durant les phases précédentes, les nouveaux utilisateurs n'étant pas encore en mesure de donner un feedback consistant. Le cas échéant, c'est également le moment d'esquisser les objectifs de développement ultérieurs.

14.4.1 Communication interne

A la direction UNIL, aux facultés, aux bibliothèques, et au Ci on présentera les résultats de l'enquête globale, dans le cadre de grandes présentations incluant des possibilités de discussion (à la BCU p.ex. dans le cadre d'un jeudi de l'info). On y fera également un bilan des premiers contenus SERVAL et des premières réactions internes et externes.

14.4.2 Communication externe

Le site Web SERVAL aura pris son aspect final (on ne dira pas définitif, car rien sur le Web n'est jamais définitif). L'administrateur SERVAL procédera au contrôle de positionnement sur Internet et prendra le cas échéant des contacts avec des partenaires comme Google pour optimiser les résultats.

Comme auparavant, le rapport de Finalisation sera envoyé aux professionnels intéressés, tout particulièrement nos voisins de l'arc lémanique (en particulier l'UNIGE) et les responsables de projets semblables et/ou complémentaires à la Confédération (BN, Consortium des Bibliothèques Suisses, projet e-Archiving, RERO). Si d'autres usagers en Suisse se lancent dans l'utilisation de FEDORA,

nous créerons un groupe FEDORA suisse, dans lequel le site de Lausanne pourrait jouer un rôle intéressant.

Au niveau des conférences, on peut envisager de faire une bonne présentation SERVAL une fois le serveur en place. Beaucoup de projet de ce type sont déjà en exploitation, nous n'avons donc pas le privilège de pouvoir présenter un projet révolutionnaire au moment de la réalisation (notre objectif étant la réalisation d'un produit d'avenir basé sur des technologies éprouvées, et non de créer un système expérimental). La technologie utilisée est cependant suffisamment intéressante pour mériter d'être présentée dans un cadre adéquat, réunissant des professionnels des systèmes d'information ou du Web sémantique.

Durant la phase Introduction, la question du e-Archiving du canton de Vaud devra être réglée et des contrats signés pour les coopérations avec RERO DOC et la BN.

Planing de la communication du projet SERVAL

	Mai.07	6.2007	7.2007	8.2007	Sept.07	10.2007	11.2007	12.2007	1.2008	2.2008	3.2008	4.2008	5.2008	6.2008
Phase/ public				Réalisation						Introduction			Finalisation	
UNIL Direction et Facultés		Présentation Concept			Présentation Concept aux Facultés	Présentation Concept aux Facultés	Présentation Concept aux Facultés		Présentation SERVAL en production					
UNIL/CHUV chercheurs	Enquête	Présentation Concept		Présentation Concept aux ingénieurs pédagogi- ques			Feedback des chercheurs pour rapport Réalisation		Mise en place de la helpline	Présentation SERVAL en production	Enquête en ligne			Présentation des résultats de l'enquête UNISCOPE
BCU	Evaluation de l'enquête auprès des groupes- pilotes. Sentiers de l'info	Présentation Concept.		Formation resp. Numérisation et resp. Photographie		Formation de l'administrat eur SERVAL				Présentation SEVAL en production. Sentiers de l'info			Contrôle du positionnement sur Internet	Discussion des premiers résultats
CHUV		Présentation Concept à ADIFAC								Présentation SERVAL en production				
BiUM	Evaluation de l'enquête auprès des groupes-pilotes	Présentation Concept				Formation de l'administrat eur SERVAL				Présentation SERVAL en production				Discussion des premiers résultats
Externe		Envoi du Concept aux partenaire régionaux/ nationaux						Site Internet SERVAL devient Accès public SERVAL. Déclaration de SERVAL sur OpenDOAR, Google etc.	Envoi du rapport Réalisation aux partenaire régionaux/ nationaux			UNISCOPE, Auditoire, Fréquence Banane	Envoi du rapport Introduction aux partenaire régionaux/ nationaux	Envoi du rapport Finalisation aux partenaire régionaux/ nationaux

15 Risques encourus et problèmes connus

15.1 Montée en charge

Durant la phase Concept, peu de tests de montée en charge ont pu être effectués. Il reste des incertitudes quant à la réaction du serveur en cas de charge massive, p.ex. par l'import des fichiers TIFF de la numérisation 4DigitalBooks de la BCU. En principe, les TIFF ne seront pour cette raison pas déposés dans SERVAL, mais référencés, les fichiers se trouvant sur un serveur à bandes.

15.2 Insécurités actuelles du e-Archiving

Actuellement, SERVAL ne remplit pas les normes OAIS pour le e-archiving. Des discussions doivent être poursuivies concernant la mise en place d'une infrastructure et/ou la participation à une organisation permettant d'assurer l'archivage pérenne.

15.3 Méthode collaborative de l'acquisition des documents

Les processus collaboratifs d'acquisition des documents présentés au chapitre 10 doivent encore être testés et probablement affinés.

15.4 Serveurs peu fréquentés - obligation de déposer

Le problème des serveurs peu fréquentés est très présent dans la littérature. SERVAL intègre toute une palette d'outils qui devraient permettre d'optimiser le résultat : facilité d'importer documents et métadonnées, engagement d'un administrateur poussant les chercheurs au dépôt et effectuant de nombreuses saisies lui-même. Une réforme des règlements de facultés en direction d'un dépôt obligatoire doit tout de même être envisagée.

15.5 Falsification de documents en format Word

Tout utilisateur déposant sur SERVAL sera authentifié. La modification d'un document ne sera donc possible que par des personnes ayant accès en écriture, donc les auteurs.

16 Budget

Étant donné le financement partiel du projet, les fonds débloqués par le Rectorat pour l'année 2007 sont engagés pour le poste d'informaticienne spécialiste au Ci, occupé depuis le 1^{er} octobre 2006 par Mme Urcelay, ingénieure en informatique EPFL. La BCU créera dès septembre 2007 un poste d'Administrateur central SERVAL. Dès janvier 2008, le solde du crédit du Rectorat après paiement du salaire de Mme Urcelay sera engagé dans une aide à la saisie de métadonnées et documents, afin de présenter une collection déjà bien achalandée dès la mise en service de SERVAL.

16.1 Personnel

Bleu : financé par le crédit extraordinaire de la Direction UNIL

Vert : financé par la BCU

Violet : financé par la BiUM

Rose : financé par le budget ordinaire du Ci

Fonction	Financé par	Charges	Coût de base par année	Coût 2007
Coordinateur technique	Crédit extraordinaire accordé par la direction de l'UNIL	1 x 100 %	CHF 110'000 par année	CHF 110'000
Appui à la saisie des métadonnées	BCU	2 x 50 %	CHF 60'000 par année (dès octobre 2007)	CHF 15'000
Cheffe de projet	BCU	40 % JF	Env. CHF 40'000 par année	CHF 40'000
Coordinatrice du groupe BCU	BCU	20 % CB	Env. CHF 20'000 par année	CHF 20'000
Administrateur des contenus SERVAL	BCU	80-100%	Env. CHF 100'000 par année (dès octobre 2007)	CHF 25'000
Bibliothécaires	BCU	env. 40 % autres BU	Env. CHF 40'000 par année (dès juillet 2006)	CHF 20'000

Coodination du groupe CHUV	BiUM	Env. 20 %	Env. CHF 20'000	CHF 20'000
Appui à la saisie des métadonnées pour le CHUV	BiUM	4 x 20 %	Env. CHF 80'000 par année	CHF 80'000
Conseil en droit d'auteur	Mandat existant à l'EPFL dont nous pourrions ponctuellement profiter	-	CHF 10'000 par année (dès 2008)	CHF 0
			CHF 480'000/an	CHF 330'000

16.2 Matériel

Type de matériel	Financé par	Coût total	Coût 2007
Serveurs	Centre Informatique	CHF 50'000/an	CHF 50'000
Sauvegardes	Centre Informatique	CHF 25'000/an	CHF 25'000
Maintenance	Centre Informatique	CHF 25'000/an	CHF 25'000
		CHF 100'000/an	CHF 100'000

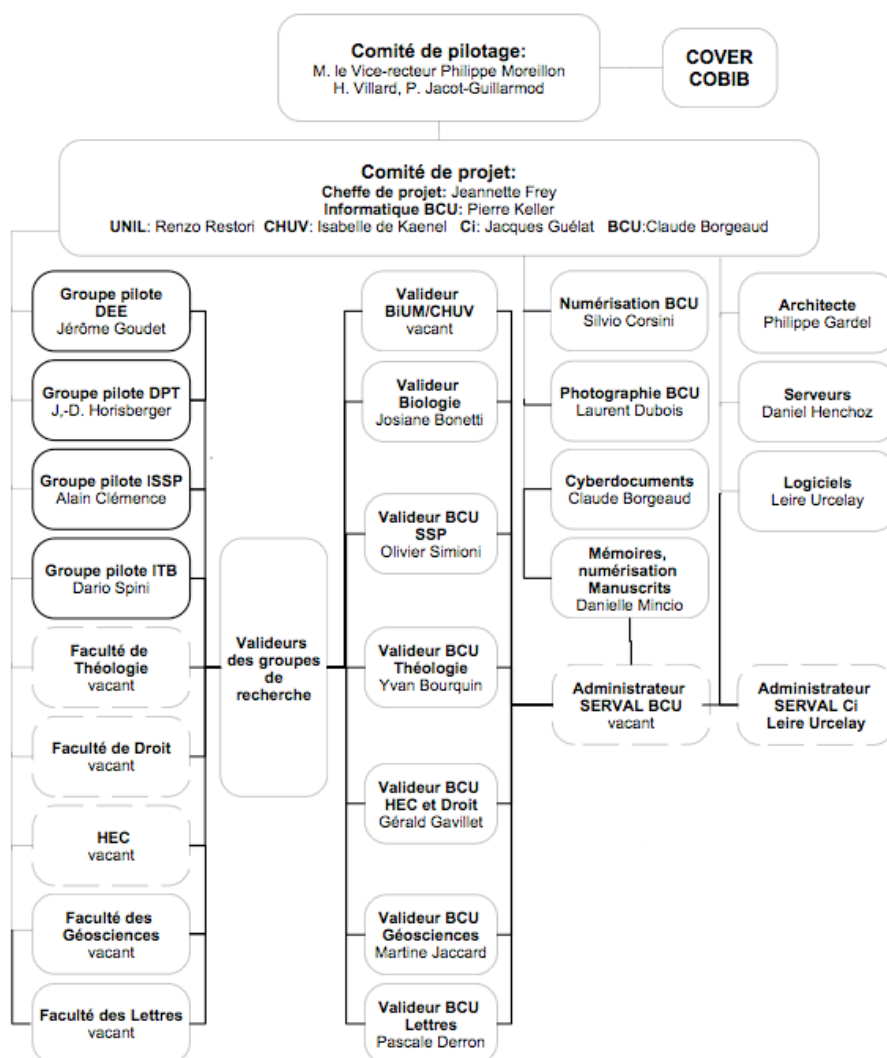
16.3 Logiciels

Type de logiciel	Financé par	Coût total	Coûts 2007
Open Source	-	-	-
Support	Centre Informatique	CHF 20'000/an	CHF 20'000
		CHF 20'000/an	CHF 20'000

Coût total		CHF 600'000/an	CHF 450'000
------------	--	----------------	-------------

17 Organigramme

Durant la phase Réalisation, les groupes-pilotes sont déjà opérationnels, et les responsables de la numérisation BCU le sont également. Les groupes-pilotes désigneront des validateurs dans leur domaine. Les bibliothécaires-validateurs à la BCU, à la bibliothèque de Biologie et à la BiUM participeront également aux tests. L'administrateur SERVAL BCU sera engagé en automne, un administrateur SERVAL BiUM nommé. Pour la phase Réalisation, nous proposons donc l'organigramme suivant:



18 Planification



18.1 *Réalisation (implémentation)*

Si le rapport de Concept est accepté, le Ci écrira le cahier de spécification du système d'après le présent rapport, le réalisera et le mettra en exploitation pour les groupes-pilotes et les responsables de la numérisation BCU.

Les groupes-pilotes, les bibliothécaires en charge des projets de numérisation ainsi que le photographe de la BCU seront formés à l'utilisation des interfaces BIB resp. FEZ et les utiliseront pour saisir des métadonnées et déposer des documents en guise de test.

L'enquête auprès des membres des groupes-pilotes sera terminée, les résultats seront mis en commun et analysés. Suivant les résultats, on développera un formulaire pour une enquête en ligne tout-UNIL.

La BCU recrutera un administrateur SERVAL. L'administrateur SERVAL commencera son travail en alimentant SERVAL avec les métadonnées et les documents des thèses et les métadonnées des anciennes publications UNIL. Un administrateur SERVAL sera également nommé à la BiUM.

Le Concept SERVAL sera d'abord communiqué à la Direction UNIL, aux responsables d'ADIFAC, ainsi qu'à la BiUM et à la BCU. Par la suite, il sera également présenté à toutes les facultés UNIL, à la COBIB et à la CODUL. Le Concept sera également communiqué aux responsables de la formation dans les bibliothèques partenaires et aux ingénieurs pédagogiques UNIL, ainsi qu'aux partenaires régionaux et nationaux.

Pour approfondir la question du e-archiving, nous passerons en revue l'*Audit Checklist* d'OCLC et CRL.

18.2 Rappel en grandes lignes des phases ultérieures du projet

18.2.1 Introduction

Si le rapport de Réalisation donne satisfaction, le Ci et la BCU procéderont début 2008 à la formation des collaborateurs de la BCU et de la BiUM à l'utilisation du système, pour être en mesure à la rentrée 2008 de proposer à la communauté scientifique lausannoise un système en production. Des manuels complets sont rédigés à l'intention des collaborateurs chargés de la maintenance du serveur, de l'acquisition des métadonnées et des documents.

Durant la phase d'introduction, la possibilité de profiter du serveur institutionnel et les formations *ad hoc* seront communiquées par tous les canaux de communication existant sur le campus (cours d'introduction, communiqués de presse, séances etc.)

Le système en production ainsi que les processus d'acquisition de documents et de métadonnées seront décrits dans un rapport remis au comité de pilotage début 2008.

18.2.2 Finalisation

En phase de Finalisation, l'organisation du projet est dissoute et les résultats du projet sont remis aux institutions assurant la maintenance des systèmes et des services. Les manuels sont distribués aux collaborateurs chargés des contrôles et de l'information aux chercheurs. Un rapport final décrivant les expériences faites au cours du projet est rendu au comité de pilotage en avril 2008.

La flexibilité de l'architecture FEDORA/FEZ rend SERVAL ouvert à des extensions futures, telles la création de périodiques en lignes, les *Overlay Journals*⁶⁶, les métriques sur les documents en Open Access⁶⁷ les technologies de réutilisation

⁶⁶ Voir l'article de Herbert van de Sompel, Carl Lagoze, Jeroen Bekaert, Xiaoming Liu, Sandy Payette, Simeon Warner : *An Interoperable Fabric for Scholarly Value Chain*, in **D-Lib Magazine** vol. 12 (10), octobre 2006, en ligne sous : <http://www.dlib.org/dlib/october06/vandesompel/10vandesompel.html> [vu le 25 avril 2007], et le site du projet d'*Overlay Journals* RIOJA , sous http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/rioja.aspx [vu le 25 avril 2007].

⁶⁷ La littérature sur les nouvelles métriques en rapport avec l'introduction des serveurs institutionnels et thématiques : Marko A. Rodriguez, Johan Bollen, Herbert Van de Sompel, *A Practical Ontology for the Large-Scale Modeling of Scholarly Artifacts and their Usage*, in **Proceedings of the Joint Conference on Digital Libraries**, Vancouver, Juin 2007 ; Johan Bollen et Herbert Van de Sompel, *Usage Impact Factor : the effects of sample characteristics on usage-based impact metrics* (cs.DL/0610154) ; Johan Bollen et Herbert Van de Sompel. *An Architecture for the aggregation and analysis of scholarly usage data*, in : **Joint Conference on Digital Libraries (JCDL2006)**, p. 298-307, juin 2006. Johan Bollen, Herbert Van de Sompel, *Mapping the structure of*

des contenus comme ORE⁶⁸ ou d'autres application du Web sémantique. Ces extensions pourront être décrites dans le rapport de finalisation si l'on envisage une continuation du projet.

science through usage, in : **Scientometrics**, 69 (2), 2006. Johan Bollen, Mark A. Rodriguez, Herbert Van de Sompel, *Journal Status*, in : **Scientometrics**, 69 (3), décembre 2006 (arxiv.org :cs.DL/0601030). Johan Bollen, Herbert Van de Sompel, Joan Smith, Rick Luce, *Toward alternative metrics of journal impact : a comparison of download and citation data*, in : **Information Proceeding and Management**, 41 (6) : 1419-1440, 2005.

⁶⁸ Site du projet ORE : <http://www.openarchives.org/ore/> [vu le 1^{er} mai 2007].

19 Conclusion

SERVAL, tel qu'il est présenté dans ce Concept pose une base à la fois solide et flexible pour la soumission, la validation, le stockage, la recherche et la mise à disposition des documents scientifiques UNIL/CHUV, ainsi que pour les documents numériques des e-collections et les données issues de la rétronumérisation BCU et BiUM.

Bien que limité à certains types de documents durant l'élaboration du projet, SERVAL pourra, en phase d'exploitation, recevoir des collections plus larges et également les collections d'autres partenaires UNIL intéressés.

Au vu de l'évolution favorable du contexte et du bon déroulement des travaux effectués jusqu'à l'heure actuelle, nous proposons au Comité de pilotage d'accepter le présent rapport de Concept et de donner le feu vert pour la phase Réalisation/implémentation du projet SERVAL.

Nous proposons à la Direction UNIL d'agender avec les Facultés, dans le courant 2007, des présentations du projet SERVAL, ainsi que des discussions concernant la modification des règlements de dépôt de thèses en faveur d'un dépôt automatique de la version électronique avec la version imprimée. La discussion pourrait également porter sur une obligation de dépôt de la version préprint des publications scientifiques UNIL sur SERVAL.

La BCU fera le nécessaire pour dégager une place d'administrateur SERVAL afin de garantir le bon fonctionnement des nouveaux workflows mis en place et une saisie de métadonnées et de documents en bonne quantité et qualité.

Le choix de la classification doit être confirmé ou infirmé par le comité de pilotage, car la CUS est engagée depuis peu dans des projets qui utilisent d'autres classifications. Une harmonisation est envisageable.

20 Annexes

20.1 Formulaire de l'enquête SERVAL

Enquête Serval	
SECTIONS	
1. Profil de la personne 2. Pratiques de recherches bibliographiques 2.1. Pratiques de recherches bibliographiques électroniques 2.2. Pratiques de recherches d'informations OA 3. Pratiques de publication 3.1. Pratiques publications imprimées 3.2. Pratiques de dépôt / publications OA 4. "Dépôt institutionnel" de l'UNIL	
1. Profil de la personne	
1.1 Quel est votre affiliation ?	
Institut/Département/Laboratoire	
Faculté	
1.2 Quel est votre statut ? (un seul statut, en cas de cumul, noter le plus élevé)	
• Etudiant	
• Assistant	
• Doctorant	
• Chercheur	
• Enseignant : maître assistant, professeur, etc.	
• Directeur d'institut/département/laboratoire	
• Autre (à préciser)	
1.3 Age	
<30 ans	
>31 <45	
> 45	
Enquete_serval.doc	
1	
V0.5	

1.4 Quels sont vos domaines d'activités et de recherche ?	
• Littérature, Linguistique, Langues étrangères	
• Philosophie	
• Mathématiques	
• Sciences Physiques	
• Chimie	
• Géologie	
• Géographie	
• Environnement, études Urbaines	
• Biologie	
• Informatique	
• Sciences et Technologies de l'Ingénieur	
• Sciences Vétérinaires	
• Médecine / Santé	
• Education	
• Economie	
• Commerce, Management, Tourisme	
• Sciences Politiques, Management public	
• Sociologie, anthropologie	
• Psychologie, Sciences comportementales et cognitives	
• Droit	
• Journalisme, Information, Communication, Bibliothéconomie	
• Arts, Muséologie, photographie, cinéma	
• Sport	
• Histoire / Archéologie	
• Théologie, sciences des religions	

2. Pratiques de recherche d'informations

2.1. Pratiques de recherche d'informations bibliographiques électroniques

2.1.1 Pour vos recherches bibliographiques sur Internet quels outils utilisez-vous de préférence ? (plusieurs réponses possibles)	
• Catalogues : RERO, etc.	
• Perunil	
• Collections de revues : Sciencedirect, JSTOR	
• Bases de données bibliographiques : Ovid, CSA etc.	
• Web of Science	
• Moteurs de recherche généralistes : Google, Yahoo...	
• Moteurs de recherche spécialisés Google Scholar, Scirus	
• Dépôts institutionnels : Rerodoc, Oaister	
• Autre(si possible préciser)	

2.1.2 Quels critères de recherche utilisez-vous ? (une réponse par ligne, en précisant 1- rarement, 2- parfois, 3- souvent, 4- toujours)	
• Auteur	
• Mots du titre	
• Descripteurs (termes contrôlés)	
• Classification (structure)	
• Mots libres plein texte	
• Autre (références liées...)	

2.1.3 Quels types de document cherchez-vous en ligne ?	
• Articles	
• Chapitres de livres, livres	
• Matériel d'enseignement (cours, exercices)	
• Thèses	
• Mémoires	
• Rapports techniques, guidelines	
• Images et vidéos	
• Documents de conférence, congrès : actes, présentations powerpoint	
• Autres (préciser)	

2.1.4 Généralement trouvez vous ce que vous cherchez sur Internet ?	
• Tout le temps	
• La plupart du temps	
• Parfois	
• Rarement	

2.1.5 Lorsque Google vous ramène 1 million de références, comment faites-vous pour trouver des références pertinentes ?	
• Je considère les 10 premières comme les plus pertinentes	
• J'utilise les fonctions de recherche avancée	
• J'ajoute des termes de recherche pour préciser, affiner la recherche	

2.1.6 Utilisez-vous des systèmes d'alertes (mail, fils rss) ?	
• Oui	
• Non	

2.1.7 Si oui, lesquels ?	
• Mail	
• Fils rss	
• Autre	

2.1.8 Comment utilisez-vous les documents électroniques ? (plusieurs réponses possibles)	
• Consultation des documents électroniques sur l'écran	
• Sauvegarde des documents dans votre ordinateur	
• Impression d'une partie des documents	
• Impression de tout le document	

2.1.9 Cherchez vous à savoir si l'information retrouvée provient d'une ressource gratuite ou payante (sous licence)	
• Oui	
• Non	

2.2. Pratiques de recherche d'informations bibliographiques en accès libre

2.2.1 Connaissez-vous le mouvement "Open Acces" ou Libre accès à l'information scientifique ?	
• Non	
• J'en ai déjà entendu parler	
• Oui, je consulte des sites OA	
• Oui, je consulte et dépose des documents sur des sites OA	

2.2.2 Sur quel site consultez-vous ces informations en libre accès ? (plusieurs réponses possibles)	
• Site institutionnel (lesquels ?)	
• Site personnel	
• Périodiques "Open Access" (lesquels?)	
• Autre	

3. Pratiques de publication

3.1 Pratiques de publication

3.1.1 Combien de publications (articles, conférences, académiques) produisez-vous (en tant qu'auteur principal ou en collaboration) en moyenne par année ?	
• 0-1 par an	
• 2-4 par an	
• 5-7 par an	
• 8-10 par an	
• Plus de 10 par an	

3.1.2 Comment décidez-vous du mode de publication ?	
• Je décide moi-même du canal de publication de mes articles.	
• Le canal de publication est imposé par mon directeur/institut.	
• Autre mode de fonctionnement.	

3.1.3 Quelle importance accordez-vous au facteur d'impact lorsque vous projetez de publier ?	
• Je ne considère que les revues pertinentes dans mon domaine, sans regarder le facteur d'impact	
• Je regarde le facteur d'impact relatif, en fonction du domaine de la revue et non pas sa valeur absolue	
• Je regarde la valeur absolue de l'Impact Factor de la revue	

3.1.4 En ce qui concerne le copyright, quelle est généralement votre position ?	
• Je ne m'occupe pas de la procédure de la soumission d'article	
• Je me contente de regarder ce que l'éditeur prévoit en matière de copyright.	
• Les éditeurs avec lesquels je travaille ne demandent pas la cession du copyright	
• J'accepte contre mon gré le transfert de copyright	
• Je signe sans objection l'accord de transfert de copyright	
• Je ne suis pas au courant des conditions contractuelles proposées par les éditeurs.	

3.1.5 Au cas où vous distribuez vos prépublications, comment le faites-vous ? (plusieurs réponses possibles)	
• Je les publie sur un site web personnel	
• Je les publie sur un site web institutionnel	
• Je les publie sur un site de pré publications propre à ma discipline	
• Je les envoie par email en fichier attaché	
• J'envoie des copies papier	
• Autre	

3.1.6 Dans quels formats sauvegardez-vous votre production scientifique ? (plusieurs réponses possibles)	
• Texte (format word ou autre TT)	
• PDF	
• Postscript	
• Excel	
• Powerpoint	
• Images	
• Vidéos	
• Sons	
• Data	
• Code source	
• Binaire	
• Autre	

3.1.7 Où et comment sauvegardez-vous votre production scientifique ? (plusieurs réponses possibles)	
• Stockage sur le disque dur de l'ordinateur	
• Stockage centralisé sur un serveur du service ou de l'institution	
• Sauvegarde sur supports externes	
• Autre	

3.1.8 Sauvegardez vous toutes les versions de vos publications ?	
• Oui, toutes les versions	
• Je ne garde que la version finale remise à l'éditeur	
• Je ne garde que la version finale (pdf) mise en page par l'éditeur	
• Je ne garde aucune version électronique de mes publications	

3.1.9 Savez-vous que la majorité des éditeurs vous permet de déposer votre version finale simultanément dans un dépôt institutionnel (serveur de l'institution) ?	
• oui	
• non	

3.2.0 Avez-vous déjà publié un article dans une revue électronique OA?	
• oui, plusieurs fois	
• oui, une fois	
• non	

3.2 Pratiques de dépôt OA / publication en OA

pour ceux qui ont déclaré avoir déposé des documents sur des sites OA à la question 2.2.1 ou 3.1.5

3.2.1 Depuis combien de temps pratiquez-vous le dépôt/publication dans des sites OA ?	
• L'année dernière	
• 2-3 ans	
• 3-5 ans	
• Plus de 5 ans	

3.2.2 Quel(s) genre(s) de publications déposez-vous dans des sites OA ? (plusieurs réponses possibles)	
• Articles destinés à des revues avec comité de lecture (peer reviewed)	
• Articles destinés à des revues sans comité de lecture	
• Chapitres de livres, livres	
• Matériel d'enseignement (cours, exercices)	
• Thèses	
• Mémoires	
• Rapports techniques, guidelines	
• Images et vidéos	
• Documents de conférence, congrès : actes, présentations powerpoint	
• Autre (préciser)	

3.2.3 Par qui est effectué le dépôt de vos documents dans les sites OA ?	
• Par vous-même	
• Par un co-auteur	
• Par le secrétariat de l'institut	
• Par la bibliothèque de l'institut	
• Autre	

3.2.4 Quel est le temps nécessaire moyen pour effectuer un dépôt pour le premier article ?	
• Moins de 15 minutes	
• Entre 15 et 30 minutes	
• Entre 30 et 60 minutes	
• Plus d'une heure	

3.2.5 Quel est le temps nécessaire moyen pour effectuer un dépôt pour les articles suivants ?	
• Moins de 15 minutes	
• Entre 15 et 30 minutes	
• Entre 30 et 60 minutes	
• Plus d'une heure	

3.2.6 Savez-vous qui a le droit de copyright de votre dernier article publié sur un site OA ?	
• Vous-même	
• Votre éditeur	
• Votre institution	
• Vous ne savez pas	

3.2.7 Pour quelles raisons publiez-vous dans des sites OA ? (plusieurs réponses possibles)	
• Par principe	
• Existence de sites OA dans votre domaine de recherche	
• Incitations de l'administration	
• Incitations des bibliothèques	
• Encouragements des collègues et co-auteurs	
• Les articles en accès en libre sont plus cités que les articles publiés dans les journaux payants	
• Autre (à préciser)	

3.2.8 Avec quelles affirmations concernant le dépôt de publications en libre accès êtes-vous d'accord ? (une réponse par ligne en précisant : 1- Pas du tout d'accord, 2- Plutôt pas d'accord, 3- Plutôt d'accord, 4- Tout à fait d'accord)	
• L'accessibilité de mon travail augmente	
• Mon travail est archivé et consultable en permanence	
• Mes documents publiés sont faciles à trouver	
• Mon travail est publié à côté d'autres travaux de recherche de qualité	
• Le nombre de citations de mon travail augmente	
• L'impact de ma recherche augmente	
• Je conserve le copyright sur mon travail	

3.2.9 Quels désavantages ou difficultés avez-vous rencontrés en déposant votre travail dans des sites OA ? (une réponse par ligne en précisant : 1- Pas du tout d'accord, 2- Plutôt pas d'accord, 3- Plutôt d'accord, 4- Tout à fait d'accord)	
• Déposer mon travail dans des archives ouvertes est difficile et/ou prend du temps	
• Sans validation par les pairs ou sans système de contrôle, la qualité du contenu peut être mise en doute	
• L'évaluation de la Recherche ne prend pas en compte le travail déposé dans les archives ouvertes	
• Mon travail pourrait être plagié	
• Autre	

4. "Dépôt institutionnel" de l'UNIL

4.1 L'Université de Lausanne est en train de créer un serveur pour offrir aux chercheurs (UNIL + CHUV) la possibilité déposer leurs publications sur un site central en OA. C'est le projet SERVAL. Quels sont, selon vous, les principaux avantages d'un tel dépôt institutionnel ? (plusieurs réponses possibles, mais pas plus de 3-4)	
• Rendre les publications accessibles gratuitement sans barrière commerciale	
• Rendre les publications accessibles plus rapidement	
• Rendre les publications accessibles sur Internet alors qu'elles ne le seraient pas autrement	
• Rendre plus visible l'éventail des recherches menées au sein de l'institution (CHUV + UNIL)	
• Centraliser les publications académiques / scientifiques de l'institution	
• Disposer d'un archivage pérenne	
• Permettre un meilleur suivi des publications personnelles	
• Permettre une mesure de la consultation (statistiques d'utilisation)	
• Rendre accessibles des documents pas mis à disposition via le processus normal de publication (ex. PPT, images, vidéos).	
• Rendre les publications accessibles sans effort d'entretien d'un site personnel	
• Autres (à préciser)	

4.2 A quelles conditions accepteriez-vous de déposer votre production scientifique (publiée ou non publiée) sur ce site institutionnel et de la rendre ainsi accessible sur Internet ? (plusieurs réponses possibles, mais pas plus de 3)	
• Garder la possibilité de publier en parallèle dans une revue	
• Avoir la garantie que le document est archivé sur une longue durée	
• Bénéficier d'une indexation et apparaître dans les moteurs de recherche	
• Bénéficier d'un processus de validation/ contrôle de qualité du contenu des documents déposés	
• Bénéficier d'un processus de validation/ contrôle de qualité de la description des documents déposés (métadonnées)	
• Bénéficier d'une protection contre le plagiat	
• Pouvoir fixer des limites d'accès : période d'embargo, restriction en fonction du périmètre : UNIL, hors UNIL	
• Autre (à préciser)	
• Pas de condition	

4.3 Quels types de documents aimeriez-vous déposer sur ce site institutionnel ?	
• Tout type de document	
• Articles destinés à des revues avec comité de lecture (peer reviewed)	
• Articles destinés à des revues sans comité de lecture	
• Chapitres de livres, livres	
• Matériel d'enseignement (cours, exercices)	
• Thèses	
• Mémoires	
• Rapports techniques, guidelines	
• Images et vidéos	
• Documents de conférence, congrès : actes, présentations powerpoint	
• Résultats/données d'expériences	
• Autre (préciser)	

4.4 Avez-vous déjà (vous-même ou votre institut/unité de recherche) envisagé de créer une revue en ligne, en Open Access?	
• oui	
• non	

Commentaires	
Le projet SERVAL vous intéresse-t-il ?	

20.2 *Détail de l'Australian Standard Research Classification dans FEZ :*

Premier niveau	Deuxième niveau	Troisième niveau
210000 Science - General		
220000 Social Sciences\, Humanities and Arts - General		
230000 Mathematical Sciences	230100 Mathematics	230101 Mathematical Logic\, Set Theory\, Lattices And Combinatorics
		230102 Number Theory And Field Theory
		230103 Rings And Algebras
		230104 Category Theory\, K Theory\, Homological Algebra
		230105 Group Theory And Generalisations (Incl. Topological Groups And Lie Groups)
		230106 Real and Complex Functions
		230107 Differential\, Difference and Integral Equations
		230108 Harmonic and Fourier Analysis
		230109 Functional Analysis
		230110 Calculus of Variations and Control Theory
		230111 Geometry
		230112 Topology and Manifolds
		230113 Dynamical Systems
		230114 Functions of Several Complex Variables
		230115 Approximation Theory
		230116 Numerical Analysis
		230117 Operations Research
		230118 Optimisation
		230119 Systems Theory and Control
		230199 Mathematics not elsewhere classified
	230200 Statistics	230201 Probability Theory
		230202 Stochastic Analysis and Modelling
		230203 Statistical Theory
		230204 Applied Statistics
		230299 Statistics not elsewhere classified
	239900 Other Mathematical Sciences	239901 Biological Mathematics
		239902 Actuarial Mathematics
		239903 Risk Theory
		239904 Theory of Graduation
		239999 Mathematical Sciences not elsewhere classified
240000 Physical Sciences	240100 Astronomical Sciences	240101 Astronomy and Astrophysics
		240102 Cosmic Ray Physics
		240199 Astronomical Sciences not elsewhere classified
	240200 Theoretical and Condensed Matter Physics	240201 Theoretical Physics
		240202 Condensed Matter Physics - Structural Properties

		240203 Condensed Matter Physics - Electronic and Magnetic Properties\, Superconductivity
		240204 Condensed Matter Physics - Other
	240300 Atomic and Molecular Physics\, Nuclear and Particle Physics\, Plasma Physics	240301 Atomic and Molecular Physics
		240302 Nuclear and Particle Physics
		240303 Plasmas and Electrical Discharges
		240304 Other Plasma Physics
	240400 Optical Physics	240401 Optics and Opto-electronic Physics
		240402 Quantum Optics and Lasers
		240499 Optical Physics not elsewhere classified
	240500 Classical Physics	240501 Acoustics and Acoustical Devices\, Waves
		240502 Fluid Physics
		240503 Thermodynamics and Statistical Physics
		240504 Electrostatics and Electrodynamics
		240599 Classical Physics not elsewhere classified
	249900 Other Physical Sciences	249901 Biophysics
		249902 Medical Physics
		249903 Instruments and Techniques
		249999 Physical Sciences not elsewhere classified
250000 Chemical Sciences	250100 Physical Chemistry (incl. Structural)	250101 Chemical Thermodynamics and Energetics
		250102 Chemistry of Catalysis
		250103 Colloid and Surface Chemistry
		250104 Chemical Spectroscopy
		250105 Structural Chemistry
		250106 Mechanisms of Reactions
		250107 Electrochemistry
		250199 Physical Chemistry not elsewhere classified
	250200 Inorganic Chemistry	250201 Transition Metal Chemistry
		250202 Main Group Metal Chemistry
		250203 Solid State Chemistry
		250204 Bioinorganic Chemistry
		250205 Non-Metal Chemistry
		250206 f-Block Chemistry
		250299 Inorganic Chemistry not elsewhere classified
	250300 Organic Chemistry	250301 Organic Chemical Synthesis
		250302 Biological and Medical Chemistry
		250303 Physical Organic Chemistry
		250399 Organic Chemistry not elsewhere classified
	250400 Analytical Chemistry	250401 Separation Science
		250402 Analytical Spectrometry
		250403 Electroanalytical Chemistry
		250404 Flow Analysis
		250405 Sensor (Chemical and Bio-) Technology
		250406 Immunological and Bioassay Methods
		250407 Other Instrumental Methods
		250409 Quality Assurance\, Traceability and Metrological Chemistry
		250499 Analytical Chemistry not elsewhere classified
	250500 Macromolecular Chemistry	250501 Synthesis of Macromolecules
		250502 Physical Chemistry of Macromolecules
		250503 Characterisation of Macromolecules

		250504 Polymerisation Mechanisms
		250599 Macromolecular Chemistry not elsewhere classified
	250600 Theoretical and Computational Chemistry	250601 Quantum Chemistry
		250602 Statistical Mechanics
		250603 Reaction Kinetics and Dynamics
		250604 Radiation and Matter
		250605 Theory of Materials
		250606 Macromolecular Design
		250699 Theoretical and Computational Chemistry not elsewhere classified
	259900 Other Chemical Sciences	259901 Organometallic Chemistry
		259902 Environmental Chemistry (incl. Atmospheric Chemistry)
		259903 Industrial Chemistry
		259904 Supramolecular Chemistry
		259999 Chemical Sciences not elsewhere classified
260000 Earth Sciences	260100 Geology	260101 Mineralogy and Crystallography
		260102 Igneous and Metamorphic Petrology
		260103 Vulcanology
		260104 Sedimentology
		260105 Petroleum Geology
		260106 Ore Deposit Petrology
		260107 Structural Geology
		260108 Geotectonics
		260109 Geochronology
		260110 Biostratigraphy
		260111 Other Stratigraphy (incl. Sequence Stratigraphy)
		260112 Palaeontology
		260113 Palynology
		260114 Geomorphology
		260115 Glaciology
		260199 Geology not elsewhere classified
	260200 Geophysics	260201 Gravity
		260202 Geomagnetism
		260203 Electromagnetism
		260204 Petrophysics
		260205 Explosion Seismology
		260206 Earthquake Seismology
		260207 Radiometrics
		260299 Geophysics not elsewhere classified
	260300 Geochemistry	260301 Geochronology and Isotope Geochemistry
		260302 Exploration Geochemistry
		260303 Organic Geochemistry not elsewhere classified
		260304 Inorganic Geochemistry not elsewhere classified
		260399 Geochemistry not elsewhere classified
	260400 Oceanography	260401 Biological Oceanography
		260402 Chemical Oceanography
		260403 Physical Oceanography
		260499 Oceanography not elsewhere classified
	260500 Hydrology	260501 Groundwater Hydrology
		260502 Surfacewater Hydrology
		260599 Hydrology not elsewhere classified
	260600 Atmospheric Sciences	260601 Meteorology
		260602 Climatology (incl. Palaeoclimatology)

		260603 Ionospheric and Magnetospheric Physics
		260699 Atmospheric Sciences not elsewhere classified
	269900 Other Earth Sciences	269901 Physical Geography
		269999 Earth Sciences not elsewhere classified
270000 Biological Sciences	270100 Biochemistry and Cell Biology	270101 Analytical Biochemistry
		270102 Cell Metabolism
		270103 Protein Targeting and Signal Transduction
		270104 Membrane Biology
		270105 Cellular Interactions (incl. Adhesion\, Matrix\, Cell Wall)
		270106 Cell Development (incl. Cell Division and Apoptosis)
		270107 Cell Neurochemistry
		270108 Enzymes
		270199 Biochemistry and Cell Biology not elsewhere classified
	270200 Genetics	270201 Gene Expression
		270202 Genome Structure
		270203 Population and Ecological Genetics
		270204 Anthropological Genetics
		270205 Genetic Development (incl. Sex Determination)
		270206 Genetic Immunology
		270207 Quantitative Genetics
		270208 Molecular Evolution
		270209 Meiosis and Recombination
		270210 Neurogenetics
		270299 Genetics not elsewhere classified
	270300 Microbiology	270301 Bacteriology
		270302 Archaeobacteriology
		270303 Virology
		270304 Infectious Agents
		270305 Mycology
		270306 Microbial Genetics
		270307 Microbial Ecology
		270308 Microbial Systematics\, Taxonomy and Phylogeny
		270399 Microbiology not elsewhere classified
	270400 Botany	270401 Plant Systematics\, Taxonomy and Phylogeny
		270402 Plant Physiology
		270403 Plant Pathology
		270404 Phycology
		270499 Botany not elsewhere classified
	270500 Zoology	270501 Animal Systematics\, Taxonomy and Phylogeny
		270502 Neurobiology
		270503 Animal Anatomy and Histology
		270504 Invertebrate Biology
		270599 Zoology not elsewhere classified
	270600 Physiology	270601 Animal Physiology - Biophysics
		270602 Animal Physiology - Cell
		270603 Animal Physiology - Systems
		270604 Comparative Physiology
		270699 Physiology not elsewhere classified
	270700 Ecology and Evolution	270701 Freshwater Ecology
		270702 Marine and Estuarine Ecology (incl. Marine Ichthyology)
		270703 Terrestrial Ecology
		270704 Landscape Ecology
		270705 Palaeoecology

		270706 Life Histories (incl. Population Ecology)
		270707 Sociobiology and Behavioural Ecology
		270708 Conservation and Biodiversity
		270709 Biogeography
		270799 Ecology and Evolution not elsewhere classified
	270800 Biotechnology	270801 Gene Therapy
		270802 Diagnostic Applications
		270803 Transgenesis
		270804 Genetic Technologies - Transformation\, Site-directed Mutagenesis\, etc.
		270805 Genetic Engineering and Enzyme Technology
		270899 Biotechnology not elsewhere classified
	279900 Other Biological Sciences	279901 Global Change Biology
		279999 Biological Sciences not elsewhere classified
280000 Information\, Computing and Communication Sciences	280100 Information Systems	280101 Information Systems Organisation
		280102 Information Systems Management
		280103 Information Storage\, Retrieval and Management
		280104 Computer-Human Interaction
		280105 Interfaces and Presentation (excl. Computer-Human Interaction)
		280106 Interorganisational Information Systems
		280107 Global Information Systems
		280108 Database Management
		280109 Decision Support and Group Support Systems
		280110 Systems Theory
		280111 Conceptual Modelling
		280112 Information Systems Development Methodologies
		280199 Information Systems not elsewhere classified
	280200 Artificial Intelligence and Signal and Image Processing	280201 Expert Systems
		280202 Computer Graphics
		280203 Image Processing
		280204 Signal Processing
		280205 Text Processing
		280206 Speech Recognition
		280207 Pattern Recognition
		280208 Computer Vision
		280209 Intelligent Robotics
		280210 Simulation and Modelling
		280211 Virtual Reality and Related Simulation
		280212 Neural Networks\, Genetic Algorithms and Fuzzy Logic
		280213 Other Artificial Intelligence
	280300 Computer Software	280301 Programming Techniques
		280302 Software Engineering
		280303 Programming Languages
		280304 Operating Systems
		280305 Multimedia Programming
		280399 Computer Software not elsewhere classified
	280400 Computation Theory and Mathematics	280401 Analysis of Algorithms and Complexity
		280402 Mathematical Logic and Formal Languages
		280403 Logics and Meanings of Programs

		280404 Numerical Analysis
		280405 Discrete Mathematics
		280406 Mathematical Software
		280499 Computation Theory and Mathematics not elsewhere classified
	280500 Data Format	280501 Data Structures
		280502 Data Storage Representations
		280503 Files
		280504 Data Encryption
		280505 Data Security
		280506 Coding and Information Theory
		280599 Data Format not elsewhere classified
	289900 Other Information\, Computing and Communication Sciences	289999 Other Information\, Computing and Communication Sciences
290000 Engineering and Technology	290100 Industrial Biotechnology and Food Sciences	290101 Fermentation\, Biotechnology and Industrial Microbiology
		290102 Food Engineering
		290103 Food Processing
		290104 Other Food Sciences
		290105 Other Industrial Biotechnology
	290200 Aerospace Engineering	290201 Aerodynamics
		290202 Flight Dynamics
		290203 Aerospace Structures
		290204 Aircraft Performance
		290205 Flight Control Systems
		290206 Aerospace Electrical Systems
		290207 Satellite\, Space Vehicle and Missile Design
		290299 Aerospace Engineering not elsewhere classified
	290300 Manufacturing Engineering	290301 Robotics and Mechatronics
		290302 Flexible Manufacturing Systems
		290303 CAD/CAM Systems
		290304 Control Engineering
		290305 Welding Technology
		290306 Textile Technology
		290307 Printing Technology
		290308 Packaging\, Storage and Transportation
		290309 Safety and Quality
		290399 Manufacturing Engineering not elsewhere classified
	290400 Automotive Engineering	290401 Automotive Engineering
	290500 Mechanical and Industrial Engineering	290501 Mechanical Engineering
		290502 Industrial Engineering
	290600 Chemical Engineering	290601 Chemical Engineering Design
		290602 Process Control and Simulation
		290603 Membrane and Separation Technologies
		290699 Chemical Engineering not elsewhere classified
	290700 Resources Engineering	290701 Mining Engineering
		290702 Mineral Processing
		290703 Petroleum and Reservoir Engineering
		290704 Geomechanics

	290799 Resources Engineering not elsewhere classified
290800 Civil Engineering	290801 Structural Engineering
	290802 Water and Sanitary Engineering
	290803 Transport Engineering
	290804 Construction Engineering
	290805 Geotechnical Engineering
	290899 Civil Engineering not elsewhere classified
290900 Electrical and Electronic Engineering	290901 Electrical Engineering
	290902 Integrated Circuits
	290903 Other Electronic Engineering
291000 Geomatic Engineering	291001 Geodesy
	291002 Surveying
	291003 Photogrammetry and Remote Sensing
	291004 Spatial Information Systems
	291005 Navigation and Position Fixing
	291006 Cartography
	291099 Geomatic Engineering not elsewhere classified
291100 Environmental Engineering	291101 Environmental Engineering Modelling
	291102 Bio-remediation
	291103 Environmental Engineering Design
	291104 Environmental Technologies
	291199 Environmental Engineering not elsewhere classified
291200 Maritime Engineering	291201 Naval Architecture
	291202 Ship and Platform Hydrodynamics
	291203 Ship and Platform Structures
	291204 Marine Engineering
	291205 Ocean Engineering
	291206 Special Vehicles
	291299 Maritime Engineering not elsewhere classified
	291300 Metallurgy
	291302 Physical Metallurgy
291400 Materials Engineering	291401 Polymers
	291402 Composite Materials
	291403 Alloy Materials
	291404 Ceramics
	291405 Timber
	291406 Pulp and Paper
	291407 Plastics
	291499 Materials Engineering not elsewhere classified
291500 Biomedical Engineering	291501 Clinical Engineering
	291502 Rehabilitation Engineering
	291503 Biomaterials
	291504 Biomechanical Engineering
	291599 Biomedical Engineering not elsewhere classified
291600 Computer Hardware	291601 Arithmetic and Logic Structures
	291602 Memory Structures
	291603 Input\, Output and Data Devices
	291604 Logic Design
	291605 Processor Architectures
	291699 Computer Hardware not elsewhere classified

	291700 Communications Technologies	291701 Antenna Technology
		291702 Optical and Photonic Systems
		291703 Digital Systems
		291704 Computer Communications Networks
		291705 Microwave and Millimetrewave Technology
		291706 Broadband Network Technology
		291707 Modem Technology
		291708 Baseband Technology
		291709 Satellite Communications
		291710 Radio Communications and Broadcasting not elsewhere classified
		291799 Communications Technologies not elsewhere classified
	291800 Interdisciplinary Engineering	291801 Fluidization and Fluid Mechanics
		291802 Heat and Mass Transfer Operations
		291803 Turbulent Flows
		291804 Nanotechnology
		291899 Interdisciplinary Engineering not elsewhere classified
	299900 Other Engineering and Technology	299901 Agricultural Engineering
		299902 Combustion and Fuel Engineering
		299903 Biosensor Technologies
		299904 Engineering/Technology Instrumentation
		299999 Engineering and Technology not elsewhere classified
300000 Agricultural\, Veterinary and Environmental Sciences	300100 Soil and Water Sciences	300101 Soil Physics
		300102 Soil Biology
		300103 Soil Chemistry
		300104 Land Capability and Soil Degradation
		300105 Applied Hydrology (Drainage\, Flooding\, Irrigation\, Quality\, etc.)
		300199 Soil and Water Sciences not elsewhere classified
	300200 Crop and Pasture Production	300201 Plant Biochemistry and Physiology
		300202 Plant Nutrition
		300203 Plant Improvement (Selection\, Breeding and Genetic Engineering)
		300204 Plant Protection (Pests\, Diseases and Weeds)
		300205 Agronomy
		300206 Post Harvest Technologies (Transportation and Storage)
		300299 Crop and Pasture Production not elsewhere classified
	300300 Horticulture	300301 Plant Improvement (Selection\, Breeding and Genetic Engineering)
		300302 Plant Growth and Development
		300303 Plant Protection (Pests\, Diseases and Weeds)
		300304 Post Harvest Technologies
		300305 Oenology and Viticulture
		300399 Horticulture not elsewhere classified
	300400 Animal Production	300401 Animal Breeding
		300402 Animal Reproduction
		300403 Animal Nutrition
		300404 Animal Husbandry
		300405 Animal Protection (Pests and Pathogens)
		300406 Animal Growth and Development

		300499 Animal Production not elsewhere classified
	300500 Veterinary Sciences	300501 Veterinary Medicine
		300502 Anaesthesiology and Intensive Care
		300503 Epidemiology
		300504 Immunology
		300505 Anatomy and Physiology
		300506 Pathology
		300507 Microbiology (excl. Virology)
		300508 Parasitology
		300509 Radiology and Imaging
		300510 Virology
		300511 Surgery
		300512 Pharmacology
		300599 Veterinary Sciences not elsewhere classified
	300600 Forestry Sciences	300601 Nutrition and Physiology
		300602 Tree Improvement (Selection\, Breeding and Genetic Engineering)
		300603 Pests\, Health and Diseases
		300604 Management and Environment
		300605 Fire Management
		300606 Agroforestry
		300699 Forestry Sciences not elsewhere classified
	300700 Fisheries Sciences	300701 Physiology and Genetics
		300702 Pests and Diseases
		300703 Aquaculture
		300704 Ecosystem Studies and Stock Assessment
		300705 Evaluation of Management Strategies
		300799 Fisheries Sciences not elsewhere classified
	300800 Environmental Sciences	300801 Environmental Management and Rehabilitation
		300802 Wildlife and Habitat Management
		300803 Natural Resource Management
		300804 Environmental Impact Assessment
		300805 Conservation
		300899 Environmental Sciences not elsewhere classified
	300900 Land\, Parks and Agriculture Management	300901 Farm Management\, Rural Management and Agribusiness
		300902 Land and Parks Management
		300903 Sustainable Development
		300999 Land\, Parks and Agriculture Management not elsewhere classified
	309900 Other Agricultural\, Veterinary and Environmental Sciences	309901 Fertilisers and Agrochemicals (Application etc.)
		309902 Education and Extension
		309999 Agricultural\, Veterinary and Environmental Sciences not elsewhere classified
310000 Architecture\, Urban Environment and Building	310100 Architecture and Urban Environment	310101 Architecture
		310102 Heritage and Conservation
		310103 Urban and Regional Planning
		310104 Landscape Planning
		310105 History of the Built Environment
		310106 Interior and Environmental Design

		310199 Architecture and Urban Environment not elsewhere classified
	310200 Building	310201 Building Science and Techniques
		310202 Building Construction Management
		310203 Building Construction Economics (incl. Quantity Surveying)
		310299 Building not elsewhere classified
	319900 Other Architecture\, Urban Environment and Building	319999 Other Architecture\, Urban Environment and Building
320000 Medical and Health Sciences	320100 Medicine - General	
	320200 Immunology	320201 Allergy
		320202 Cellular Immunology
		320203 Humoral Immunology and Immunochemistry
		320204 Immunogenetics
		320205 Transplantation Immunology
		320206 Tumor Immunology
		320207 Autoimmunity
		320299 Immunology not elsewhere classified
	320300 Medical Biochemistry and Clinical Chemistry	320301 Clinical Chemistry
		320302 Medical Biochemistry - Carbohydrates
		320303 Medical Biochemistry - Lipids
		320304 Medical Biochemistry - Nucleic Acids
		320305 Medical Biochemistry - Proteins and Peptides
		320306 Medical Biochemistry - Phospholipids
		320307 Medical Biochemistry - Other
	320400 Medical Microbiology	320401 Medical Bacteriology
		320402 Medical Virology
		320403 Medical Infection Agents (incl. Prions)
		320404 Medical Mycology
		320405 Medical Parasitology
		320499 Medical Microbiology not elsewhere classified
	320500 Pharmacology and Pharmaceutical Sciences	320501 Pharmaceutical Sciences and Pharmacy
		320502 Basic Pharmacology
		320503 Clinical Pharmacology and Therapeutics
		320504 Toxicology (incl. Clinical Toxicology)
		320599 Pharmacology not elsewhere classified
	320600 Medical Physiology	320601 Human Biophysics
		320602 Cell Physiology
		320603 Systems Physiology
		320699 Medical Physiology not elsewhere classified
	320700 Neurosciences	320701 Autonomic Nervous System
		320702 Central Nervous System
		320703 Peripheral Nervous System
		320704 Cellular Nervous System
		320705 Sensory Systems
		320799 Neurosciences not elsewhere classified
	320800 Dentistry	320801 Oral Surgery
		320802 Dentistry Support
		320803 Dental Technology
		320804 Dental Therapy
		320899 Dentistry not elsewhere classified
	320900 Optometry	320901 Optical Technology

	320999 Optometry not elsewhere classified
321000 Clinical Sciences	321001 Anaesthesiology
	321002 Dermatology
	321003 Cardiology (incl. Cardiovascular Diseases)
	321004 Endocrinology
	321005 Fetal Development and Medicine
	321006 Gastroenterology and Hepatology
	321007 Geriatrics and Gerontology
	321008 Haematology
	321009 Intensive Care
	321010 Infectious Diseases
	321011 Medical Genetics
	321012 Nephrology and Urology
	321013 Neurology and Neuromuscular Diseases
	321014 Obstetrics and Gynaecology
	321015 Oncology and Carcinogenesis
	321016 Ophthalmology and Vision Science
	321017 Orthopaedics
	321018 Otorhinolaryngology
	321019 Paediatrics
	321020 Pathology
	321021 Psychiatry
	321022 Radiology and Organ Imaging
	321023 Radiotherapy and Nuclear Medicine
	321024 Rehabilitation and Therapy - Occupational and Physical
	321025 Rehabilitation and Therapy - Hearing and Speech
	321026 Reproduction
	321027 Respiratory Diseases
	321028 Rheumatology and Arthritis
	321029 Surgery
	321030 Venereology
	321099 Clinical Sciences not elsewhere classified
321100 Nursing	321101 Midwifery
	321102 Clinical Nursing - Primary (Preventative)
	321103 Clinical Nursing - Secondary (Acute Care)
	321104 Clinical Nursing - Tertiary (Rehabilitative)
	321105 Mental Health Nursing
	321106 Aged Care Nursing
	321199 Nursing not elsewhere classified
321200 Public Health and Health Services	321201 Environmental and Occupational Health and Safety
	321202 Epidemiology
	321203 Health Information Systems (incl. Surveillance)
	321204 Mental Health
	321205 Nutrition and Dietetics
	321206 Preventive Medicine
	321207 Indigenous Health
	321208 Primary Health Care
	321209 Family Care
	321210 Community Child Health
	321211 Residential Client Care
	321212 Care for Disabled
	321213 Human Bioethics
	321214 Health and Community Services
	321215 Health Care Administration
	321216 Health Promotion

		321217 Health Counselling
		321299 Public Health and Health Services not elsewhere classified
	321300 Complementary/alternative Medicine	321301 Oriental Medicine and Treatments
		321302 Chiropractic
		321303 Naturopathy
		321399 Complementary/Alternative Medicine not elsewhere classified
	321400 Human Movement and Sports Science	321401 Exercise Physiology
		321402 Biomechanics
		321403 Motor Control
		321404 Sport and Exercise Psychology
		321405 Sports Medicine
		321499 Human Movement and Sports Science not elsewhere classified
	329900 Other Medical and Health Sciences	329901 Podiatry
		329902 Medical Biotechnology
		329903 Therapies and Therapeutic Technology
		329999 Medical and Health Sciences not elsewhere classified
330000 Education	330100 Education Studies	330101 Educational Psychology
		330102 History and Philosophy of Education
		330103 Sociology of Education
		330104 Educational Policy\, Administration and Management
		330105 Educational Counselling
		330106 Comparative Education
		330107 Educational Technology and Media
		330108 Special Education
		330109 Assessment and Evaluation
		330199 Education Studies not elsewhere classified
	330200 Curriculum Studies	330201 Curriculum Studies - English Education
		330202 Curriculum Studies - Mathematics Education
		330203 Curriculum Studies - Science Education
		330204 Curriculum Studies - Economics\, Commerce\, Management and Services Education
		330205 Curriculum Studies - Other Social Sciences\, Humanities and Arts Education
		330206 Curriculum Theory and Development
		330299 Curriculum Studies not elsewhere classified
	330300 Professional Development of Teachers	330301 Teacher Education - Early Childhood
		330302 Teacher Education - Primary
		330303 Teacher Education - Secondary
		330304 Teacher Education - Vocational Education and Training
		330305 Teacher Education - Higher Education
		330306 Teacher Education - Special Education
		330307 Teacher Education - Nursing and Health
		330399 Professional Development of Teachers not elsewhere classified
	339900 Other Education	339999 Other Education
340000 Economics	340100 Economic Theory	340101 Microeconomic Theory
		340102 Macroeconomic Theory
		340103 Mathematical Economics
		340199 Economic Theory not elsewhere classified
	340200 Applied Economics	340201 Agricultural Economics

		340202 Environment and Resource Economics
		340203 Finance Economics
		340204 Health Economics
		340205 Industry Economics and Industrial Organisation
		340206 International Economics and International Finance
		340207 Labour Economics
		340208 Macroeconomics (incl. Monetary and Fiscal Theory)
		340209 Public Sector Economics
		340210 Welfare Economics
		340211 Transport Economics
		340212 Economic Geography
		340213 Economic Development and Growth
		340214 Urban and Regional Economics
		340299 Applied Economics not elsewhere classified
	340300 Economic History and History of Economic Thought	340301 Economic History
		340302 History of Economic Thought
	340400 Econometrics	340401 Economic Models and Forecasting
		340402 Econometric and Statistical Methods
		340403 Time-Series Analysis
		340404 Cross-Sectional Analysis
		340405 Panel Data Analysis
		340499 Econometrics not elsewhere classified
	349900 Other Economics	349901 Political Economy
		349902 Comparative Economic Systems
		349999 Economics not elsewhere classified
350000 Commerce\, Management\, Tourism and Services	350100 Accounting\, Auditing and Accountability	350101 Financial Accounting
		350102 Management Accounting
		350103 Auditing and Accountability
		350104 Taxation
		350105 Bookkeeping
		350106 International Accounting
		350107 Other Accounting
	350200 Business and Management	350201 Human Resources Management
		350202 Business Information Systems (incl. Data Processing)
		350203 Industrial Relations
		350204 Marketing and Market Research
		350205 Sales and Distribution
		350206 Advertising and Public Relations
		350207 Office Services
		350208 Organisational Planning and Management
		350209 Small Business Management
		350210 Quality Management
		350211 Innovation and Technology Management
		350212 International Business
		350213 Electronic Commerce
		350299 Business and Management not elsewhere classified
	350300 Banking\, Finance and Investment	350301 Finance
		350302 Financial Econometrics
		350303 Financial Institutions (incl. Banking)

		350304 Insurance Studies
		350399 Banking\, Finance and Investment not elsewhere classified
	350400 Transportation	350401 Aeronautical Transportation
		350402 Aircrew Training
		350403 Nautical Transportation
		350404 Marine Crew Training
		350405 Road and Rail Transportation
		350499 Transportation not elsewhere classified
	350500 Tourism	350501 Tourism Policy and Planning
		350502 Tourism Resource Appraisal
		350503 Impacts of Tourism
		350504 Tourist Behaviour
		350505 Tourism Economics
		350506 Tourism Forecasting
		350507 Tourism Management
		350508 Tourism Marketing
		350599 Tourism not elsewhere classified
	350600 Services	350601 Hospitality Management
		350602 Food and Hospitality Services
		350603 Real Estate and Valuation
		350604 Professional Services (Legal\, Management Consulting\, etc.)
		350605 Other Retail Services
		350606 Wholesale Services
		350699 Services not elsewhere classified
	359900 Other Commerce\, Management\, Tourism and Services	359999 Other Commerce\, Management\, Tourism and Services
360000 Policy and Political Science	360100 Political Science	360101 Australian Government and Politics
		360102 Comparative Government and Politics
		360103 Federalism and Intergovernmental Relations
		360104 Political Theory and Political Philosophy
		360105 International Relations
		360199 Political Science not elsewhere classified
	360200 Policy and Administration	360201 Public Policy
		360202 Public Administration
		360203 Research\, Science and Technology Policy
		360204 Defence Studies
		360299 Policy and Administration not elsewhere classified
	369900 Other Policy and Political Science	369999 Other Policy and Political Science
370000 Studies in Human Society	370100 Sociology	370101 Social Theory
		370102 Social Policy and Planning
		370103 Race and Ethnic Relations
		370104 Urban Sociology and Community Studies
		370105 Applied Sociology\, Program Evaluation and Social Impact Assessment
		370106 Sociological Methodology and Research Methods
		370107 Social Change
		370199 Sociology not elsewhere classified
	370200 Social Work	370201 Clinical Social Work Practice
		370202 Social Program Evaluation
		370203 Social Policy
		370204 Counselling\, Welfare and Community Services

		370299 Social Work not elsewhere classified
	370300 Anthropology	370301 Biological (Physical) Anthropology
		370302 Social and Cultural Anthropology
		370303 Linguistic Anthropology
		370399 Anthropology not elsewhere classified
	370400 Human Geography	370401 Urban and Regional Studies
		370402 Social and Cultural Geography
		370403 Recreation and Leisure Studies
		370499 Human Geography not elsewhere classified
	370500 Demography	370501 Population Trends and Policies
		370502 Migration
		370503 Fertility
		370504 Family and Household Studies
		370599 Demography not elsewhere classified
	370600 History and Philosophy of Science and Medicine	370601 History and Philosophy of Science and Technology
		370602 Sociology and Social Studies of Science and Technology
		370603 History and Philosophy of Medicine
		370699 History and Philosophy of Science and Medicine not elsewhere classified
	379900 Other Studies in Human Society	379901 Gender Specific Studies
		379902 Aboriginal Studies
		379999 Studies in Human Society not elsewhere classified
380000 Behavioural and Cognitive Sciences	380100 Psychology	380101 Sensory Processes\, Perception and Performance
		380102 Learning\, Memory\, Cognition and Language
		380103 Biological Psychology (Neuropsychology\, Psychopharmacology\, Physiological Psychology)
		380104 Personality\, Abilities and Assessment
		380105 Social and Community Psychology
		380106 Developmental Psychology and Ageing
		380107 Health\, Clinical and Counselling Psychology
		380108 Industrial and Organisational Psychology
		380109 Psychological Methodology\, Design and Analysis
		380199 Psychology not elsewhere classified
	380200 Linguistics	380201 Applied Linguistics and Educational Linguistics
		380202 Computational Linguistics
		380203 Discourse and Pragmatics
		380204 Laboratory Phonetics and Speech Science
		380205 Language in Culture and Society (Sociolinguistics)
		380206 Language in Time and Space (incl. Historical Linguistics\, Dialectology)
		380207 Linguistic Structures (incl. Grammar\, Phonology\, Lexicon\, Semantics)
		380208 Lexicography
		380299 Linguistics not elsewhere classified
	380300 Cognitive Science	380301 Philosophy of Cognition
		380302 Linguistic Processes (incl. Speech Production and Comprehension)
		380303 Computer Perception\, Memory and Attention
		380304 Neurocognitive Patterns and Neural Networks
		380305 Knowledge Representation and Machine Learning
		380306 Planning and Problem Solving
		380399 Cognitive Science not elsewhere classified

	389900 Other Behavioural and Cognitive Sciences	389999 Other Behavioural and Cognitive Sciences
390000 Law\, Justice and Law Enforcement	390100 Law	390101 Administrative Law
		390102 Comparative Law
		390103 Constitutionalism and Constitutional Law
		390104 Commercial and Contract Law
		390105 Corporations and Associations Law
		390106 Criminal Law
		390107 Environmental and Natural Resources Law
		390108 Family Law
		390109 Civil Law
		390110 Indigenous Law
		390111 International Law
		390112 Law and Literature
		390113 Legal History
		390114 Intellectual Property
		390115 Property Law and Conveyancing (excl. Intellectual Property)
		390116 Labour Law
		390117 Tort Law
		390118 Taxation Law
		390119 Equity Law
		390120 Trust Law
		390199 Law not elsewhere classified
	390200 Professional Development of Law Practitioners	390201 Legal Practice
		390202 Adjudication
		390203 Legal Education
		390204 Professional Ethics
		390299 Professional Development of Law Practitioners not elsewhere classified
	390300 Justice and Legal Studies	390301 Justice Systems and Administration
		390302 Jurisprudence and Legal Theory
		390303 Human Rights
		390304 Counselling and Mediation Services
		390305 Law and Society
		390399 Justice and Legal Studies not elsewhere classified
	390400 Law Enforcement	390401 Criminology
		390402 Evidence and Procedure
		390403 Police Administration\, Procedures and Practice
		390404 Detection and Prevention of Crime\, Security Services
		390405 Correctional Theory and Services\, Penology
		390499 Law Enforcement not elsewhere classified
	399900 Other Law\, Justice and Law Enforcement	399901 History and Philosophy of Law and Justice
		399999 Law\, Justice and Law Enforcement not elsewhere classified
400000 Journalism\, Librarianship and Curatorial Studies	400100 Journalism\, Communication and Media	400101 Journalism
		400102 Technical Writing
		400103 Professional Creative Writing
		400104 Communication and Media Studies

		400199 Journalism\, Communication and Media not elsewhere classified
	400200 Librarianship	400201 Librarianship
	400300 Curatorial Studies	400301 Museum Studies
		400302 Archival Studies
		400303 Materials Conservation
		400399 Curatorial Studies not elsewhere classified
	409900 Other Journalism\, Librarianship and Curatorial Studies	409999 Other Journalism\, Librarianship and Curatorial Studies
410000 The Arts	410100 Performing Arts	410101 Music
		410102 Drama\, Theatre and Performance Studies
		410103 Dance
		410104 Indigenous Performing Arts
		410199 Performing Arts not elsewhere classified
	410200 Visual Arts and Crafts	410201 Graphic Arts
		410202 Fine Arts (incl. Sculpture and Painting)
		410203 Photography
		410204 Crafts
		410299 Visual Arts and Crafts not elsewhere classified
	410300 Cinema\, Electronic Arts and Multimedia	410301 Film and Video
		410302 Cinema Studies
		410303 Multimedia
		410304 Other Cinema and Electronic Arts
	410400 Design Studies	410401 Design Theory
		410402 Design Innovation
		410403 Design Studio Practice
		410404 Design Management and Professional Practice
		410405 Design History and Cultural Theory
		410499 Design Studies not elsewhere classified
	419900 Other Arts	419901 Art History and Appreciation
		419999 The Arts not elsewhere classified
420000 Language and Culture	420100 Language Studies	420101 English
		420102 English as a Second Language
		420103 Aboriginal Languages
		420104 Latin
		420105 Greek
		420106 French
		420107 German
		420108 Italian
		420109 Spanish
		420110 Russian
		420111 Other European Languages
		420112 Chinese Languages
		420113 Japanese
		420114 Indonesian Languages
		420115 Malaysian Languages
		420116 Indian Languages
		420117 Korean
		420118 Vietnamese Languages
		420119 Thai Languages
		420120 Other Asian Languages

		420121 Comparative Language Studies
		420122 Oracy and Oral Traditions
		420199 Language Studies not elsewhere classified
	420200 Literature Studies	420201 British and Irish
		420202 Australian and New Zealand
		420203 North American
		420204 South American
		420205 Latin and Classical Greek
		420206 French
		420207 German
		420208 Italian
		420209 Russian
		420210 Other European
		420211 Chinese
		420212 Japanese
		420213 Indonesian
		420214 Indian
		420215 Other Asian
		420216 Comparative Literature Studies
		420217 Textual Transmission and the Material Record
		420218 Literary Theory
		420219 Stylistics
		420220 Folklore\, Myth and Mythologies
		420299 Literature Studies not elsewhere classified
	420300 Cultural Studies	420301 Cultural Policy Studies
		420302 Cultural Theory
		420303 Culture\, Gender\, Sexuality
		420304 Screen and Media Culture
		420305 Aboriginal Cultural Studies
		420306 Postcolonial and Global Cultural Studies
		420307 Consumption and Everyday Life
		420308 Multicultural\, Intercultural and Cross-cultural Studies
		420399 Cultural Studies not elsewhere classified
	429900 Other Language and Culture	429999 Other Language and Culture
430000 History and Archaeology	430100 Historical Studies	430101 History - Australian
		430102 History - Asian
		430103 History - Pacific
		430104 History - North American
		430105 History - Latin American
		430106 History - African
		430107 History - British
		430108 History - European
		430109 History - Middle Eastern
		430110 History - Classical Greek and Roman
		430111 History - Other
		430112 Biography
		430199 Historical Studies not elsewhere classified
	430200 Archaeology and Prehistory	430201 Archaeology of Hunter-Gatherer Societies (incl. Pleistocene Archaeology)
		430202 Archaeology of Agricultural and Pastoral Societies
		430203 Archaeology of Complex Societies - Europe\, the Mediterranean and the Levant
		430204 Archaeology of Complex Societies - Asia\, Africa\, Oceania and the Americas

		430205 Historical Archaeology (incl. Industrial Archaeology)
		430206 Maritime Archaeology
		430207 Archaeological Science
		430299 Archaeology and Prehistory not elsewhere classified
	439900 Other History and Archaeology	439999 Other History and Archaeology
440000 Philosophy and Religion	440100 Philosophy	440101 Aesthetics
		440102 Epistemology
		440103 Ethical Theory
		440104 Applied Ethics (incl. Bioethics and Environmental Ethics)
		440105 History of Philosophy and History of Ideas
		440106 Logic
		440107 Metaphysics
		440108 Philosophy of Language
		440109 Philosophy of Mind (excl. Cognition)
		440110 Social Philosophy
		440111 Phenomenology
		440112 Hermeneutic Theory
		440113 Comparative Philosophy
		440114 Philosophy of Action
		440115 Philosophy of Specific Cultures (e.g. Ancient Greek\, Chinese)
		440199 Philosophy not elsewhere classified
	440200 Religion and Religious Traditions	440201 Hindu Studies
		440202 Buddhist Studies
		440203 Jewish Studies
		440204 Christian Theology (incl. Biblical Studies and Church History)
		440205 Islamic Studies
		440206 Studies in Other Religious Traditions
		440207 Religion and Society
		440208 Psychology of Religion
		440209 Philosophy of Religion
		440299 Religion and Religious Traditions not elsewhere classified
	449900 Other Philosophy and Religion	449999 Other Philosophy and Religion

20.3 Glossaire

AAI	AAI : <i>Authentication and Authorization Infrastructure</i> . Le projet AAI de Switch (Swiss education and Research Network) vise à simplifier l'accès inter-organisationnel aux services en réseau en utilisant un concept appelé <i>Federated identity management</i> .
AI	Format Adobe Illustrator : une version strictement limitée et très simplifiée du format EPS. Si EPS peut contenir pratiquement n'importe quelle commande PS, AI est limité à un petit nombre de commandes de dessin. [voir sous http://www.fileformat.info/format/ai/]
AIFF	<p>AIFF est un format de fichier audio numérique développé par Apple pour stocker les sons sur les ordinateurs de la marque. AIFF est un acronyme qui signifie Audio Interchange File Format.</p> <p>Ce format n'est pas sans rappeler le format WAV de Microsoft.</p> <p>Les fichiers AIFF portent généralement l'extension .aif ou .aiff.</p> <p>Les données sont codées en PCM big-endian sans compression. Ainsi, une piste CD Audio, codée en 16 bits, stéréo 44,1 kHz aura un bitrate de 1411,2 kbit/s. Il existe néanmoins un format compressé (AIFF-C ou AIFC) qui supporte une compression pouvant aller jusqu'à un rapport 1/6. [Wikipedia]</p>
API	Une API (application programming interface) est une interface de code source fournie par un ordinateur ou une bibliothèque de programme en vue de supporter des requêtes pour des services qui peuvent l'employer par des programmes informatiques. [Wikipedia]
ASCII	<p>La norme ASCII [aski:] (<i>American Standard Code for Information Interchange</i>) est la norme de codage de caractères en informatique la plus connue et la plus largement compatible. C'est également la variante américaine du codage de caractères ISO/CEI 646. ASCII contient les caractères nécessaires pour écrire en anglais. Elle a été inventée par l'américain Bob Bemer en 1961. Elle est à la base de nombreuses autres normes (Unicode, ISO 8859-1, Windows-1252), qui l'étendent.</p> <p>L'ASCII définit 128 caractères numérotés de 0 à 127 et codés en binaire de 0000000 à 1111111. Sept bits suffisent donc pour représenter un caractère codé en ASCII. Toutefois, les ordinateurs travaillant presque tous sur huit bits (un octet) depuis les années 1970, chaque caractère d'un texte en ASCII est stocké dans un</p>

	<p>octet dont le 8e bit est 0.</p> <p>Les caractères de 0 à 31 ainsi que le 127 ne sont pas affichables et correspondent à des directives de terminal. Le caractère 32 est l'espace blanc. Les autres correspondent aux chiffres arabes, aux lettres latines majuscules et minuscules et à quelques symboles de ponctuation. [Wikipedia]</p>
BiBTeX	<p>BibTeX est un logiciel et un format de fichier conçu par Oren Patashnik et Leslie Lamport en 1985 pour LaTeX. Il sert à gérer et traiter des bases bibliographiques. [Wikipedia]</p>
BMP	<p>Le format BMP, Windows Bitmap, est un format d'image numérique ouvert développé par Microsoft et IBM. C'est un des formats d'images les plus simples à développer et à utiliser pour programmer. Il est lisible par quasiment tous les visualiseurs et éditeurs d'images. [Wikipedia]</p>
checksum	<p>La somme de contrôle est un concept de la théorie des codes utilisé pour les codes correcteurs, elle correspond à un cas particulier de contrôle par redondance. Elle est largement utilisée en informatique et en télécommunications numériques.</p> <p>Les codes utilisant les sommes de contrôle permettent de valider un message. Si le nombre d'altérations durant la transmission est suffisamment petite, alors elles sont détectées. L'utilisation d'une unique somme de contrôle permet la détection mais non la correction des erreurs.</p> <p>Une somme de contrôle est un moyen simple pour garantir l'intégrité de données en détectant les erreurs lors d'une transmission de données dans le temps (sur un support de données) ou dans l'espace (télécommunications). Le principe est d'ajouter aux données des éléments dépendant de ces dernières - on parle de redondance - et simple à calculer. Cette redondance accompagne les données lors d'une transmission ou bien lors du stockage sur un support quelconque. Plus tard, il est possible de réaliser la même opération sur les données et de comparer le résultat à la somme de contrôle originale, et ainsi conclure sur la corruption potentielle du message.</p> <p>Un cas particulier répandu dans l'industrie est celui du bit de parité. C'est une somme de contrôle dans le cas où l'alphabet comporte deux lettres zéro et un. [Wikipedia]</p>
CRL	<p><i>Center for Research Libraries</i>. Le CRL est un consortium nord-américain d'universités, de collèges et de bibliothèques indépendantes. Le consortium achète et conserve des journaux, des périodiques, des documents, des archives et d'autres ressources traditionnelles et numériques pour la recherche et</p>

	<p>l'enseignement. Ces ressources sont ensuite mise à disposition aux institutions membres de manière coopérative, par prêt interurbain ou accès électronique. Voir le site du CRL sous : http://www.crl.edu/content.asp?l1=1</p>
<i>datastream</i>	<p>Une séquence d'octets. Ce terme est utilisé à la place de « fichier » pour décrire une séquence d'octet qui n'est qu'une partie d'un fichier. On utilise ce terme lorsque l'on veut accentuer le fait qu'une portion de fichier peut être consommée de manière immédiate, en n'apparaissant jamais aux yeux du consommateur comme un fichier sauvegardé. [http://www.w3.org/TR/PNG-Glossary.html]</p>
Dublin Core	<p><i>Dublin Core</i> est un schéma de métadonnées générique qui permet de décrire une ressource numérique ou physique et d'établir des relations avec d'autres ressources. Il comprend 15 éléments de description formels (titre, auteur, éditeur), intellectuels (sujet, description, langue...) et relatifs à la propriété intellectuelle.</p> <p>La signification de certains éléments peut être précisée à l'aide de raffinements. Un raffinement restreint la signification d'un élément, mais sans la changer fondamentalement. L'utilisation des raffinements est facultative. [Wikipedia]</p>
EPS	<p><i>Encapsulated PostScript</i> est un format créé par Adobe Systems en langage PostScript qui permet de décrire des images qui peuvent être constituées d'objets vectoriels et/ou bitmap. [Wikipedia]</p>
FEDORA	<p><i>Flexible Extensible Digital Object and Repository Architecture</i>. Un logiciel open source comprenant une architecture flexible et orientée service pour la gestion et la mise à disposition de contenus numériques.</p>
FOXML	Fedora Object XML
FRBR	Functional Requirements for Bibliographic Records.
GIF	<p>Le <i>Graphics Interchange Format</i> (littéralement « format d'échange de graphiques »), plus connu sous l'acronyme GIF (prononcé jif ou guif), est un format d'image numérique couramment utilisé sur le World Wide Web. GIF a été mis au point par CompuServe en 1986 pour permettre le téléchargement d'images en couleur. Ce format utilise l'algorithme de compression LZW, nettement plus efficace que l'algorithme RLE utilisé par la plupart des formats alors disponibles (PCX, ILBM puis BMP).</p> <p>En théorie GIF supporte 16 777 216 nuances de couleur : 8 bits par composante RVB, soit 224 nuances. Toutefois GIF n'enregistre pas directement la couleur de chaque pixel, car pour chaque image, une palette de 2 à 256 couleurs est construite. Ensuite</p>

	<p>chaque pixel de l'image fait référence à l'une des entrées de la palette. Cette méthode limite donc à 256 le nombre maximal de couleurs différentes présentes dans chaque calque d'une même image (GIF89a supportant les calques multiples). On parle donc de format 8 bits car chaque pixel est codé sur 8 bits, représentent donc au maximum 256 valeurs. L'usage d'une palette permettait un affichage beaucoup plus rapide sur les ordinateurs de l'époque dont les cartes graphiques contenaient elles-mêmes une palette d'au plus 256 couleurs.</p> <p>La limitation à 256 couleurs n'est pas gênante pour les logos, les graphiques et la plupart des images synthétiques, ainsi que les photographies noir et blanc. En revanche une photographie couleur de qualité nécessite plus de nuances.</p> <p>GIF permet de spécifier qu'une entrée de la palette est transparente. C'est notamment utile lorsqu'une image non rectangulaire est intégrée à un document comme une page Web : on voit le document à travers les pixels transparents. GIF propose un mode entrelacé permettant de commencer par transmettre quelques lignes d'une image, puis les lignes placées entre elles. Ce mode permet de donner plus rapidement un aperçu de l'image lorsque la transmission est lente.</p> <p>En 1989, le format GIF a été étendu (format GIF89a au lieu de GIF87a) pour permettre le stockage de plusieurs images dans un fichier. Ceci permet de créer des diaporamas, voire des animations si les images sont affichées à un rythme suffisamment soutenu. Chaque image d'une animation peut avoir sa propre palette, ce qui permet de créer des images contenant 16 777 215 couleurs simultanément, mais d'une taille non-négligeable). [Wikipedia]</p>
header	<p>En informatique, le header (« en-tête ») se réfère à la part de données d'un fichier placé au début d'un bloc sauvegardé ou transmis, et contenant des informations sur la manière de lire le bloc de données. [Wikipedia]</p>
html	<p>L'<i>Hypertext Markup Language</i>, généralement abrégé HTML, est le langage informatique créé et utilisé pour écrire les pages Web. HTML permet en particulier d'insérer des hyperliens dans du texte, donc de créer de l'hypertexte, d'où le nom du langage. Techniquement, HTML est une application du <i>Standard Generalized Markup Language</i> (SGML). Le développement de HTML a été interrompu en 1999 au profit de celui du XHTML, qui est une application de l'<i>Extensible Markup Language</i> (XML). Le W3C a cependant relancé le développement HTML en 2007, suite notamment aux demandes des fabricants de navigateurs. [Wikipedia]</p>

<p>ISO 14721:2003</p>	<p>L'ISO 14721:2003 spécifie un modèle de référence pour un système ouvert d'archivage d'information (OAIS). Le but de l'ISO 14721:2003 est d'établir un système pour archiver l'information, aussi bien numérique que physique, avec une structure d'organisation composée de personnes qui acceptent la responsabilité de préserver l'information et de la rendre disponible à un groupement désigné.</p> <p>Le modèle de référence aborde toute une série de fonctions liées à la conservation d'informations archivées comprenant le versement des données, le stockage, la gestion des données, l'accès et la diffusion. Il aborde également la migration d'informations numériques vers des supports et des formats nouveaux, les modèles de données utilisés pour représenter l'information, le rôle du logiciel dans la conservation de l'information, ainsi que l'échange d'informations numériques entre archives. Il identifie les interfaces internes et externes des fonctions d'archivage ainsi qu'un certain nombre de services de haut niveau relatifs à ces interfaces. Le présent modèle de référence présente divers exemples concrets, ainsi que quelques recommandations de "bonnes pratiques". Il définit les responsabilités minimales auxquelles une archive doit souscrire pour être qualifiée d'OAIS, ainsi qu'un profil maximal, en s'appuyant sur un large ensemble de termes et de concepts utiles.</p> <p>Le modèle d'OAIS exposé dans l'ISO 14721:2003 peut être applicable à toutes les archives. Il s'applique plus particulièrement aux organisations chargées de rendre l'information disponible sur le long terme. Ces organisations peuvent avoir d'autres responsabilités, telles que le traitement et la diffusion de données en réponse à des obligations programmatiques. [Site ISO]</p>
<p>Java</p>	<p>Java est une technologie développée par Sun Microsystems : (la technologie Java™). Elle correspond à plusieurs produits et spécifications de logiciels qui, ensemble, constituent un système pour développer et déployer des applications. Java est utilisée dans une grande variété de plates-formes depuis les systèmes embarqués et les téléphones mobiles jusqu'aux serveurs et aux superordinateurs. Java est largement présente dans les téléphones mobiles, les serveurs Web et les applications d'entreprise, et moins utilisée pour les interfaces graphiques, bien que l'on rencontre sur le Web des applets Java. Depuis des années, Sun Microsystems appelle Java la « technologie Java » dans son ensemble. En pratique, beaucoup de programmeurs utilisent le mot « Java » pour désigner le langage de programmation, tandis que la plate-forme d'exécution est appelée « JRE » (Java Runtime Environment, environnement d'exécution Java) et le système de compilation : « JDK » (Java Development</p>

	Kit) plutôt que « compilateur Java ». Java est sous licence GNU GPL depuis novembre 2006. [Wikipedia]
JHOVE	JSTOR/Harvard Object Validation Environment. Projet collaboratif entre JSTOR et la bibliothèque de l'Université de Harvard pour développer un framework extensible pour la validation des formats. JHOVE propose des fonctionnalités pour l'identification, la validation et la caractérisation d'objets numériques spécifiquement au format. [Voir le site JHOVE sous : http://hul.harvard.edu/jhove/]
JPEG	<p>acronyme de Joint Photographic Experts Group, nom d'un comité créé en 1986 résultant de la fusion de plusieurs groupes (d'où le joint) de professionnels de l'industrie de l'image. Ce comité a donné son nom à la norme ouverte de compression d'images numériques JPEG. Ensuite cette norme a donné son nom au format de données défini et au format de fichier le plus utilisé pour contenir ces données.</p> <p>Généralement, le terme JPEG fait référence à la norme de compression JPEG, formellement ISO/CEI 10918-1 ou UIT-T Recommandation T.81. Cette norme décrit une méthode de compression basée sur la DCT définie initialement par le projet européen ESPRIT PICA (Photographic Image Coding Algorithm). Elle spécifie uniquement la transformation d'une image brute en une suite de bits. Elle ne spécifie donc pas directement comment stocker les informations sur ses dimensions, son auteur, etc. Cela est le rôle d'un format de fichier. On utilisera généralement le JPEG File Interchange Format (JFIF) pour cela.</p> <p>Le JPEG est un format à perte, qui élimine donc des informations, mais un des points forts de JPEG est que son taux de compression est réglable. Un compromis doit cependant être fait entre le taux de compression et la qualité de l'image comprimée. En d'autres termes, le taux de compression ne doit pas être trop élevé, ni l'opération de compression être trop souvent répétée, sous peine de nuire gravement à la qualité générale de l'image. Certains logiciels offrent plusieurs choix pré-programmés de compression et d'autres permettent de l'affiner très précisément.</p> <p>Le format JPEG sauvegarde davantage d'informations couleur que le format GIF et garantit de ce fait un nombre élevé de couleurs. La compression flexible rend possible une réduction de la taille du fichier JPEG sans avoir trop d'impact sur la qualité de l'image. [Wikipedia]</p>
JPEG2000	JPEG 2000 ou ISO/CEI 15444-1 est une norme commune de l'ISO et de l'UIT-T. C'est un standard de compression d'images défini par le comité Joint Photographic Experts Group. JPEG 2000 est

capable de travailler avec ou sans pertes, utilisant une transformation en ondelettes (méthode d'analyse mathématique du signal). En compression irréversible, JPEG 2000 est plus performante que la méthode de compression JPEG ISO/CEI 10918-1 (le JPEG classique). On obtient donc des fichiers d'un poids inférieur pour une qualité d'image égale. De plus, les contours nets et contrastés sont mieux rendus.

La norme est divisée en 12 sous-parties. Bien que l'appellation JPEG 2000 couvre l'ensemble de ces sous-parties, c'est en général à la première partie que l'on fait référence, le « cœur » du système. JPEG normalise uniquement l'algorithme et le format de décodage. Le processus de codage est laissé libre à la compétition des industriels et universitaires, du moment que l'image produite est décodable par un décodeur standard. La norme propose un jeu de fichiers de tests appelés fichiers de conformance qui permettent de vérifier qu'un décodeur respecte bien la norme. Un décodeur est alors dit conforme s'il est capable de décoder tous les fichiers de conformance. C'est l'objet de la partie 4 du standard.

- * Partie 1 : cœur du système de codage
- * Partie 2 : extensions
- * Partie 3 : Motion JPEG 2000
- * Partie 4 : conformance
- * Partie 5 : logiciel de référence
- * Partie 6 : compound image file format
- * Partie 7 : a été abandonnée
- * Partie 8 : JPSEC sécurité
- * Partie 9 : JPIP protocole interactif
- * Partie 10 : JP3D imagerie volumétrique
- * Partie 11 : JPWL wireless
- * Partie 12 : ISO Base Media File Format (commun avec MPEG-4)

Les performances en compression de JPEG 2000, bien que meilleures que JPEG à bas débit ne sont pas révolutionnaires. Ce sont plutôt la multitude de nouvelles caractéristiques telles la scalabilité, les régions d'intérêt, la résistance aux erreurs de transmission, le codage sans pertes, la versatilité de l'organisation du codestream, ainsi que les diverses extensions visant une application (interactivité, sécurité, sans fil, etc.) qui font l'intérêt de la norme.

De par ses fonctionnalités avancées, sa capacité à gérer les images de grande taille, ainsi que d'excellentes performances à haut débit, JPEG 2000 a un fort potentiel pour les professionnels de l'image. Sa carrière dans le grand public risque par contre

	d'être moins brillante que celle de son prédécesseur JPEG. [Wikipedia]
LaTeX	<p>LaTeX a été créé par Leslie Lamport, et est un système logiciel de composition de documents, ou plus exactement : une collection de macro-commandes destinées à faciliter l'utilisation du « processeur de texte » TeX.</p> <p>Du fait de sa relative simplicité, il est devenu la méthode privilégiée d'écriture de documents scientifiques employant TeX ; la version actuelle est Logo LaTeX2_ (2 epsilon). Il est particulièrement utilisé dans les domaines techniques et scientifiques pour la production de documents de taille moyenne ou importante (thèse ou livre). Néanmoins, il peut aussi être employé pour générer des documents de types variés (par exemple des lettres ou des transparents). [Wikipedia] (voir aussi sous « TeX »).</p>
LDA	Loi sur le Droit d'Auteur (dans ce rapport : La Loi Suisse sur le Droit d'Auteur).
LDAP	<p><i>Lightweight Directory Access Protocol</i> (LDAP) est un protocole permettant l'interrogation et la modification des services d'annuaire. Ce protocole repose sur TCP/IP. Un annuaire LDAP respecte généralement le modèle X.500 édicté par l'UIT-T : c'est une structure arborescente dont chacun des nœuds est constitué d'attributs associés à leurs valeurs.</p> <p>Le nommage des éléments constituant l'arbre (racine, branches, feuilles) reflète souvent le modèle politique, géographique ou organisationnel de la structure représentée. La tendance actuelle est d'utiliser le nommage DNS pour les éléments de base de l'annuaire (racine et premières branches). Les branches plus profondes de l'annuaire peuvent représenter des personnes (people), des unités organisationnelles (organizational units), des groupes (groups). [Wikipedia]</p>
METS	<i>Metadata Encoding and Transmission Standard</i> . Le schéma METS est un standard pour encoder des métadonnées descriptives, administratives et structurales, pour des objets d'une bibliothèque numérique. METS utilise le langage XML schema du consortium W3C. Le standard est maintenu par le Network Development and MARC Standard Office de la Bibliothèque du Congrès. [Voir le site METS sous : http://www.loc.gov/standards/mets/]
MIME type	Les MIME Types sont une possibilité qu'ont les logiciels de savoir à quel type de fichiers ils ont à faire sans avoir à regarder dans les contenus du fichier et sans dépendre d'éléments dépendant des

	<p>systèmes d'exploitations comme les extensions (.doc, ou .jpeg) [Voir sous : http://www.fileformat.info/info/mimetype/index.htm]</p>
MODS	<p><i>Metadata Object Description Schema</i>. Un XML Schema développé par la Bibliothèque du Congrès et le Network Development and MARC Standards Office avec un groupe d'experts. Il contient un sous-ensemble de champs d'une notice MARC 21 permettant soit d'extraire des données de notices MARC 21, soit de créer des descriptions de ressources originales. [Voir le site MODS, sous : http://www.loc.gov/standards/mods/]</p>
MOV	<p>Extension du format Quicktime. QuickTime est un framework multimédia développé par Apple. Cette technologie a été créée en 1989, puis mise sur le marché en 1991 pour le Macintosh. Elle fut portée sur Microsoft Windows en 1992 afin d'accélérer son adoption par l'industrie. Un fichier QuickTime contient une ou plusieurs pistes, chacune comporte un type de données particuliers : audio, vidéo, effet ou texte (pour des sous-titres par exemple). Chaque piste contient une piste media, soit le stream codé numériquement (avec un codec tel Cinepak, Sorenson, MP3, JPEG, DivX...) ou une référence à un media situé dans un autre fichier ou sur un réseau.</p> <p>En interne, les fichiers QuickTime maintiennent ce format dans un arbre d'atomes. Chacun utilise un identifiant de quatre octets pour déterminer sa structure. Un atome peut être le parent d'autres atomes ou peut contenir des données (mais il ne peut être les deux à la fois). [Wikipedia]</p>
MPEG	<p>MPEG, acronyme de <i>Moving Picture Experts Group</i> est le groupe de travail SC 29/WG 11 du comité technique mixte JTC 1 de l'ISO et de la CEI pour les technologies de l'information chargé du développement de normes internationales pour la compression, la décompression, le traitement et le codage de la vidéo, de l'audio et de leur combinaison, de façon à satisfaire un large panel d'applications.</p> <p>Les réunions du Moving Picture Experts Group ont démarré en 1988 dans le but de développer un premier standard, MPEG-1, pour des applications de stockage audio/vidéo du type Video CD. MPEG a ensuite rapidement produit un second standard, MPEG-2, visant essentiellement les applications liées à la télévision numérique. D'autres familles de standards ont depuis été produites, MPEG rassemblant un nombre croissant de spécialistes provenant de l'industrie de l'électronique, de l'informatique et des télécommunications.</p> <p>Les formats produits par MPEG sont ouverts, mais non libres : leur utilisation est soumise au paiement de redevances.</p> <p>MPEG a développé les standards suivants :</p> <ul style="list-style-type: none"> * MPEG-1 : leur premier standard audio et vidéo utilisé plus

	<p>tard comme standard des Vidéo CDs . Ce format offre une résolution à l'écran de 352 x 240 pixels à 30 images par seconde ou de 352 x 288 à 25 images par seconde avec un débit d'environ 1,5 Mbit/s..Il inclut le populaire format audio MPEG-1 Layer 3 (MP3).</p> <ul style="list-style-type: none"> * MPEG-2 : standard couvrant le codage de l'audio et la vidéo, ainsi que leur transport pour la télévision numérique : télévision numérique par satellite, télévision numérique par câble, télévision numérique terrestre, et (avec quelques restrictions) pour les vidéo-disques DVD ou SVCD. C'est notamment le format utilisé jusqu'à présent pour la TV sur ADSL. Les débits habituels sont de 2 à 6 Mbit/s pour la résolution standard (SD), et de 15 à 20 Mbit/s pour la haute résolution (HD) * MPEG-4 : norme comblant le vide des bas débits (jusqu'à 2 Mbit/s) pour lesquels MPEG-2 n'avait pas été développé. Il permet entre autres de coder des objets vidéo/audio, le contenu 3D et supporte le DRM. La partie 2 de MPEG-4 (Visual) est compatible avec la partie baseline de H.263 et a connu du succès grâce à l'implémentation DivX ainsi que dans les téléphones mobiles. La partie 10 appelée MPEG-4 AVC permet des gains d'un facteur 2 à 3 par rapport à MPEG-2 et a déjà été retenu comme le successeur de celui-ci pour la TV haute définition, la TV sur ADSL et la TNT. L'extension de cette partie, appelée Scalable Video Coding (SVC), est en cours de définition et devrait être finalisée en 2007. Elle permettra de proposer différents niveaux de qualité à partir d'un même flux codé. * MPEG-7 : un standard pour décrire et chercher du contenu multimédia. * MPEG-21 : un standard proposant une architecture pour l'interopérabilité et l'utilisation simple de tous les contenus multimédia. * MPEG-A : tourné vers les applications multimédia. En cours de standardisation. <p>[Wikipedia][Voir aussi sous : http://www.fileformat.info/format/mpeg/egff.htm]</p>
<p>OA</p>	<p>Open Access Initiative. Initiative lancée en 2002 et encourageant l'accès gratuit par l'Internet à la part de littérature scientifique que les scientifiques produisent sans attentes d'être payés. Il s'agit d'abord des articles publiés dans des revues spécialisées peer-reviewed, mais également des préprints que les auteurs souhaitent communiquer à leurs collègues.</p> <p>Par "open access" on entend l'accès gratuit à cette littérature par Internet public, ce qui permet de lire, télécharger, copier, distribuer, imprimer, rechercher ou établir un lien sur le full text de ces articles, de les indexer, de les inclure dans des banques de données ou de les utiliser à toutes fin utiles et légales sans barrières financières, légales ou techniques autres que celles que d'avoir accès à l'Internet. La seule contrainte pour la reproduction</p>

	<p>et la distribution et la seule règle de copyright dans ce domaine est de donner à l'auteur le contrôle sur l'intégrité de son travail et le droit d'être correctement reconnu et cité.</p>
OAI	<p>L'Open <i>Archives</i> Initiative développe et promeut les standards d'interopérabilité dans le dessein de faciliter la diffusion efficace des contenus. Elle prend racines dans l'effort de renforcer l'accès aux archives électroniques en vue d'améliorer la disponibilité des publications scientifiques. Les standards développés sont indépendants du type de contenus et des mécanismes économiques entourant les contenus. Ils se basent sur le Dublin Core et XML.</p> <p>Actuellement le comité exécutif de l'OAI est formé par Carl Lagoze (Senior Research Associate, Cornell Computing and Information Science) et Herbert van de Sompel (Digital Library Researcher à la Research Library of the Los Alamos National Laboratory).</p>
OAI-ORE	<p>Open Archive Initiative – Object Reuse and Exchange. Un projet de l'Open Archives Initiative, actuellement en cours. ORE développe des spécifications qui permettront à des archives distribuées d'échanger des informations sur leurs objets numériques. Ces spécifications inclueront des approches pour représenter les objets numériques et les services des serveurs institutionnels pour faciliter l'accès et la saisie de ces représentations. Les spécifications permettront de créer une nouvelle génération de services entre serveurs institutionnels rehaussant la valeur intrinsèque des objets numériques au-delà des limites des serveurs particuliers. [Voir le site OAI-ORE sous : http://www.openarchives.org/ore/]</p>
OAI-PMH	<p>L'Open Archive Initiative Protocol for Metadata Harvesting est le protocole de moisson des données élaborées par l'OAI. La dernière version est la 2.0.</p>
OAIS	<p>OAIS (Open Archival Information System) est un modèle conceptuel destiné à la gestion, à l'archivage et à la préservation longue de documents numériques. La mise au point de OAIS a été pilotée par le CCSDS (Comité Consultatif pour les Systèmes de Données Spatiales). L'OAIS est enregistré norme ISO sous la référence 14721:2002.</p> <p>Le modèle OAIS constitue une référence décrivant dans les grandes lignes les fonctions, les responsabilités et l'organisation d'un système qui voudrait préserver de l'information, en particulier des données numériques, sur le long terme, pour en garantir l'accès à une communauté d'utilisateurs identifiés. Le long terme est défini comme suffisamment long pour être soumis à l'impact des évolutions technologiques.</p>

	<p>Les principales parties de la norme OAIS sont :</p> <ul style="list-style-type: none"> • le glossaire et la définition des principaux concepts • les responsabilités liées à la mise en place d'une archive OAIS • deux modèles détaillés, le modèle fonctionnel et le modèle d'information • les perspectives de pérennisation • l'interopérabilité des archives OAIS. <p>[Wikipédia]</p>
OAISTER	<p>OAISTER est un catalogue collectif de ressources numériques donnant accès à ces ressources par harvesting des métadonnées par OAI-PMH. Actuellement, OAISTER enregistre 11'315'096 notices provenant de 769 serveurs. [Voir le site OAISTER sous : http://www.oaister.org/]</p>
Objet	<p>Concept propre à la programmation par objet (du terme anglo-saxon <i>Object-Oriented Programming</i> ou OOP), un paradigme de programmation, qui consiste en la définition et l'assemblage de briques logicielles appelées objets ; un objet représente un concept, une idée ou une chose du monde réel, de l'entreprise ou celui des idées.</p> <p>Concrètement, un objet est une structure de données valuées et qui répond à un ensemble de messages. Cette structure de données définit son état tandis que l'ensemble des messages qu'il comprend décrit son comportement :</p> <ul style="list-style-type: none"> * Les données ou champs qui décrivent sa structure interne sont appelées ses attributs ; * L'ensemble des messages forme ce que l'on appelle l'interface de l'objet ; c'est seulement au travers de celui-ci que les objets interagissent entre eux. La réponse à la réception par un objet d'un message est appelée une méthode (méthode de mise en œuvre du message) ; elle décrit comment est réalisé le message. <p>Les attributs (ou plus exactement leur représentation informatique) et les méthodes sont cachés ; ils forment une boîte noire. C'est le principe d'encapsulation. Son avantage principal réside dans la capacité à pouvoir modifier la structure interne des objets ou les méthodes associées aux messages sans en impacter les utilisateurs des objets.</p> <p>Exemple d'un objet représentant un nombre complexe donné. Celui-ci peut aussi bien être représenté sous forme cartésienne (réel, image) que sous forme trigonométrique ou exponentielle (module, angle). Cette représentation, quelle qu'elle soit, est cachée et est interne à l'objet sous forme d'une structure de données. Toutefois, celui-ci peut proposer deux messages</p>

	<p>permettant chacun de lire une représentation différente du nombre complexe. En utilisant les seuls messages que comprend notre nombre complexe, les objets appelants sont assurés de ne pas être impactés lors d'un changement de sa structure interne. Cette dernière n'est accessible que par (et donc aussi au travers) les méthodes des messages. [Voir sous : http://www.techno-science.net/?onglet=glossaire&definition=5393]</p>
OCLC	<p>Online Computer Library Center. Organisation de services bibliothécaires en ligne à but non-lucratif qui a pour but de favoriser l'accès public à l'information mondiale en réduisant les coûts d'accès à cette information. [Voir le site d'OCLC sous : http://www.oclc.org/about/default.htm]</p>
OpenDOAR	<p>Le Directory of Open Access Repositories, qui recense les serveurs proposant des contenus en Open Access. Ce recensement n'est pas fait de manière automatique. On s'inscrit et l'équipe d'OpenDOAR vérifie les informations reçues dans la soumission. Voir le site d'OpenDOAR sous : http://www.opendoar.org/</p>
ORE	<p>Voir « OAI-ORE »</p>
pathname	<p>Le <i>pathname</i> – « nom du chemin » est défini comme la suite de caractères à entrer par un utilisateur pour identifier un fichier. Normalement le pathname contient l'identité du hardware et/ou des dossiers et le nom spécifique du fichier. Exemple : C :\mondossier\monfichier.doc.</p>
PDF	<p>Le <i>Portable Document Format</i> (communément abrégé PDF) est un format de fichier informatique créé par Adobe Systems. L'avantage du format PDF est qu'il préserve les polices, les images, les objets graphiques et la mise en forme de tout document source, quelles que soient l'application et la plate-forme utilisées pour le lire. Les fichiers PDF peuvent être créés avec des options personnalisées, tant aux niveaux de la compression des images et des textes, de la qualité d'impression du fichier, que du verrouillage (interdiction d'impression, de modification...).</p> <p>Le format de fichier permet en effet d'intégrer dans le fichier PDF les polices utilisées pour la création du document. Ces polices sont cependant parfois soumises aux droits d'auteurs et ne sont pas toujours redistribuables. La police doit alors être remplacée par une police équivalente (mais pas toujours identique).</p> <p>Le format PDF peut aussi être interactif : il est en effet possible (grâce à des logiciels de génération de PDF tels Acrobat Profesional ou Scribus) d'incorporer des champs de textes, des menus déroulants, des choix, des calculs, etc. On parle alors de formulaire PDF.</p>

La génération de documents PDF est possible à l'aide de logiciels spécialisés, d'imprimantes virtuelles mais elle est également possible automatiquement dans certaines suites bureautiques. Le PDF s'est imposé comme format d'échange et d'archivage de documents électroniques.

Pour garantir la portabilité, quelques précautions sont cependant à prendre lors de la génération du fichier PDF. Il s'agit d'utiliser (et d'intégrer dans le PDF) des polices redistribuables. Pour ceci, la spécification PDF indique qu'un ensemble minimal de 14 polices est fourni en standard (il s'agit de Times, Helvetica, Courier, Symbol et Zapf Dingbats avec leurs variantes italiques et grasses quand disponible). Ces dernières, quand elles sont utilisées, garantissent alors la portabilité du document sans en surcharger la taille.

Le PDF est consultable sur de très nombreux appareils communicants (ordinateurs, assistants personnels numériques / PDA, nouveaux téléphones hybrides...). En effet, le lecteur diffusé gratuitement par Adobe, nommé « Adobe Reader » est disponible sur de très nombreuses plateformes et systèmes d'exploitation : Windows, Mac OS, Linux, Palm OS, Pocket PC, Symbian OS, Sun Solaris Sparc, IBM AIX, HP-UX, OS/2 / Warp ... De nombreux autres lecteurs dont certains sont des logiciels libres, existent également.

PDF propose un système de gestion des droits d'auteurs, qui permet de limiter la lecture ou la modification du document. A l'image de l'industrie de la musique ou du cinéma, l'industrie du livre s'intéresse à cette technologie car elle lui permettrait de diffuser des oeuvres soumises au droit d'auteur en ayant la garantie que le fichier ne se diffuserait pas en dehors des ordinateurs "autorisés".

Outre le débat qui entoure aujourd'hui les systèmes de gestion numérique des droits, certains s'interrogent sur l'ergonomie du procédé : lire un livre sur écran d'ordinateur n'est pas la solution la plus agréable ni celle qui répond actuellement aux habitudes du lecteur, qui désire pouvoir lire sans contrainte de batterie et en tournant ses pages plutôt que faire défiler un ascenseur.

De nouveaux procédés, comme le papier électronique sortent actuellement des laboratoires et rentrent en production, donnant un nouveau souffle à ces débouchés du PDF.

PDF est un format ouvert, c'est-à-dire que son créateur Adobe Systems autorise des programmes tiers à réutiliser son format. Il se réserve cependant la propriété de nombreux brevets déposés, et donc le droit de demander des royalties. Il reste très tolérant à

	<p>ce sujet : en effet, les spécifications sont publiques et utilisables librement (certains éléments sont à disposition sur le site Adobe) et gratuitement (à l'exception de quelques très importantes sociétés commerciales auxquelles Adobe demande alors une participation financière).</p> <p>Il existe plusieurs versions du format PDF, la dernière étant la version numérotée 1.7. On trouve aussi une version dite « rastérisées » (la rastérisation consiste à transformer les textes modifiables en images matricielles figées accompagnées des images d'illustration) nommée PDF-IT et PDF-X. Cette version est principalement utilisée dans l'industrie graphique lors de la séparation des couleurs en quadrichromie, au niveau du RIP. Deux sous-ensembles du format PDF ont été normalisés par l'ISO, il s'agit des format PDF/A-1 (PDF for Archive, référencé par la norme ISO 19005-1) et PDF/X (PDF for Exchange). Deux autres sous-ensembles sont actuellement en proposition pour devenir des standards : PDF/E (PDF for Engineering) et PDF/UA (PDF for Universal Access).</p> <p>Le 29 janvier 2007, Adobe a annoncé son intention de soumettre la spécification complète du format PDF (dans sa version 1.7) à l'AIIM pour une publication par le comité ISO. La durée de cette standardisation devrait être d'environ 3 ans. [Wikipedia]</p>
PID	Persistent Identifier (identificateur permanent)
PNG	<p>Le <i>Portable Network Graphics</i> est un format d'images numériques ouvert, qui a été créé pour remplacer le format propriétaire GIF, dont la compression était soumise à un brevet. Le PNG est un format non-destructeur spécialement adapté pour publier des images simples comprenant des aplats de couleurs. Il a été normalisé par l'ISO (ISO/CEI 15948 : 2004). PNG est particulièrement approprié lorsqu'il s'agit d'enregistrer des images synthétiques destinées au Web comme des graphiques, des icônes, des images représentant du texte (bonne conservation de la lisibilité), ou des images avec peu de dégradés. Le PNG surpasse régulièrement le format GIF tant en ce qui concerne la taille (avec une palette de couleurs bien choisie) que la qualité puisqu'il n'est pas limité à 256 couleurs.</p> <p>Les caractéristiques de PNG lui permettent d'enregistrer des photographies sans perte de données, mais au détriment de la taille du fichier résultant qui reste la plupart du temps très supérieure à celle de formats spécifiques aux photographies comme JPEG ou JPEG2000. [Wikipedia]</p>
PDD	Un format Adobe Photoshop, en mode bitmap.

<p>Plug-in</p>	<p>En informatique, le terme anglais plugin est employé pour désigner un programme qui interagit avec un logiciel principal, appelé programme hôte, pour lui apporter de nouvelles fonctionnalités.</p> <p>La plupart du temps, ces programmes sont caractérisés de la façon suivante :</p> <ul style="list-style-type: none"> * Ils ne peuvent fonctionner seuls car ils sont uniquement destinés à apporter une fonctionnalité à un ou plusieurs logiciels * Ils sont mis au point par des personnes n'ayant pas nécessairement de relation avec les auteurs du logiciel principal. <p>[Wikipedia]</p>
<p>Postscript (PS)</p>	<p>Le PostScript est un langage de programmation spécialisé dans la description de page, mis au point par Adobe. Il repose sur des formulations vectorielles de ses éléments.</p> <p>Ce langage inter-plateformes permet d'obtenir un fichier unique comportant tous les éléments décrivant la page (textes, images, polices, couleurs, etc.).</p> <p>PostScript est devenu pratiquement un standard, la plupart des imprimantes récentes peuvent traiter directement le format PostScript (NB : sur les imprimantes plus anciennes, il fallait utiliser un filtre logiciel en entrée pour convertir le langage PostScript au format raster compréhensible par les anciennes imprimantes).</p> <p>Postscript est un langage complet, qui permet le codage de tout algorithme. Bien qu'il soit tout à fait possible d'écrire directement de tels programmes, ils sont en général fabriqués par d'autres programmes, des pilotes d'impression par exemple.</p> <p>Le Postscript est indissociable de l'environnement dans lequel il sera exécuté. Étant donné le caractère totalement dynamique de ce langage, il sera en fait interprété.</p> <p>L'interpréteur est composé d'une pile, et d'un ou plusieurs dictionnaires. La pile sert de stockage temporaire pour les paramètres des fonctions, puis pour leurs résultats. Les dictionnaires permettent le stockage des variables, ainsi que du code des fonctions.</p> <p>Un programme PostScript est composé d'une séquence de mots séparés par des espaces, tabulation (TAB), retour chariot (CR), avance de ligne (LF), ou commentaires. L'interpréteur analyse chaque mot du programme PostScript séquentiellement en fonctionnant comme un calculateur en notation polonaise inverse, c'est-à-dire que chaque mot est évalué, puis le (ou les) résultat de cette évaluation est placé au sommet de la pile, et ainsi de suite.</p>

	PostScript est sous licence de la société Adobe. Néanmoins il existe un interpréteur PostScript libre, Ghostscript. [Wikipedia]
PPT	Power Point : Logiciel et format de Microsoft permettant de créer des présentations sous forme de diaporamas.
PPUR	Presses Polytechniques et Universitaires Romandes.
PubMed	Un service de l'U.S. National Library of Medicine incluant plus de 17 millions de citations de MEDLINE et d'autres périodiques de sciences de la vie, pour des articles du biomédical. PubMed propose des liens sur les articles en full-text et d'autres ressources. Voir le site de PubMed sous : http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed .
PubMed Central	<i>PubMed Central</i> est une archive numérique libre de périodiques du domaine du biomédical et des sciences de la vie de <i>U.S. National Institutes of Health (NIH)</i> , développé et géré par le <i>NIH's National Center for Biotechnology Information (NCBI)</i> dans la <i>National Library of Medicine (NLM)</i> . PubMed Central préserve et maintient un accès sans restrictions à la littérature électronique. Voir le site PubMed Central sous : http://www.pubmedcentral.nih.gov/about/intro.html
PSD	Le format .psd (Photoshop Document) est le format natif du logiciel Adobe Photoshop. Il est reconnu par plusieurs logiciels de traitement d'image. C'est le format de travail de référence de photoshop car il permet de conserver les calques, les couches de transparence etc. [Wikipedia]
RDF	<i>Resource Description Framework (RDF)</i> est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées, de façon à permettre le traitement automatique de telles descriptions. Développé par le W3C, RDF est le langage de base du Web sémantique. Une des syntaxes (sérialisation) de ce langage est RDF/XML. En annotant des documents non structurés et en servant d'interface pour des applications et des documents structurés (par ex. bases de données, GED, etc.) RDF permet une certaine interopérabilité entre des applications échangeant de l'information non formalisée et non structurée sur le Web. Un document structuré en RDF est un ensemble de triplets. Un triplet RDF est une association : { sujet, objet, prédicat}

	<ul style="list-style-type: none"> • Le sujet représente la ressource à décrire • Le prédicat représente un type de propriété applicable à cette ressource • L'objet représente une donnée ou une autre ressource <p>Le sujet, et l'objet dans le cas où c'est une ressource, peuvent être identifiés par une URI ou être des nœuds anonymes. Le prédicat est nécessairement identifié par une URI.</p> <p>Les documents RDF peuvent être écrits en différentes syntaxes, y compris en XML. Mais RDF en soi n'est pas un dialecte XML. Il est possible d'avoir recours à d'autres syntaxes pour exprimer les triplets. RDF est simplement une structure de données constituée de nœuds et organisée en graphe. Bien que RDF/XML — sa version XML proposée par le W3C — ne soit qu'une sérialisation du modèle, elle est souvent appelée RDF. Un abus de langage désigne à la fois le graphe de triplets et la présentation XML qui lui est associée.</p> <p>Un document RDF ainsi formé correspond à un multi-graphe orienté étiqueté. Chaque triplet correspond alors à un arc orienté dont le label est le prédicat, le nœud source est le sujet et le nœud cible est l'objet.</p> <p>La structure de RDF est extrêmement générique et sert de base à un certain nombre de schémas ou vocabulaires dédiés à des applications spécifiques. Une partie de ces vocabulaires est spécifiée par le W3C, comme les langages d'ontologie RDFS et OWL, ou le langage SKOS pour la représentation des thésaurus et autres vocabulaires structurés. D'autres vocabulaires RDF, sans être spécifiés par le W3C, sont néanmoins utilisés largement et constituent des standards de fait dans la communauté du Web Sémantique, comme par exemple FOAF qui est destiné à la description des personnes. [Wikipedia]</p>
REST	<p>REST (<i>Representational state transfer</i>) est une manière de construire une application pour les systèmes distribués comme le World Wide Web. Le terme a été inventé par Roy Fielding.</p> <p>REST n'est pas un protocole ou un format, c'est une architecture, c'est l'architecture originale du Web, bâtie sur quelques principes simples :</p> <ul style="list-style-type: none"> * l'URI est important : connaître l'URI doit suffire pour accéder à la ressource; * HTTP fournit toutes les opérations nécessaires (GET, POST, PUT et DELETE, essentiellement); * chaque opération est auto-suffisante : il n'y a pas d'état; * utilisation des standards hypermedia : HTML ou XML qui

permettent de faire des liens vers d'autres ressources et d'assurer ainsi la navigation dans l'application REST.

Cette architecture n'est pas limitée à la réalisation d'application pour un utilisateur humain. Elle est de plus en plus utilisée pour la réalisation de services Web destinés à la communication entre machines. Dans ce cadre là, les requêtes et les réponses sont typiquement encodées en XML. REST dans ce cas là se pose en alternative à RPC et SOAP, alternative censée être plus simple à mettre en œuvre. Les systèmes qui suivent les principes REST de Fielding sont souvent appelés RESTful.

La thèse de Roy Fielding précise les avantages de cette architecture par rapport à d'autres architectures d'applications web. Citons entres autres :

- * L'absence d'état sur le serveur conduit à une consommation de mémoire inférieure et donc à une capacité plus grande de répondre à un grand nombre de requêtes simultanées.

- * L'absence d'état sur le serveur rend le fonctionnement plus simple à appréhender. Le résultat d'une requête ne dépend pas de variables cachées difficilement identifiables. Cela conduit à une mise au point plus simple.

- * L'absence d'état sur le serveur permet une répartition des requêtes sur plusieurs serveurs avec une meilleure granularité et de manière plus souple. Cela permet aussi une meilleure tolérance aux pannes d'un des serveurs.

- * Le respect de la philosophie du protocole HTTP (à la différence de SOAP) conduit à une architecture plus cohérente et plus simple.

- * l'utilisation d'URI comme représentant d'une ressource, permet la mise en place de serveurs cache.

Le principal désavantage de REST est la nécessité pour le client de conserver localement toutes les données nécessaires au bon déroulement d'une requête, ce qui induit une consommation en bande passante réseau plus grande. Attention, il est cependant possible de coupler une application web REST à un service extérieur assurant la permanence des données, par exemple une base de données. On pourrait cependant considérer que l'utilisation d'un tel service pour gérer des données relatives à une session ouverte par le client serait en violation de la philosophie de REST. L'utilisation d'un tel service extérieur ne doit permettre que de conserver des données destinées à survivre à la session.

Plusieurs gros services sont aujourd'hui accessibles en REST comme Amazon.

Par ailleurs beaucoup d'applications, bien que ne respectant pas

	scrupuleusement toutes les contraintes de l'architecture REST sont largement inspirées par elle. [Wikipedia]
Shibboleth	Système d'authentification et d'autorisation défini par Internet2 utilisé en Suisse par Switch-AAI
SOAP	<p><i>Simple Object Access Protocol</i> est un protocole de RPC orienté objet bâti sur XML. Il permet la transmission de messages entre objets distants, ce qui veut dire qu'il autorise un objet à invoquer des méthodes d'objets physiquement situés sur une autre machine. Le transfert se fait le plus souvent à l'aide du protocole HTTP, mais peut également se faire par un autre protocole, comme SMTP.</p> <p>Le protocole SOAP est composé de deux parties :</p> <ul style="list-style-type: none"> • une enveloppe, contenant des informations sur le message lui-même afin de permettre son acheminement et son traitement, • un modèle de données, définissant le format du message, c'est-à-dire les informations à transmettre. <p>SOAP a été initialement défini par Microsoft et IBM, mais est devenu une référence depuis une recommandation du W3C, utilisée notamment dans le cadre d'architectures de type SOA (Service Oriented Architecture) pour les Web Services. Le protocole SOAP emploie des métadonnées. [Wikipedia]</p>
SSL	<i>Secure Socket Layer</i> (SSL), est un protocole de sécurisation des échanges sur Internet, développé à l'origine par Netscape (SSL version 2 et SSL version 3). Il a été renommé en Transport Layer Security (TLS) par l'IETF suite au rachat du brevet de Netscape par l'IETF en 2001. Le groupe de travail correspondant à l'IETF a permis la création de la RFC 2246. [Wikipedia]
tag	<p>Balise. Les langages de balisage (sous-classe des langages de description) représentent une classe de langages spécialisés dans l'enrichissement d'information textuelle. Ils opèrent grâce aux balises, unités sémantiques délimitant chacune un ensemble à l'intérieur d'un fichier texte, souvent en unicode.</p> <p>L'inclusion de balises permet de transférer à la fois la structure du document et son contenu. Cette structure est compréhensible par un programme informatique, ce qui autorise un affichage personnalisé selon des règles pré-établies ; la typographie (en premier lieu la fonte) et d'autres éléments de présentation peuvent changer. On peut de plus inclure des éléments non-textuels.</p> <p>Les langages de balisages les plus utilisés sur le Web sont des langages dérivés de SGML, comme HTML, XML et XSLT.</p>

	[Wikipedia]
TeX	<p>Système logiciel de composition de documents, largement utilisé par les scientifiques, particulièrement en mathématiques, physique et science informatique. Il a été créé à partir de 1977 par le mathématicien et informaticien Donald Knuth, excédé par la qualité de la typographie de l'époque, lorsqu'il était en train de réviser la première édition de <i>The Art of Computer Programming</i>. TeX interprète un langage de description formel de document défini par D. Knuth ; ce n'est pas un formateur de texte intuitif (WYSIWYG pour « What You See Is What You Get ») où le document en préparation apparaît à l'écran comme il est sensé être une fois imprimé, mais un logiciel traduisant du texte en un document graphique. L'utilisation directe de TeX étant assez ardue, il a été étendu en LaTeX, écrit à l'origine par Leslie Lamport, qui constitue en fait un jeu de macro-commandes basées sur TeX.</p>
TIFF	<p>Le <i>Tag(ged) Image File Format</i> généralement abrégé TIFF est un format de fichier pour image numérique. Le TIFF non compressé est un format courant et lu par tous les logiciels de traitement d'image matricielle. TIFF est un format extrêmement flexible :</p> <ul style="list-style-type: none"> * Il est notoirement connu pour permettre l'enregistrement des données multi-octets au format big endian ou little endian. * Il permet d'utiliser de nombreux types de compression, avec ou sans perte de données : brut, PackBits, LZW, CCITT Fax 3 et 4, JPEG. * Il supporte de nombreux codages des pixels, de 1 à 64 bits par pixel, signé ou non, ainsi que les formats en virgule flottante 32 et 64 bits définis par l'IEEE. * Il supporte de nombreux espaces colorimétriques : noir et blanc, monochrome, palette de couleurs (de toute taille), RVB, YCbCr, CMJN, CIE Lab. * Il supporte de nombreuses informations additionnelles sur les couleurs utiles à la calibration colorimétrique : correction gamma, etc. * Il supporte le stockage d'image par bloc, et aussi de multiples images par fichier, des images alternatives en basse résolution, des annotations sous forme de courbes et de texte, etc. <p>Cette considérable flexibilité fait que TIFF est utilisé dans des applications très diverses, des scanners industriels aux appareils photo numériques en passant par les imprimantes. Cela fait également que tous les logiciels ne sont pas capables d'afficher tous les types d'images TIFF. [Wikipedia]</p>
URI	<p>Un <i>Uniform Resource Identifier</i>, soit littéralement identifiant uniforme de ressource, est une courte chaîne de caractères</p>

	<p>identifiant une ressource Web physique ou abstraite, et dont la syntaxe respecte une norme d'Internet mise en place pour le World Wide Web (voir RFC 3986).</p> <p>Le sigle URI est généralement utilisé pour désigner une telle chaîne de caractères. Par exemple <code>urn:ietf:rfc:2396</code> est un URI identifiant le RFC 2396. Les URI sont la technologie de base du World Wide Web car tous les hyperliens du Web sont exprimés sous forme d'URI. Un URI peut être de type "locator" ou "name" ou les deux.</p> <p>Un Uniform Resource Locator (voir « URL ») est un URI qui, outre qu'il identifie une ressource Web, fournit les moyens d'agir sur une ressource ou d'obtenir une représentation de la ressource en décrivant son mode d'accès primaire ou "emplacement" réseau. Par exemple, l'URL <code>http://www.wikipedia.org/</code> est un URI qui identifie une ressource (page d'accueil Wikipedia) et implique qu'une représentation de cette ressource (comme le code usuel HTML de la page d'accueil, comme caractères encodés) peut être obtenu via HTTP d'un réseau hôte appelé <code>www.wikipedia.org</code>.</p> <p>Un Uniform Resource Name (voir « URN ») est un URI qui identifie une ressource par son nom dans un espace de noms. Un URN peut être employé pour parler d'une ressource sans que cela préjuge de son emplacement ou de la manière de la référencer. Par exemple, l'URN <code>urn:isbn:0-395-36341-1</code> est un URI qui, comme un International Standard Book Number (ISBN), autorise quelqu'un à faire référence à un livre, mais ne suggère où et comment en obtenir une copie réelle. [Wikipedia]</p>
URL	<p>Uniform Resource Locator. L'URL est l'adresse virtuelle d'une ressource sur le Web, exprimé par une séquence de caractères, conformément à un format standard. L'URL a été une innovation fondamentale dans l'histoire de l'Internet. Il a été créé par Tim Berners-Lee en 1991 pour permettre aux auteurs de documents d'établir des hyperliens sur le World Wide Web.</p>
URN	<p>Uniform Resource Name, traduit littéralement de l'anglais par « nom uniforme de ressource », est le nom d'un standard informatique dans le domaine de l'Internet qui concerne principalement le World Wide Web.</p> <p>Le document de base est le RFC 2141 URN Syntax publié en 1997 par l'Internet Engineering Task Force. Il donne une syntaxe de chaîne de caractères utilisable pour identifier une ressource (un document, une image, un enregistrement sonore, etc.) globalement, durant toute son existence, indépendamment de sa localisation ou de son accessibilité par Internet. Les Uniform Resource Names sont des Uniform Resource Identifiers dont la</p>

	<p>méthode est urn.</p> <p>Le sigle URN est généralement utilisé, en français comme en anglais, pour désigner une telle chaîne de caractères. Par exemple urn:ietf:rfc:2141 est un URN identifiant le RFC 2141. Les URN sont des Uniform Resource Identifiers (URI) et en respectent donc les règles syntaxiques. Les URN ont la syntaxe suivante :</p> <p>urn:NID:NSS</p> <ul style="list-style-type: none"> • urn est la méthode d'URI des URN. • NID (Namespace Identifier) est un identificateur d'espace de nommage. • NSS (Namespace specific String) est la partie spécifique à l'espace de nommage identifié par le NID. L'interprétation syntaxique de cette partie dépend de l'espace de nommage. <p>L'usage de minuscules ou de majuscules ne fait pas de différence pour l'écriture de la méthode urn ni pour le NID. Il peut en revanche faire une différence pour le NSS. [Wikipedia]</p>
UTF-8	<p><i>L'UCS transformation format 8 bits</i> est un format de codage de caractères défini pour les caractères Unicode (UCS). Chaque caractère est codé sur une suite d'un à quatre octets. UTF-8 a été conçu pour être compatible avec certains logiciels originellement prévus pour traiter des caractères d'un seul octet. UTF-8 est standardisé dans la RFC 3629 (UTF-8, a transformation format of ISO 10646). [Wikipedia]</p>
WAV	<p>WAV (ou WAVE), une contraction de <i>WAVEform audio format</i>, est un standard pour stocker l'audio numérique de Microsoft et IBM. C'est le format le plus courant pour l'audio non compressé sur les plates-formes de Microsoft, mais il est bien courant sur les systèmes GNU/Linux aussi.</p> <p>Le format WAV ne correspond à aucun format d'encodage spécifique, il s'agit d'un conteneur capable de recevoir des formats aussi variés que le MP3, le WMA, l'ATRAC3, l'ADPCM, le PCM. C'est ce dernier qui est cependant le plus courant, et c'est pour cela que l'extension .wav est souvent - et donc à tort - considérée comme correspondant à des fichiers "sans pertes" (communément désignés par le mot anglais lossless). Le format WAV est standardisé sous Windows ; son pendant sous plate-forme Macintosh est l'AIFF/AIFC. [Wikipedia]</p>
WordPerfect	<p>Un format de traitement de texte du logiciel WordPerfect de Corel.</p>

XACML	L'eXtensible Access Control Markup Language est un XML-Schema qui standardise la représentation et le traitement de politiques d'autorisation. Ce standard, établi par le consortium OASIS sert à émettre des règles par l'application desquelles on peut diriger l'accès d'un sujet aux ressources d'un système. [Wikipedia]
XLS	.xls est un suffixe de fichier tableur. Ces fichiers peuvent être ouvert avec des logiciels tableurs comme Microsoft Excel ou OpenOffice.org calc. [Wikipedia]
XML	<p>XML (Extensible Markup Language ou langage de balisage extensible) est un standard du World Wide Web Consortium qui sert de base pour créer des langages de balisage spécialisés: c'est un « métalangage ». En ce sens, XML permet de définir un vocabulaire et une grammaire associée sur base de règles formalisées. L'objectif initial de XML était de faciliter le partage de textes et d'informations structurées, par exemple au travers de l'Internet, en séparant le contenu (les données) du contenant (la présentation des données).</p> <p>Les dialectes XML (XSLT, XML Schema, XHTML, RDF/XML, SOAP, SMIL, MathML, SVG) sont décrits de façon formelle : une structure de données simple est définie avec une DTD (Document Type Definition), une structure de données détaillée est définie avec un XML Schema ou tout autre DSDL (Document Schema Definition Languages, c'est-à-dire langage de description de schéma). [vu le 14.2.2006]</p>
XSLT	<i>L'eXtensible Stylesheet Language Transformation</i> est un logiciel qui convertit un document XML en un autre format comme HTML, PDF ou texte. Il peut également être utilisé pour convertir un document XML en un document XML possédant un ensemble différent de balises XML (un schema différent).
Z39.87	NISO Z39.87 est un standard qui définit un ensemble d'éléments de métadonnées pour les images numériques rastérisées, pour permettre aux usagers de développer, échanger et interpréter des fichiers d'images numériques. Le dictionnaire a été conçu pour faciliter l'intéropérabilité entre les systèmes, les services et les logiciels ainsi que pour soutenir la gestion et l'accès à long terme aux collection d'images numériques. [NISO]

20.4 Contacts

Nom, prénom	E-mail	Fonction (dans ce projet)	Téléphone	Bâtiment	Service
Beetschen, Floriane	Floriane.Beetschen@unil.ch	Unisciences - public scientifique prof	692 22 72	Amphimax	UNIL/Unicom
Bonetti, Josiane	Josiane.Bonetti@unil.ch	Bibliothèque resp. Biologie	692 41 04	Biophore	UNIL/BIB. Biologie
Borgeaud, Claude	Claude.Borgeaud@bcu.unil.ch	Thèses électroniques	692 48 12	Unithèque	BCU/Dorigny
Bourquin, Yvan	Yvan.Bourquin@bcu.unil.ch	Bibliothécaire resp. Théologie	692 48 31	Unithèque	BCU/Dorigny
Clémence, Alain	Alain.Clemence@unil.ch	Décanat SSP, ISSP	692 32 40	Anthropole	UNIL/ISSP
Corsini, Silvio	Silvio.Corsini@bcu.unil.ch	Resp. projet numérisation BCU	692 47 85	Unithèque	BCU/Dorigny
Derron, Pascale	Pascale.Derron@bcu.unil.ch	Bibliothécaire resp. Lettres	692 48 41	Unithèque	BCU/Dorigny
Dubois, Laurent	Laurent.Dubois@bcu.unil.ch	Photographie	692 47 74	Unithèque	BCU/Dorigny
Frey, Jeannette	Jeannette.Frey@bcu.unil.ch	chefe de projet SERVAL	692 48 28	Unithèque	BCU/Dorigny
Furst, Suzanne	Suzanne.Furst@unil.ch	Fonds des thèses UNIL	692 20 30	Unicentre	UNIL/Aff. étudiantes
Gardel, Philippe	Philippe.Gardel@unil.ch	Ci, Architecture SERVAL	692 22 34	Amphimax	UNIL/Ci
Gavillet, Gérald	Gerald.Gavillet@bcu.unil.ch	Bibliothécaire resp. HEC et Droit	692 48 80	Internef	BCU/BDSE
Goudet, Jérôme	Jerome.Goudet@unil.ch	DEE	692 42 42	Biophore	UNIL/DEE
Gozzo, Tessa	Tessa.Gozzo@unil.ch	Inventaire des fonds patrimoniaux UNIL	692 20 82	Unicentre	UNIL/soutien enseignement
Grosse, Jérôme	Jerome.Grosse@unil.ch	Responsable de la communication	692 22 71	Amphimax	UNIL/Unicom
Guélat, Jacques	Jacques.Guelat@unil.ch	Responsable conseils et études	692 22 31	Amphimax	UNIL/Ci
Hajjaoui, Akram	Akram.Hajjaoui@unil.ch	Responsable groupe de développements	692 22 53	Amphimax	UNIL/Ci
Henchoz, Daniel	Daniel.Henchoz@unil.ch	Responsable serveurs SERVAL	692 22 20	Amphimax	UNIL/Ci
Horisberger, Jean-Daniel	Jean-Daniel.Horisberger@unil.ch	DPT	692 53 62	Bugnon 27	UNIL/DPT
Iriarte, Pablo	Pablo.Iriarte@chuv.ch	Informatique bibliothécaire CHUV	314 73 98	CHUV	CHUV/BiUM
Jaccard, Martine	Martine.Jaccard@bcu.unil.ch	Bibliothécaire resp. Géosciences	692 48 39	Unithèque	BCU/Dorigny
Jacot-Guillarmod, Pascal	Pascal.Jacot-Guillarmod@unil.ch	Directeur du Ci	692 22 01	Amphimax	UNIL/Ci
de Kaenel, Isabelle	Isabelle.De-Kaenel@chuv.ch	Directrice de la BiUM	314 50 82	CHUV	CHUV/BiUM
Keller, Pierre	Pierre.Keller@bcu.unil.ch	Chef de projet informatique BCU	692 48 13	Unithèque	BCU/Dorigny
Meyer, Yannick	Yannick.Meyer@unil.ch	chef de projet Unimédia	692 22 73	Amphimax	UNIL/Unicom
Mincio, Danielle	Danielle.Mincio@bcu.unil.ch	Responsable PAC à la BCU/Dorigny	692 47 83	Unithèque	BCU/Dorigny
Plaza, Patricia	Patricia.Plaza@unil.ch	Cheffe de projet GED	692 20 17	Unicentre	UNIL/archives
Pochon, Mireille	Mireille.Pochon@bcu.unil.ch	e-journals, BCU/Dorigny	692 48 29	Unithèque	BCU/Dorigny
Robert, Olivier	Olivier.Robert@unil.ch	Chef de services	692 20 27	Unicentre	UNIL/archives
Rusconi, Federica	Federica.Rusconi@bcu.unil.ch	Resp. collection de musique	316 78 64	Palais de Rumine	BCU/Riponne
Sahy, Pascale	Pascale.Sahy@bcu.unil.ch	Resp. médiathèque BCU/Dorigny	692 47 69	Unithèque	BCU/Dorigny
Schranz, Anne	Anne.Schranz@bcu.unil.ch	Resp. dépôt légal vaudois BCU/Riponne	316 78 54	Palais de Rumine	BCU/Riponne

Simioni, Olivier	Olivier.Simioni@bcu.unil.ch	Resp. e-books BCU	692 48 09	Unitèque	BCU/Dorigny
Spini, Dario	Dario.Spini@unil.ch	ITB	692 38 44	Provence	UNIL/ITB
Strahm, Jean-Jacques	Jean-Jacques.Strahm@unil.ch	bases de données images UNIL	692 22 84	Internef	UNIL/bur 246
Tulipani, Corinne	Corinne.Tulipani@unil.ch	Fonds des publications UNIL	692 20 50	Unicentre	UNIL/Aff. Générales
Urcelay, Leire	Leire.Urcelay@unil.ch	Informaticienne spécialiste SERVAL	692 22 40	Amphimax	UNIL/Ci
Villard, Hubert	Hubert.Villard@bcu.unil.ch	Directeur de la BCU	692 48 00	Unicentre	BCU/Dorigny