

# Etude de concept élaboration d'un modèle pour un archivage à long terme centralisé des données scientifiques primaires et secondaires en Suisse

situation, besoins, desiderata, modèles,  
protagonistes, conditions cadre et contexte européen

**Report****Author(s):**

Keller-Marxer, Peter

**Publication date:**

2008

**Permanent link:**

<https://doi.org/10.3929/ethz-a-006070907>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

Etude de concept

# **Elaboration d'un modèle pour un archivage à long terme centralisé des données scientifiques primaires et secondaires en Suisse**

Situation, besoins, desiderata, modèles, protagonistes,  
conditions cadre et contexte européen

Rapport établi à la demande de



ETH-Bibliothek, Ecole poly-  
technique fédérale, Zurich

**E-lib.ch**

Elektronische Bibliothek Schweiz  
Bibliothèque électronique suisse  
Swiss electronic library

Projet Bibliothèque électronique suisse,  
Conférence universitaire suisse

DOCUMENT INTERNE

080401-01– Version 1.4 – 2008-12-22 –(version finale)

© 2008 ikeep Ltd.

## Auteurs

Peter Keller-Marxer, Dr. phil. nat.  
Niklaus Bütikofer, lic. phil.

ikeep AG  
Morgenstrasse 129  
CH-3018 Bern  
+41 31 998 42 80  
peter.keller@ikeep.com

## Versions du document :

Version	Date	Etat	Auteurs	Remarques
1.0	2008-10-01	Esquisse interne	PK,NB	Révision en interne
1.1	2008-10-25	Esquisse interne	PK,NB	Version augmentés après révision en interne
1.2	2008-11-06	Esquisse	PK,NB	Version proposée au client pour révision
1.3	2008-12-09	Version finale de l'esquisse	PK,NB	Adaptations après révision par le client au 2.12.08
1.4	2008-12-22	Version finale	PK,NB	Petites corrections demandées par le client, 22.12.08

Traduction française de l'étude : *Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitaler Primär- und Sekundärdaten der Forschung für die Schweiz. Situation, Bedarf, Nutzenziele, Modelle, Anspruchsgruppen, Rahmenbedingungen und europäisches Umfeld.* ikeep Ltd. 2008

NB: En cas d'ambiguïté, c'est la version originale de cette étude, rédigée en allemand, qui fait foi.

Traduction : Hubert Villard, Belmont/Lausanne (été 2009)

# Table des matières

<b>1</b>	<b>Résumé</b>	<b>1</b>
<b>2</b>	<b>Introduction, tour d'horizon</b>	<b>4</b>
2.1	Quelques concepts	5
2.2	Situation initiale	6
2.2.1	Utilisation secondaire	6
2.2.2	Intégrité de la recherche	7
2.2.3	Lois sur les archives et obligation de proposer	8
2.3	Thématique centrale du rapport	9
2.4	Modèles	11
2.5	Structure du rapport	13
<b>3</b>	<b>Situation, besoins et périmètre</b>	<b>14</b>
3.1	Données scientifiques primaires	14
3.1.1	Brève description	14
3.1.2	Utilisation possible des données conservées ou archivées	15
3.1.3	Besoins et degré d'utilisation	16
3.1.3.1	Directives sur l'intégrité de l'ESF et de l'OCDE	17
3.1.3.2	Dispositions concernant la conservation en Suisse et en Allemagne	17
3.1.3.3	Identification des données issues de la recherche	19
3.1.3.4	Conservation en vue d'une exploitation secondaire	20
3.1.4	Etat des mesures d'archivage	21
3.2	Les publications électroniques	24
3.2.1	Brève description	24
3.2.2	Utilisation possible des données conservées ou archivées	24
3.2.3	Besoins et taux d'utilisation	25
3.2.4	Etat des mesures d'archivage	25
3.3	Supports numériques pour l'enseignement	26
3.3.1	Brève description	26
3.3.2	Utilisation possible des données conservées ou archivées	26
3.3.3	Besoins et degré d'utilisation	26
3.3.4	Etat des mesures d'archivage	27
3.4	Documents rétronumérisés	27
3.4.1	Brève description	27
3.4.2	Utilisation possible des données conservées ou archivées	28
3.4.3	Besoins et degré d'utilisation	28
3.4.4	Etat des mesures d'archivage	28
<b>4</b>	<b>Protagonistes</b>	<b>29</b>

4.1	Cadre juridique .....	33
4.1.1	La Loi fédérale sur l'archivage.....	34
4.1.2	Le droit cantonal.....	36
4.1.3	La législation sur les bibliothèques.....	36
<b>5</b>	<b>Modèles pour les données primaires .....</b>	<b>37</b>
5.1	Modèle «Conservation avec autocontrôle » .....	37
5.1.1	Objectifs stratégiques.....	38
5.1.2	Exposé des motifs .....	39
5.1.3	Objectifs opérationnels.....	40
5.1.3.1	Coûts et financement .....	44
5.1.4	Explications .....	45
5.1.5	Elargissement vers OAIS des objectifs opérationnels .....	47
5.1.5.1	Coûts et financement .....	48
5.1.6	Vue d'ensemble .....	48
5.2	Modèle «Archivage avec obligation de dépôt» .....	53
5.2.1	Objectifs stratégiques.....	55
5.2.2	Justification.....	56
5.2.3	Objectifs opérationnels.....	56
5.2.3.1	Coûts et financement .....	59
<b>6</b>	<b>Modèle pour les données secondaires.....</b>	<b>61</b>
6.1	Objectifs stratégiques.....	61
6.2	Motivation .....	62
6.3	Objectifs opérationnels pour <i>eJournals</i> et <i>eBooks</i> .....	65
6.3.1	Coûts et financement.....	68
6.4	Objectifs opérationnels pour d'autres types de données .....	69
6.4.1	Bases de données scientifiques.....	69
6.4.2	Objets de <i>eLearning</i> .....	70
6.4.3	Copies <i>master</i> de rétronumérisations .....	71
<b>7</b>	<b>Modèles alternatifs.....</b>	<b>74</b>
7.1	Comité stratégique / Task Force nationale .....	74
7.2	Centre national de compétences.....	75
7.3	Réseau national de compétences .....	75
7.4	Répertoire, inventaire et portail.....	76
<b>8</b>	<b>Environnement européen .....</b>	<b>77</b>
8.1	Les projets EU dans le cadre des FP6 et FP7 .....	78
8.2	Serveurs institutionnels .....	79
8.3	Les bibliothèques nationales de dépôt.....	80
8.4	L'Alliance for Permanent Access to the Records of Science .....	81
8.4.1	PARSE.insight.....	83
8.5	L'Allianz-Initiative Digitale Information en Allemagne .....	83
8.6	Autres activités du même ordre.....	85

8.6.1	Stratégies, plans d'action et infrastructures nationales .....	86
8.6.1.1	Sur le plan international .....	86
8.6.1.2	Sur le plan national .....	86
8.6.2	Conservation des données primaires propres à certaines disciplines .....	90
8.6.2.1	Sciences sociales .....	90
8.6.2.2	Sciences naturelles .....	92
8.6.3	Archivage de publications électroniques .....	93
8.6.3.1	Pays-Bas : e-Depot Koninklijke Bibliotheek .....	93
8.6.3.2	Allemagne : Kopal .....	94
8.6.3.3	Suisse : e-Helvetica .....	94
8.6.3.4	France : Hyper Article en Ligne (HAL) .....	94
8.6.3.5	Sur le plan international : Portico .....	95
<b>9</b>	<b>Bibliographie .....</b>	<b>96</b>



# 1 Résumé

Une part prépondérante de la recherche scientifique actuelle s'appuie sur des données électroniques : les données sous forme numérique provenant de mesures, d'analyses et de simulations (données primaires), une fois saisies, documentées et stockées, constituent la matière première des dépouillements et des résultats d'expériences qui, à leur tour, sont eux aussi rédigés et publiés de plus en plus souvent sous forme électronique (données secondaires) pour venir irriguer de nouveaux projets de recherche (utilisation secondaire).

Nombre de publications scientifiques et de supports d'enseignement ne sont plus publiés ou utilisés que sous forme électronique (*eJournals*, *eBooks*, bases de données). Ceci rend nécessaire la mise en place de stratégies, de méthodes et de solutions techniques qui garantissent, pour la recherche future, la disponibilité de ces données secondaires une fois que les licences d'utilisation seront échues ou que les éditeurs auront supprimé les accès en ligne ou cessé leur activité.

La manière de considérer les données primaires a changé : au terme du projet de recherche, elles ne sont plus vues comme un artefact obsolète, mais comme la base même qui permet de vérifier l'intégrité des résultats obtenus et qui peut ensuite alimenter une utilisation secondaire ultérieure.

En Suisse, un bon nombre d'instituts de recherche ainsi que tous les organes d'encouragement ont publié des prescriptions – fondées sur le nouvel article 11a de la Loi sur la recherche valable depuis mars 2008 – qui exigent notamment une conservation sécurisée et documentée des données primaires bien après l'aboutissement du projet de recherche. En Suisse, le délai de conservation doit être «adéquat pour le domaine considéré», alors que l'Allemagne et d'autres pays prescrivent un délai explicite d'au moins dix ans.

Mais la responsabilité de cette conservation sécurisée n'incombe aujourd'hui qu'aux chercheurs eux-mêmes, qui généralement, en raison de la forte mobilité que connaît leur secteur d'activité, ne disposent pas des moyens techniques et organisationnels nécessaires pour assumer une telle responsabilité sur le long terme.

L'initiative européenne «Alliance for Permanent Access to the Records of Science in Europe» et l'«Allianz-Initiative ,Digitale Information'» lancée par neuf sociétés de recherche allemandes répondent à un urgent besoin de traiter de la conservation des données primaires afin d'assurer l'intégrité de la recherche et de garantir leur réexploitation ultérieure.

La Conférence universitaire suisse, la Conférence des recteurs des universités suisses, le Conseil des écoles polytechniques ainsi que l'Office fédéral de la formation et de la technologie ont inscrit la thématique de l'archivage à long terme des données primaires et secondaires dans un sous-projet du projet national «E-lib.ch» (Bibliothèque électronique suisse). La Conférence des



bibliothèques universitaires a chargé la Bibliothèque de l'ETH de développer des modèles pour une prestation nationale et centralisée dans ce domaine.

Sur mandat de la Bibliothèque de l'ETH, nous en avons examiné les conditions cadre, les besoins et les bénéfices potentiels et, après évaluation de projets et de solutions opérationnelles en Europe, nous avons esquissé deux modèles complémentaires pour les données primaires et un pour les données secondaires.

Le modèle proposé pour les données secondaires reprend pour l'essentiel la stratégie en vigueur en Allemagne, mais va plus loin en évaluant des synergies avec des partenaires potentiels dans le domaine de l'archivage de supports de *eLearning* et de copies «master» d'objets numériques.

Le premier modèle pour les données primaires va dans le sens d'une approche de type «self deposit» et met l'accent de manière pragmatique sur les besoins relevés dans l'activité de recherche relativement aux nouvelles directives sur l'intégrité de la science. Il ne prévoit aucune réglementation, mais compte sur la responsabilité personnelle et sur l'initiative spontanée des chercheurs et des institutions de recherche. C'est à eux qu'il appartient de définir et de mettre en œuvre, pour les données primaires qu'ils déposent dans l'archive, les niveaux minimaux d'exigence en termes de délai de conservation, de volume, de documentation et de format, ceci conformément aux prescriptions d'intégrité valables dans chaque cas et en fonction de la valeur que les données peuvent revêtir pour une utilisation secondaire future. Le modèle prévoit également une assistance en matière de conseils au travers d'un réseau de compétences.

Le modèle offre à la fois aux chercheurs et aux institutions de recherche, à côté d'une conservation intégrale et sécurisée des données, des possibilités d'exploitation supplémentaires comme le fait de définir soi-même les autorisations d'accès aux données déposées et de les rendre librement disponibles pour la collectivité («Open Access»), la constitution de catalogues thématiques ou institutionnels de ces données ainsi que la possibilité de les citer. Ce dernier point mérite attention car les données primaires ne sont généralement pas publiées, ce qui fait que, surtout dans le cas d'une utilisation secondaire, les chercheurs n'ont aujourd'hui aucune possibilité de faire référence dans leurs publications de manière simple, univoque et compréhensible à tel corpus de données utilisé.

Le second modèle, complémentaire au premier, propose une approche réglementée en partant du constat que l'archivage de données primaires qui revêtent une valeur durable n'est à ce jour pas pris en compte par les législations gouvernementales sur l'archivage (Confédération et cantons). Il propose une démarche systématique, mais néanmoins pragmatique, pour évaluer le degré d'intérêt que présentent des données primaires pour un archivage et définit, pour les institutions de recherche, les conditions préalables qui leur permettent d'assumer leur devoir dans le cadre de la loi sur les archives. Ce modèle est aussi en adéquation avec les attentes de la collectivité qui veut que les résultats de la recherche restent compréhensibles sur le long terme pour des domaines éthiquement ou politiquement délicats, lorsque ces résultats peuvent avoir d'importantes répercussions sur l'homme et son environnement ou lorsqu'ils sont susceptibles d'influencer des décisions politiques ou l'action du législateur.

Puis nous esquissons quatre modèles alternatifs plus légers où la centrale nationale n'archive elle-même aucune donnée, mais assume une fonction stratégique de coordination et de conseil et anime un réseau national de compétences „archivage pérenne“ sur les modèles allemand et français.

L'engagement marqué des partenaires du projet «E-lib.ch» résulte de leur conviction que la conservation à long terme des informations numériques scientifiques et leur accessibilité garantie sur la durée sont les conditions essentielles d'un approvisionnement satisfaisant en matière de documentation scientifique pour les hautes écoles suisses au fil des générations.

Il concorde également avec les recommandations de l'OCDE de 2007 sur le traitement des données issues de la recherche financée par les fonds publics : les stratégies, les projets et les centres de prestation voués à la conservation et à l'utilisation secondaire des données primaires doivent être conçus, planifiés et implémentés comme des composants essentiels de l'infrastructure scientifique d'un pays.

## 2 Introduction, tour d'horizon

Le projet dans le cadre duquel la présente étude de concept a été produite est en fait un sous-projet du projet national «E-lib.ch – Elektronische Bibliothek Schweiz»<sup>1</sup>. «E-lib.ch» est soutenu par la Conférence universitaire suisse pour la période de subventionnement 2008 – 2011, ce qui permet aux institutions du domaine des Ecoles polytechniques de s'y associer via le Conseil des Hautes Ecoles, et à celles des hautes écoles spécialisées via l'Office fédéral de la formation et de la technologie (OFFT). Plusieurs sous-projets proposés par de nombreux partenaires sont en phase de réalisation dans le cadre de E-lib.ch, et ont pour but de promouvoir un accès à la fois simple et rapide à des ressources documentaires de toutes natures. La direction et la coordination du projet E-lib.ch sont rattachées à la Bibliothèque de l'ETH, et la Conférence des bibliothèques universitaires suisses (CBU) agit en tant que superviseur.

C'est la CBU qui est le mandant du présent sous-projet intitulé «Elaboration conceptuelle d'un modèle pour un archivage pérenne centralisé des données numériques primaires et secondaires pour la Suisse ». L'organe exécutif (le maître d'ouvrage) en est la Bibliothèque de l'ETH. Le mandat de la direction de la Bibliothèque de l'ETH, qui répond à une volonté de financement par la CBU, énumère les objectifs suivants [1] (en bref) :

*«L'objectif du projet est le développement d'un concept avec un ou plusieurs modèles, voire des variantes de modèles, pour l'archivage électronique fiable à long terme des données numériques primaires et secondaires des hautes écoles suisses, des centres de recherche et des bibliothèques universitaires suisses.»*

*«Le résultat du projet devra aboutir à un concept comprenant un ou plusieurs modèles, voire des variantes de modèles, pour l'archivage numérique central à long terme de données primaires et secondaires pour les hautes écoles suisses, les centres de recherche et les bibliothèques universitaires suisses. Ce concept devra par exemple prévoir des propositions pour l'élaboration de directives et de standards nationaux dans le domaine de l'archivage numérique à long terme, pour la création d'un service suisse d'archivage numérique pérenne, et pour la création d'une archive fiable et durable sur le plan national.»*

Nous avons été mandatés, dans le périmètre ainsi défini du projet, pour analyser et mettre en exergue les conditions limites du point de vue du droit et des juridictions, repérer les groupes d'intérêt potentiels, évaluer les besoins en termes d'archivage pérenne pour différentes catégories de supports, présenter l'état de l'art sur les plans européen et international, et enfin proposer quelques modèles qui soient en conformité avec le souhait de parvenir à une solution centralisée et réellement nationale, tout en restant réaliste. Précisons que cette étude ne prévoit pas la réalisation de prototypes d'expérimentation („testbeds“).

---

<sup>1</sup> <http://www.e-lib.ch>

## 2.1 Quelques concepts

On entend par « données primaires » les données originales, mais déjà épurées (résultats de mesures, d'enquête ou de simulations), qui serviront de base pour des dépouillements ultérieurs et permettront d'étayer des hypothèses de recherche. Les résultats des dépouillements en revanche, ainsi que les résultats des travaux de recherche, ne sont pas considérés comme des données primaires, car ils peuvent être reproduits à partir de ces mêmes données primaires pour autant que le protocole d'analyse soit connu.

On entend aussi par « données primaires » toutes sortes d'informations auxiliaires (métadonnées et documentation) qui font que les données de fond restent intelligibles et compréhensibles pour des tiers sur la durée (par ex. les résultats de mesures ou de relevés, les unités de mesure, les catégories de données, les étalonnages d'instruments de mesure, le contexte même de la mesure ou des relevés, le journal d'expérimentation, etc.) Mais on ne comprend pas sous ce terme tous les autres documents qui décrivent la procédure de mesure et de recherche et la rendent par là compréhensible. Pour tout ce matériau, on parlera plutôt ici de « **documentation de recherche** ».

On entend par « données secondaires » les éditions électroniques des revues scientifiques (*eJournals*) et des livres (*eBooks*) prises sous licence par les bibliothèques, les supports d'enseignement développés ou acquis par les universités (*eLearning*) ainsi que les publications sous forme électronique produites par les hautes écoles (thèses, travaux de diplôme, *preprints*, etc.)

Nous faisons une distinction entre « conservation » et « archivage » :

- On entend par **conservation** le fait de garder pour un temps des informations sous quelque forme que ce soit, et pour lesquelles les critères d'exigence en matière de format de données, de documentation et de classement sont fixés de cas en cas, soit par ceux qui les auront produites, soit au travers de recommandations non contraignantes propres à chaque domaine. La conservation peut découler d'une volonté spontanée ou être prescrite par des lois<sup>2</sup> ou des réglementations supranationales.
- **L'archivage**<sup>3</sup> consiste à conserver des données dans une forme adéquate ; dans ce cas, il existe des prescriptions contraignantes et valables pour tous les cas au sujet de la sélection des données à archiver, de leur documentation, de leur classement et des techniques utilisées pour leur traitement. L'archivage peut être un acte spontané ou imposé par la législation<sup>4</sup> ou par des réglementations supranationales.
- **Un archivage conforme aux lois gouvernementales sur l'archivage** voit ses modalités et son périmètre prescrits par les dispositions de ladite loi et par les directives des archives gouvernementales. Il se limite à la documentation qui, d'un commun accord entre

---

<sup>2</sup> De nombreuses législations prescrivent un délai de conservation de 1 à 20 ans. C'est le cas par exemple du Code des obligations.

<sup>3</sup> En informatique, le terme d'archivage signifie depuis plus de 50 ans „stocker et conserver en sécurité sur une mémoire morte“. Cette acception n'est pas considérée ici.

<sup>4</sup> Plusieurs législations et réglementations prescrivent un délai d'archivage de 1 à 30 ans sous une forme et un classement bien définis. Ainsi par exemple la loi sur la TVA ou les législations cantonales sur le Registre des hypothèques.

les archivistes et les producteurs d'information, est considérée comme digne d'être archivée pour le long terme.

Tant la conservation que l'archivage s'effectuent généralement en fonction d'objectifs bien définis et en supputant les besoins des futurs utilisateurs potentiels.

La conservation et l'archivage peuvent être limités dans le temps ou pas. Si le délai dépasse dix ans pour des données numériques, on parle alors d'archivage pérenne, car ce laps de temps est déjà supérieur à la durée de vie des systèmes de stockage ou de traitement des données actuels. En Suisse, l'archivage tel que prescrit par la loi a un caractère pérenne.

Nous nous limiterons dans cette étude à exposer quatre motivations qui justifient une conservation ou un archivage des données scientifiques primaires et secondaires :

- **« Intégrité de la recherche »** : conservation limitée (5 – 10 ans) des données primaires afin d'être en accord avec les lignes directrices sur l'«intégrité de la recherche» et sur la «bonne pratique scientifique».
- **Réutilisation** : conservation, souvent limitée dans le temps, des données primaires significatives pour une utilisation ultérieure.
- **Sécurité** : conservation à long terme ou archivage des publications scientifiques afin de garantir l'accès à leur version numérique après échéance de la licence ou en cas de disparition de l'éditeur.
- **Réglementations sur l'archivage** : archivage de données primaires et secondaires par les archives d'Etat conformément aux réglementations en vigueur.

Seules les deuxième et troisième raisons sont explicitement évoquées dans les objectifs de ce sous-projet de E-lib.ch. Fondamentalement, l'obligation d'archiver lors de chaque étape de l'activité de recherche vise, pour un public déterminé, à

- identifier, sélectionner et placer dans l'archive toute documentation qui a une valeur durable,
- à la classer et à la cataloguer selon des critères cohérents et systématiques,
- à la rendre accessible, lisible et utilisable dans sa forme authentique
- et à en préserver l'intelligibilité relativement à son contexte d'origine.

Les buts de la conservation ont généralement moins de portée.

## 2.2 Situation initiale

### 2.2.1 Utilisation secondaire

Les objectifs de ce sous-projet d'E-lib.ch découlent de la nécessité *de principe* qu'il y a pour la recherche de conserver durablement les données numériques primaires et d'archiver de manière pérenne les données secondaires [1]:

*«La conservation à long terme des informations numériques scientifiques (primaires et secondaires), ainsi que leur accessibilité garantie sur la durée, sont les conditions essentielles d'un approvi-*

*sionnement satisfaisant en documentation scientifique pour les hautes écoles suisses au fil des générations»*

*«Les données numériques occupent une place prépondérante dans les travaux scientifiques contemporains : sans le recours à des moyens techniques plus ou moins intensifs, le dépouillement de données numériques issues de mesures ou de relevés quantitatifs (données primaires) ne pourrait plus être maîtrisé. Les résultats de ces travaux sont à leur tour rédigés et publiés sous forme électronique (données secondaires).»*

Si les données primaires (résultats de mesures, d'enquêtes ou de simulations) sont depuis longtemps traitées quasi exclusivement sous une forme électronique, les résultats eux-mêmes de la recherche sont de plus en plus fréquemment publiés dans des revues scientifiques numériques (*eJournals*), soit en complément aux versions imprimées, soit exclusivement sous forme électronique. S'ajoutent à ces données secondaires de plus en plus d'ouvrages scientifiques (*eBooks*) ainsi que des supports d'enseignement électroniques (*eLearning*) pour lesquels il n'existe plus de version imprimée.

Ainsi on appellera *utilisation primaire* l'exploitation des données primaires, *réutilisation* leur exploitation ultérieure par ceux qui les ont produites, et *utilisation secondaire* leur exploitation ultérieure par des tiers. Pour les données secondaires, la publication des résultats de la recherche représente l'utilisation primaire (diffusion des résultats dans le monde scientifique) alors que l'utilisation de la publication comme source de connaissance ou de citation représente l'utilisation secondaire. Tant les informations primaires que les informations secondaires peuvent, à leur tour, être injectées dans de nouvelles activités de recherche.

Cela entraîne, pour les données primaires, que leur conservation ne doit pas s'arrêter au terme du projet de recherche, mais qu'elles doivent rester disponibles bien au-delà. Cela implique également qu'elles doivent être conservées sous une forme qui en permette la compréhension et l'interprétation lors d'une utilisation secondaire.

Pour les données secondaires, cela a pour conséquence que les publications numériques doivent rester disponibles même après l'échéance de la licence d'utilisation ou la disparition de l'éditeur. Cela implique également qu'elles soient archivées sous une forme qui en permette l'exploitation et la compréhension sur la durée. Cela peut aller jusqu'à la conversion périodique dans de nouveaux formats lorsque les solutions techniques retenues pour l'archivage deviennent obsolètes.

### 2.2.2 Intégrité de la recherche

Par la publication des résultats de la recherche dans leurs revues, les éditeurs – notamment leurs «editorial boards» et leurs cercles de «peer review» - font de gros efforts pour garantir l'authenticité, la qualité, la visibilité et une diffusion efficace des contenus publiés. Mais parallèlement les tentatives de tricherie, de manipulation ou de falsification des données augmentent, imputables très certainement à la pression croissante à laquelle sont soumis les chercheurs de publier pour assurer leur succès.

C'est pourquoi, durant les trois dernières années, des organisations de recherche et des associations du monde entier ont édicté des recommandations et directives en faveur de l'„intégrité dans la recherche scientifique” et de la „bonne pratique scientifique”. Parmi elles, l'OCDE et la European Science Foundation qui exigent que les données primaires qui étayent les résultats

d'une recherche soient conservées suffisamment longtemps pour que l'exactitude et la fiabilité des résultats puissent être contrôlées après leur publication, que cela soit par d'autres chercheurs ou (en cas de soupçon de fraude) par les instances universitaires.

En Suisse, en 2007 et 2008, les organes d'encouragement de la recherche – le Fonds national et les quatre académies scientifiques – ont publié de nouvelles directives sur la „bonne pratique scientifique“, que les bénéficiaires de subventions doivent s'engager à respecter conformément au nouvel article 11a de la Loi sur la recherche applicable depuis mars 2008.

Ces directives des organes d'encouragement impliquent aussi que les données primaires soient conservées de manière sécurisée et dans leur forme originale pour un laps de temps adéquatement prolongé après la fin du projet. Négliger cette obligation représente une infraction à la „bonne pratique scientifique“ et est en principe déjà considéré comme un indice de fraude lors des enquêtes consécutives à des soupçons de tricherie. De nombreuses universités et hautes écoles ont déjà suivi l'exemple des organes de subventionnement et édicté leurs propres règlements en la matière.

Le devoir généralisé de conservation, même si le délai n'est pas clairement défini en Suisse – il est par exemple de dix ans en Allemagne – crée une situation inédite pour les chercheurs du pays dans la mesure où c'est directement à eux, et non à leur institution, que les directives délèguent la responsabilité de la conservation des données primaires. En raison de la volatilité des projets de recherche et de la grande mobilité des chercheurs eux-mêmes, le devoir qui leur est imposé de conserver pour le long terme des données émanant de nombreux projets tient de la gageure.

Aussi bien l'OCDE que neuf grandes sociétés scientifiques allemandes considèrent qu'il y a urgence à régler le cas de la conservation et de l'archivage des données primaires.

### 2.2.3 Lois sur les archives et obligation de proposer

La plupart des législations gouvernementales sur les archives font référence à un concept global de documentation qui inclut toutes les informations enregistrées, indépendamment de leur support, que la Confédération et les cantons reçoivent ou produisent dans l'exercice de leur fonction publique.

Doivent ainsi par exemple être archivés, selon la loi fédérale sur les archives (LAR) [2] tous documents «qui ont une importance juridique ou administrative, ou qui ont une grande valeur d'information». Sont également archivés selon la LAR «*tous les documents de la Confédération qui ont une valeur juridique, politique, économique, historique, sociale ou culturelle*». Des citations analogues pourraient être tirées des législations cantonales.

Le fait de savoir si, dans tel ou tel cas, certaines données primaires (et des protocoles de recherche) devraient être archivées au sens des critères mentionnés est laissé à l'appréciation des services d'archives compétents. Ceux-ci sont les Archives fédérales pour l'Assemblée fédérale, le Conseil fédéral, les services du Parlement ainsi que les services centraux et décentralisés de l'Administration fédérale. Les établissements de la Confédération autonomes et les institutions fédérales assimilées organisent en revanche leurs propres centres d'archives et archivent leurs documents de manière autonome selon les principes de la LAR et en accord avec les Archives

fédérales. Font partie de ces institutions les deux écoles polytechniques, le Conseil des écoles polytechniques, ainsi que les instituts EMPA, EAWAG, WSL, PSI et Swissmedic.

Pour les universités cantonales et les hautes écoles spécialisées, ce sont les archives cantonales qui sont compétentes, bien que la plupart des universités gèrent un service d'archivage propre, soit en conservant leurs propres archives sur mandat du service d'Etat, soit pour servir d'intermédiaire entre l'université et ce dernier.

Tous les services soumis à la loi sur l'archivage sont tenus de satisfaire à l'obligation de proposer : ils doivent proposer à l'organisme d'archive compétent toute documentation dont ils n'ont plus besoin de manière régulière, afin de faire évaluer si elle doit être archivée ou pas.

Les documents qui auront été décrétés dignes d'être archivés au sens de la loi le seront définitivement. Dans certains cadres de réglementation (par ex. à la Confédération), le traitement nécessaire à l'archivage à long terme doit être pris en charge par les offices dépositaires, après le dépôt de leurs fonds. Dans certains cantons, cette tâche incombe au service d'archive compétent.

Jusqu'ici les archives d'Etat se sont peu, voire pas du tout préoccupées de la documentation issue de la recherche. Les impératifs de conservation (discutés au chapitre 3.1.3) touchant aux données primaires et destinés à garantir l'„intégrité dans la recherche“ restent de toute façon applicables, indépendamment du devoir de verser tel que l'entendent les lois sur les archives. Les deux contraintes ne s'excluent pas l'une l'autre, mais on ne peut pas non plus dire qu'elles se complètent absolument : d'une part, des données primaires qui ne sont pas réputées dignes d'archivage au sens de la loi doivent malgré tout être conservées en raison des directives sur l'intégrité (même si cette obligation est limitée dans le temps).

Mais d'un autre côté un archivage de données primaires qui méritent d'être conservées n'est pas forcément directement utile pour répondre aux directives sur l'intégrité puisque en général ces données doivent être préparées pour un archivage à long terme et converties dans des formats adaptés. C'est probablement surtout pour des formats binaires complexes qu'une telle conversion conduit à des pertes de l'information originale (par exemple en précision, types de données, structure, etc.), pertes sans doute acceptables du point de vue de l'archiviste, mais qui rendent impossible la vérification de résultats d'expériences qui avaient été obtenus à partir de données primaires non converties. Une conversion qui aboutirait à un tel état de fait serait inadmissible au sens des prescriptions actuelles sur l'intégrité, et il faudrait par conséquent conserver les données primaires dans leur forme originale en plus de l'archivage traditionnel.

### 2.3 Thématique centrale du rapport

Les objectifs de ce sous-projet de E-lib.ch mettent l'accent, en ce qui touche l'évaluation et la sélection des données primaires en vue de leur conservation, sur l'autodétermination des chercheurs et sur l'intérêt direct de ces données pour les projets de recherche à venir [1] :

*«Evaluation et sélection de données et d'objets numériques : une évaluation du contenu de données scientifiques ne peut et ne doit être entreprise par les seuls centres d'archives. Elle est pour l'essentiel l'affaire de ceux qui ont produit ces données. Ceux-ci doivent déterminer avant la livrai-*



*son de leurs données sous quelle forme ils pourraient en avoir à nouveau besoin ultérieurement, ou lesquelles ils aimeraient mettre à disposition d'autres chercheurs.»*

Au vu des lignes directrices fixées pour ce projet, le respect des cadres législatifs sur l'archivage ne fait pas partie intégrante de l'étude et relève essentiellement des archives d'Etat :

*«Une autre partie des données et objets numériques à archiver sont déjà rassemblés et pris en charge par des offices situés en amont. En règle générale, de telles démarches impliquent déjà une sélection et une évaluation qui aboutit à la décision de conserver.»*

Nous pensons cependant qu'une focalisation sur l'utilisation secondaire et sur une évaluation autodéterminée (*« l'évaluation est essentiellement l'affaire de ceux qui ont produit ces données »*) empêcherait de tirer tout le bénéfice qu'une solution d'archivage nationale serait en mesure d'apporter si, allant au-delà des besoins des chercheurs, elle s'efforçait aussi de répondre aux prescriptions de la législation gouvernementale sur les archives. Ceci est explicité par exemple au travers du critère mentionné dans la LAr de „grande valeur d'information“ attribué à des objets qui méritent d'être archivés : des données primaires qui, du point de vue de la discipline au sein de laquelle elles ont été produites, revêtent une „grande valeur d'information dans la perspective d'un usage secondaire ultérieur“, revêtent probablement aussi une „grande valeur d'information“ au sens de la LAr.

Nous nous en tenons donc, avec le modèle de base que nous proposons pour les données primaires, au point sur lequel le projet de mandat met l'accent principal, mais nous esquissons néanmoins une seconde composante du modèle qui a pour caractéristique de répondre aux exigences légales sur l'archivage.

Toujours dans le sens du mandat, nous limitons notre analyse aux données primaires, à savoir les données elles-mêmes accompagnées de la documentation nécessaire à les rendre compréhensibles et interprétables par des tiers, ou à reproduire les résultats des expériences pour les vérifier. Nous partons du principe que cette documentation est elle aussi fournie sous forme numérique (native ou rétrounumérisée).

Les *résultats d'expériences*, en revanche, ne sont pas retenus, pas plus que toute autre documentation accompagnant un projet de recherche qui permet d'en interpréter le déroulement et les étapes de dépouillement et d'analyse. Nous partons de l'idée que cette *documentation de recherche* au sens large – qui se présente par exemple sous forme de courriels ou de documents papier ou électroniques plus ou moins classés – peut être archivée si nécessaire par les instituts de recherche eux-mêmes dans leurs dépôts habituels, ou transmise aux archives d'Etat conformément à l'obligation de proposer pour être évalués et archivés comme n'importe quel autre document administratif.

La volonté exprimée dans le projet de la CBU de créer *«une archive numérique pérenne nationale à l'intention des hautes écoles, des instituts de recherche et des bibliothèques universitaires»* pour les données primaires et secondaires se limite, de notre point de vue, aux données brutes (accompagnées de leur documentation) ainsi qu'aux revues, et laisse volontairement de côté la conservation de documentation sous forme papier ou la gestion des dossiers de recherche.

Nous devons attirer l'attention, à propos de l'obligation d'archiver, sur le fait que vouloir se conformer aux *dispositions légales étatiques* soulève des problèmes juridiques et organisation-

nels complexes. Ceux-ci ne pourront être globalement résolus que de manière coordonnée par les nombreux services d'archives de la Confédération et des cantons qui seraient concernés par cette problématique.

La perspective de voir des milliers de projets de recherche, émanant de centaines de disciplines hautement spécialisées, remettre systématiquement leurs données primaires à un service d'archive d'Etat qui devrait ensuite en estimer la valeur nous paraît irréaliste. Car il faudrait en outre que les services d'archives cantonaux concernés, ainsi que ceux de nombreux instituts de recherche, puissent se mettre d'accord sur l'ensemble des conditions juridiques, organisationnelles et techniques qui devraient prévaloir en vue de la création d'une archive nationale centralisée.

## 2.4 Modèles

Conformément aux lignes directrices discutées au point 2.3, nous allons proposer un modèle pour l'archivage centralisé des publications scientifiques numériques, et deux variantes de modèle pour les données primaires. Puis nous esquisserons pour les données primaires quatre alternatives volontairement assez différentes

Le **modèle pour les publications scientifiques** (chapitre 6) correspond en grande partie à la stratégie «Allianz-Initiative ‚Digitale Information‘» de la Deutsche Forschungsgemeinschaft en Allemagne (voir chapitre 8.5). Il a pour ambition de garantir l'accès durable aux publications scientifiques numériques (*eJournals* et *eBooks*) une fois que les licences seront arrivées à terme et que les éditeurs auront supprimé les accès en ligne.

La **première variante pour les données primaires** (chapitre 5.1, modèle «Conservation avec autocontrôle») s'articule autour du principe du «self deposit», et se focalise de manière pragmatique sur le fait que les chercheurs doivent satisfaire aux nouvelles directives édictées en Suisse sur l'„intégrité dans la recherche“. Dans ce premier cas, la solution proposée ne prévoit aucune réglementation et repose par conséquent sur la responsabilité individuelle et l'initiative laissée aux chercheurs et à leurs institutions ; c'est à eux qu'il revient de définir les critères à appliquer en matière de délai de conservation, de périmètre et de format d'archivage des données primaires à conserver, en tenant compte des directives sur l'intégrité et de la valeur présumée de ces données pour une utilisation secondaire ultérieure.

Pour les chercheurs et leurs instituts, un tel service ne devrait en principe entraîner aucun surcoût, étant donné que les chercheurs sont censés prendre à leur charge la préparation des données primaires et la documentation qui les accompagne afin de se mettre en conformité avec la réglementation sur l'intégrité scientifique.

Le service propose par ailleurs une prestation de conseil et de support à la demande, mais n'offre à part cela qu'une «bitstream preservation», une certification de l'intégrité des données pour la durée du dépôt, des outils qui en facilitent la citation et, sur demande, leur catalogage en vue d'un accès public.

Il permet par ailleurs aux *instituts de recherche*, en tout temps jusqu'à expiration du délai de conservation prévu, de conduire *de leur propre gré* une évaluation du contenu des données, de

prolonger le cas échéant la durée d'archivage ou de faire remettre par le service central les données *qui ont encore de la valeur* à un service d'archives pérennes, au titre notamment de la législation gouvernementale sur les archives. Cette variante tient ainsi compte des recommandations de l'OCDE (voir chapitre 3.1.3.4) qui, au sujet de la conservation à long terme de données primaires en vue d'une utilisation secondaire, préconise de remplacer une modalité de sélection initiale précipitée par des évaluations périodiques.

Cette variante respecte fonctionnellement les RFC 4810 «Long-Term Archive Service Requirements» du «Long-Term Archive and Notary Services Charter» (LTANS), mais n'est pas totalement conforme à ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS).

La **deuxième variante pour les données primaires** (chapitre 5.2, modèle «Archivage avec obligation de dépôt») ébauche une solution réglementée en s'inspirant des législations gouvernementales sur les archives pour encourager une obligation générale de proposer pour la documentation relative à la recherche. L'entrée en force de l'obligation de proposer et l'évaluation du point de vue archivistique interviendraient au moment où les projets de recherche doivent de toute façon être évalués par les organes compétents : soit au moment de l'octroi du financement, ou de l'acceptation des résultats après évaluation, ou du *reporting*, c'est-à-dire après rédaction du rapport final.

Il revient ainsi aux experts respectifs de juger de la nécessité d'archiver (pour une courte ou longue durée) en fonction des intérêts de la discipline en question. Ces propositions sont systématiquement remises aux services d'archivage chargés de décider. Les experts et les chercheurs évalueront surtout l'utilité sur la durée des données primaires du point de vue d'une utilisation secondaire dans leur domaine de recherche, alors que le service d'archivage compétent en appréciera la valeur d'usage à long terme du point de vue des lois fédérales et cantonales sur l'archivage.

Si les données primaires méritent d'être archivées à long terme, le modèle prévoit qu'elles sont préparées par le service national d'archivage en collaboration avec les chercheurs, les instituts et les archives d'Etat en vue d'un archivage pérenne, et prises en charge par le service national qui se charge de la conservation par délégation des services officiels. Des commissions d'archivage spécifiques à chaque discipline édictent les directives et normes à respecter pour documenter les données et pour en préparer la conservation.

Ce modèle d'archivage à long terme des données primaires se conforme à la recommandation ISO 14721:2003 « Open Archival Information System Reference Model » (OAIS). Il en découle en particulier que le service national d'archivage se voit investi d'une responsabilité générale au niveau technique pour garantir le succès de cet archivage pérenne, et qu'en conséquence il doit disposer de toutes les compétences de décision nécessaires en vue d'une future conversion ou migration des données.

Ce deuxième modèle ne couvre toutefois que l'archivage de données primaires qui sont réputées dignes d'être archivées au sens de la législation sur les archives gouvernementales. Il ne propose donc pas de solution pour une conservation globale des données primaires qui répondent aux besoins des chercheurs en matière d'intégrité, qu'il s'agisse des données primaires qui ne méritent pas l'archivage, ou de toutes celles qui le mériteraient, mais qui, suite aux travaux de

préparation à l'archivage, ne seraient plus suffisamment conformes à l'original au sens des exigences d'intégrité.

Les quatre **modèles alternatifs** reposent tous sur le présupposé que le service central n'archive aucune donnée lui-même, mais assume „uniquement“ des fonctions de conseil et de coordination stratégique, tout en organisant un réseau national de compétences „Archivage pérenne“ sur les modèles allemand et français.

## 2.5 Structure du rapport

Le rapport s'articule ainsi :

- **Chapitre 3 – Situation, besoins et périmètre** : brève description des différentes catégories de documentation, mise en évidence du besoin et du bénéfice effectifs de les conserver, évaluation du degré d'utilisation supposé durant la période de conservation, description de la situation actuelle en matière de conservation et d'archivage.
- **Chapitre 4 – Protagonistes** : mise en évidence des intérêts, des rôles possibles et des conditions cadre sur le plan juridique des *stakeholders* potentiels à propos d'un service national centralisé d'archivage.
- **Chapitre 5 – Modèle pour les données primaires** : objectifs et spécifications des modèles de base que nous proposons pour un service national centralisé d'archivage pour les données scientifiques primaires en Suisse.
- **Chapitre 6 – Modèle pour les données secondaires** : objectifs et spécifications du modèle que nous proposons pour un service national centralisé d'archivage pour les publications scientifiques numériques en Suisse.
- **Chapitre 7 – Modèles alternatifs** : description des modèles alternatifs qui s'écartent notablement du modèle pour les données primaires
- **Chapitre 8 – Contexte européen** : survol, explication et typologies des projets, initiatives et stratégies pour l'archivage pérenne dans le contexte européen, avec limitation aux données primaires et secondaires. Enumération, avec brève description, de projets/réalisations qui ont le même objectif, en Suisse et à l'étranger.
- **Chapitre 9 – Bibliographie**

## 3 Situation, besoins et périmètre

Selon les catégories de documentation considérées, l'état actuel de l'archivage, les besoins ou le volume des données à prendre en compte peuvent varier fortement. Nous allons donc différencier, dans la suite de ce travail, les différentes catégories de documentation qui sont énumérées dans l'énoncé du mandat :

- Documentation liée à l'activité de recherche (données primaires et protocoles expérimentaux)
- Publications scientifiques électroniques
- Supports électroniques d'enseignement ou d'apprentissage („eLearning objects“)
- Documents provenant de rétrouvements

Les sous-chapitres qui suivent s'articulent en sections :

- Brève description de la documentation
- Utilisation potentielle de l'archive
- Besoin concret et degré d'utilisation
- Etat des mesures d'archivage

La documentation administrative au sens strict<sup>5</sup>, qui concerne l'administration courante, la planification et la gestion de l'institut lui-même, n'est pas comprise dans le périmètre de cette thématique.

### 3.1 Données scientifiques primaires

#### 3.1.1 Brève description

On appelle données primaires celles qui sont produites notamment dans le cadre d'expérimentations, de séries de mesures, de simulations par ordinateur ou de relevés statistiques. Ce qui les caractérise, c'est surtout la grande diversité de formats et de métadonnées qui s'y rapportent tout au long des différentes phases d'élaboration lors des recherches conduites dans les instituts. Dans une enquête de l'ETH Zurich [3], seulement 51% des personnes sondées ont confirmé que leurs données à archiver se présentaient dans un format standard et ouvert.<sup>6</sup>

---

<sup>5</sup> Pour l'évaluation et l'archivage des données administratives des hautes écoles qui dépendent de la Confédération, ce sont les centres d'archivage indépendants qui sont compétents selon la Loi fédérale sur l'archivage ; pour les universités cantonales, ce sont les services d'archives cantonaux.

<sup>6</sup> Il faut dire que le taux de réponse au questionnaire envoyé à tous les départements ne s'est élevé qu'à 17%.

Tout aussi diverses sont les documentations annexes, qui sont indispensables pour documenter la manière dont les données ont été acquises et rendre intelligible tout le processus de recherche. Les quantités de données également diffèrent fortement : une enquête conduite en 2004 sur la situation en Allemagne [19] indique un volume global de 100 GB par année pour les branches de sciences humaines (langues, sociologie), et, pour la quantité totale de données primaires à conserver en Allemagne, un volume de 1000 – 2000 TB par an, avec une forte tendance à la hausse.

La conservation à long terme des données primaires issues de la recherche et des protocoles d'expérimentation dans le cadre de projets financés par les deniers publics a pris une importance croissante ces dernières années, indépendamment des juridictions d'Etat, en raison des cas spectaculaires de falsification et de manipulation survenus lors de la publication d'articles scientifiques. La conservation des données de recherche aux fins de rendre possible la vérification des résultats publiés est aujourd'hui partie intégrante de la „bonne pratique scientifique“ et de „l'intégrité dans la recherche“. Il y a de plus en plus d'organes d'encouragement à la recherche de par le monde qui subordonnent désormais l'attribution de crédits à la condition que toute la documentation d'expérience soit conservée avec soin et reste accessible durant une période adéquate après la fin du projet.

#### 3.1.2 Utilisation possible des données conservées ou archivées

Une conservation à long terme ou pérenne de la documentation scientifique répond pour l'essentiel aux objectifs suivants :

- La **possibilité de comprendre et de vérifier les résultats de la recherche** est une exigence centrale pour „l'intégrité dans la recherche“ et la „bonne pratique scientifique“. Pour la majeure partie des recherches effectuées toutefois, un tel recours à la documentation selon ces critères ne devrait plus avoir de raison d'être au bout d'une dizaine d'années.
- **Réception de données non reproductibles** : bon nombre de données qui résultent de mesures dépendent des conditions du moment de leur capture, et ne sont de ce fait pas reproductibles. Quelques exemples : données météorologiques, relevés sismiques, statistiques de population, enquêtes d'opinion, études sur la diversité de la faune et de la flore, données sur la recherche climatique, astronomie. Cela leur confère sur la durée une plus grande valeur historique et scientifique qu'aux données reproductibles.
- **Utilisation secondaire** : certaines données peuvent très bien, après des années, servir à de nouvelles recherches conduites selon un point de vue renouvelé, en adoptant d'autres méthodes ou dans le but d'étayer une théorie nouvelle. Ou bien des chercheurs veulent reprendre des données issues d'autres projets pour les incorporer dans un nouveau cadre de réflexion. La constitution de séries temporelles historiques représente un troisième cas de figure, par exemple en climatologie ou en météorologie, où l'on accumule des données provenant de projets séparés dans un réservoir qui servira ensuite à produire des suites temporelles pour de nouvelles recherches. Une telle utilisation secondaire exige une documentation transparente au sujet des données et de leur capture.

- **Possibilité de faire référence aux données primaires :** l'utilisation secondaire exige que les données primaires utilisées provenant de tiers puissent être désignées de manière compréhensible et non ambiguë dans les publications, ce qui est en général impossible, puisque les données primaires ne sont généralement pas publiées. Dans ce cas, le recours à une centrale indépendante qui soit à même de proposer des identificateurs pour les lots de données qu'elle conserve ou archive simplifie sensiblement les choses et les rend plus transparentes.
- **Fondement de décisions politiques :** les résultats de la recherche sont susceptibles d'influencer des décisions politiques (pour le législateur) ou la gestion des affaires publiques (par exemple des données sur le facteur de risque des organismes génétiquement modifiés ou de la téléphonie mobile). Les principes fondamentaux de l'Etat exigent que l'action de l'administration, des autorités et du parlement lui-même puisse être toujours comprise au cours du temps quant à ses fondements. Cela peut signifier que, dans certains domaines sensibles, les données primaires qui fondent les résultats de la recherche doivent être conservées pour une longue durée, voir définitivement.
- **Histoire des sciences :** celle-ci explore le développement de la science en fonction de ses pôles d'intérêt, de ses interrogations, de ses méthodes et de ses résultats. C'est moins aux données primaires qu'à la documentation qui les accompagne qu'elle s'intéresse.

### 3.1.3 Besoins et degré d'utilisation

La fréquence d'utilisation de données scientifiques qu'on aura conservées sera fonction de la facilité d'accès à la documentation de recherche. En ce qui concerne l'utilisation secondaire (voir plus haut 3.1.2), une enquête conduite en Allemagne en 2004 [4] a montré que 60% des instituts sondés échangent régulièrement des données primaires avec d'autres institutions, mais que le recours à ces lots de données décroît fortement avec le temps pour ne jouer plus qu'un rôle anecdotique après 10 ans.

On a cependant repéré quelques exceptions d'utilisation intensive lorsqu'il s'agissait de données primaires non reproductibles (3.1.2) et dont la conservation et la diffusion étaient gérées de manière centralisée pour une discipline particulière, comme par exemple dans les quelque 50 réservoirs de données de l'International Council for Science<sup>7</sup> (ICSU) disséminés de par le monde. Les statistiques d'utilisation provenant d'autres centres d'archivage du même domaine montrent aussi dans certains cas des taux d'accès élevés. Ainsi le Economic and Social Data Service (ESDS) en Angleterre annonce, pour la période 2003-2004, 17'800 accès sur un fonds de 3'500 lots de relevés [5]. La disponibilité d'un accès en ligne est un facteur essentiel qui entraîne un haut degré d'utilisation.

Une enquête menée à l'ETH Zurich [3] révèle que 80% des sondés ont manifesté leur intérêt pour un système centralisé d'archivage à long terme de leurs données numériques, et que plus de 55% ont déclaré que, dans leur domaine, il existe des lots de données qui doivent être conservées pour très longtemps, soit bien au-delà de 10 ans. Si ces réponses ne disent rien sur le

---

<sup>7</sup> <http://www.icsu.org>



taux d'utilisation, on peut néanmoins en déduire qu'il existe un réel besoin d'accéder à ces données.

Il découle des directives sur l'intégrité de la recherche qu'il existe par ailleurs *per se* un besoin d'archivage pour environ 10 ans pour toutes les données primaires issues de recherches financées par les fonds publics.

#### 3.1.3.1 Directives sur l'intégrité de l'ESF et de l'OCDE

La European Science Foundation (**ESF**) recommande depuis 2000 dans ses directives «Good scientific practice in research and scholarship» [6] un délai de conservation d'au moins 10 ans pour les données scientifiques (Ziff. 37): *«Institutions must pay particular attention to documenting and archiving original research and scholarship data. Several codes of good practice recommend a minimum period of 10 years, longer in the case of especially significant or sensitive data.»*

L'**OCDE** recommande depuis 2007 dans ses «Principles and Guidelines for Access to Research Data from Public Funding» [7], - qui ne sont toutefois pas contraignants juridiquement parlant pour les états membres – que les données issues d'une activité scientifique financée par des fonds publics doivent être fondamentalement libres d'accès pour une certaine durée, ce qui sous-entend une conservation qui aille au-delà du terme du projet de recherche aux fins de vérification des résultats : *«Promote a culture of openness and sharing of research data among the public research communities within member countries and beyond. ... Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.»* En particulier :

*«Research institutions and government organisations should take formal responsibility for ensuring that research data are effectively preserved, managed and made accessible in order that they can be put to efficient and appropriate use over the long term.»* (lit. F)

*«Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure.»* (lit. M)

#### 3.1.3.2 Dispositions concernant la conservation en Suisse et en Allemagne

En Allemagne, c'est depuis 1998 déjà qu'une clause sur le devoir de conservation des données scientifiques figure dans les contrats passés entre la **Deutsche Forschungsgemeinschaft** (DFG) et les bénéficiaires de ses crédits de recherche [8]. C'est là une prescription que presque toutes les hautes écoles et centres de recherche du secteur public ont reprise dans leur réglementation interne.

En Suisse, le nouvel article 11a de la Loi sur la recherche (Loi fédérale sur la recherche [9] (SR 420.1) fait obligation aux organes d'encouragement de la recherche (académies des sciences et Fonds national) de veiller à ce *«que les recherches qu'elles soutiennent soient menées selon les règles de la bonne pratique scientifique»*.



Les quatre **académies suisses des sciences**<sup>8</sup> ont adopté le 28 juin 2007 des «Principes de base et procédures concernant l'intégrité dans la recherche scientifique» [10], qui contiennent également des prescriptions sur la conservation de données scientifiques et de leur documentation :

*«Pour permettre la supervision de la recherche, la reproduction des essais et l'analyse ultérieure des données selon d'autres points de vue, il convient de documenter toutes les données (y inclus les données brutes) d'une manière claire, complète et précise. Les données et matériaux doivent être conservés de sorte que soient exclus tout dommage, toute perte ou toute manipulation. Il en va ainsi non seulement pour les données manuscrites, mais aussi pour les données électroniques. Il est nécessaire de documenter les incidents particuliers, tels que par exemple la perte de données et les écarts du plan de recherche initial» (Paragr. 3.1)*

*«A la conclusion du projet, la direction du projet est responsable de la conservation des données et matériaux pendant une durée définie en fonction de la spécialité. Elle doit veiller à leur durabilité et à leur protection» (Paragr. 3.2)*

Contrairement aux directives allemandes, on ne précise ici aucun délai explicitement, mais on prévoit *«une durée adéquate en fonction de la spécialité»*.

Dans son nouveau règlement (entré en force le 1<sup>er</sup> mars 2008) [11], le **Fonds national** enjoint à ses bénéficiaires de *«réaliser les travaux de recherche... selon les règles de la bonne pratique scientifique»* (Art. 32). Il n'y est pas mentionné explicitement ce que sont ces règles (pas plus que le devoir de conservation des données scientifiques), mais il y a une allusion implicite aux règles des académies suisses, de l'OCDE et de l'ESF (voir ci-dessus 3.1.3.1). En revanche un devoir généralisé de conservation est implicitement contenu dans l'article 44 du *Règlement sur les subsides* [11]:

*«Mettre à disposition d'autres chercheurs les données recueillies durant les travaux de recherche soutenus par le FNS et les déposer dans des fichiers scientifiques reconnus, conformément aux prescriptions du FNS.»<sup>9</sup>*

Les règles édictées par les académies et le Fonds national ne valent que pour les projets qu'elles soutiennent financièrement. Mais depuis 2007 plusieurs organes liés à la recherche ont inscrit un devoir de conservation des données scientifiques dans leur propre réglementation. Ainsi, le «Reglement über die wissenschaftliche Integrität» de l'Université de Berne du 1.5.2007 [13] considère comme une faute le fait *«d'éliminer avant le terme du délai juridiquement prescrit des données primaires qui ont été archivées»*, sans pour autant que les dispositions légales soient mentionnées.

Les directives pour *«l'intégrité de la recherche et la bonne pratique scientifique»* de l'ETH Zurich [14] prescrivent que *«les données primaires doivent être stockées et sécurisées de façon telle qu'on puisse y accéder ultérieurement pour réutilisation ou vérification»*

*«Toutes les étapes de la collecte des données primaires (analyses statistiques, transformations, etc.) doivent être documentées sous une forme propre à la discipline (journal de laboratoire, autres supports) de façon telle que les résultats obtenus puissent être intégralement reproduits à partir des données primaires.» (Art. 11 al. 2)*

<sup>8</sup> Académie suisse des sciences naturelles SCNAT, Académie suisse des sciences médicales SAMW, Académie suisse des sciences humaines et sociales SAGW, Académie suisse des sciences techniques SATW

<sup>9</sup> Plus de précisions dans le *Règlement sur l'information, la valorisation et les droits relatifs aux résultats issus de la recherche*, du 17.6.08 [12]

### 3 Situation, besoins et périmètre

*«La direction de projet [...] veille notamment à ce que les données et les matériels soient conservés après le terme du projet pendant un laps de temps qui convient à la discipline en question, et, le cas échéant, qu'elles soient détruites à l'expiration du délai prescrit.»  
(Art. 11 al. 4)*

On entend par données primaires *«les données brutes recueillies de manière expérimentale qui n'ont pas été retravaillées»*, et par matériels *«les éléments de preuve ou les résultats des expériences sous une forme matérielle (par ex. des prototypes) ou immatérielle (par ex. des programmes informatiques)»*.

L'ETH transfère la responsabilité de la conservation des données aux personnes qui sont directement en charge du projet de recherche, à savoir les professeurs, les *senior scientists* ainsi que les assistants de recherche ou les post doctorants (art. 3). Bien que les règlements cités ici évoquent un délai de conservation *„reconnu comme adéquat pour la discipline“*, c'est pratiquement la règle des 10 ans valable dans les autres pays européens qui s'applique généralement.

#### 3.1.3.3 Identification des données issues de la recherche

Le problème du référencement va de pair avec le besoin de conserver les résultats de la recherche à des fins de vérification. Les auteurs d'articles scientifiques, c'est-à-dire ceux qui ont la paternité des données, de même que les utilisateurs secondaires potentiels, ont besoin d'une méthode pour se référer de manière univoque dans leurs publications à tel ou tel ensemble de données, comme cela est possible dans la littérature scientifique imprimée. Pour cela, ils doivent pouvoir d'une part faire référence à l'importance et à l'état des lots de données à un moment donné, et d'autre part identifier ceux qui les ont constitués.

Les données brutes de la recherche ne sont généralement pas publiées. On peut considérer cependant qu'on est en présence d'un référencement fiable lorsque les données brutes, une fois exploitées, sont déposées auprès d'un centre tiers, indépendant et agréé par la communauté des chercheurs, qui pourvoit chaque lot d'un identifiant univoque, le catalogue et le conserve avec toutes les garanties d'intégrité. Ainsi le besoin de pouvoir citer des données brutes non publiées rejoint-il ici les réglementations sur l'intégrité de la recherche, qui réclament une conservation qui garantisse que les données soient protégées de l'altération, de la perte ou de la manipulation.

Les premiers projets pilotes en matière de citation des données primaires sont soutenus en Allemagne par la Deutsche Forschungsgemeinschaft (DFG), comme par exemple *«CODATA – Datenpublikation und die Registrierungsagentur für wissenschaftliche Daten an der Technischen Informationsbibliothek Hannover»*<sup>10</sup> dans le domaine des géosciences.

La *«First World Conference on Research Integrity»* organisée en septembre 2007 par l'OCDE, l'ESF et le U.S. Office of Research Integrity (ORI) recommande même une inversion du principe de citation pour les données issues de la recherche ([15], p.28):

*«An important development will be the establishment of public digital repositories for primary research data with links to the published articles. In the USA an Inter Agency Working Group on Digital Data has been set up to propose such a repository system. In Europe an Alliance for*

<sup>10</sup> <http://www.tib-hannover.de/de/die-tib/projekte/codata/>

*Permanent Access to the Digital Records of Science has been created by major stakeholders in science and science information to help establish a European Digital Information Infrastructure.»*

De manière analogue aux organes publics qui financent des projets de recherche, quelques éditeurs scientifiques ont récemment édicté une „Data Availability Policy“ et n’acceptent plus que des manuscrits accompagnés de données justificatives bien documentées et fournies dans un format non propriétaire, et mises ainsi librement à la disposition d’autres chercheurs aux fins de vérification des résultats.<sup>11</sup>

#### 3.1.3.4 Conservation en vue d’une exploitation secondaire

On parle d’exploitation secondaire lorsque les données primaires sont utilisées à d’autres fins que celles pour lesquelles elles ont été générées.

Toutes les réglementations et recommandations évoquées dans les sections 3.1.3.1 et 3.1.3.2, notamment celles de l’OCDE, des académies suisses et du Fonds national, mentionnent explicitement l’exploitation secondaire comme un facteur important qui justifie la conservation durable des données primaires. Le Fonds national n’exige pas seulement que soient «*mises à disposition d’autres chercheurs les données recueillies durant les travaux de recherche*», mais aussi que celles-ci soient déposées «*dans des fichiers scientifiques reconnus, conformément aux prescriptions du FNS*» ([11], Art. 44), ce qui implique une forme durable d’archivage.

Entre la nécessité de conserver les données primaires au titre de l’intégrité scientifique et le besoin de les conserver à des fins d’exploitation secondaire, il y a au moins trois distinctions à faire : la seconde motivation ne prescrit pas de délai (puisqu’on ne peut déterminer à l’avance à quel moment on aura besoin des données), elle pose des critères plus stricts en matière d’intelligibilité des données, donc de documentation, et dépend enfin de la valeur scientifique reconnue de ces données. Car plus les coûts d’une conservation permanente seront élevés, plus on sera tenté d’y renoncer si les données primaires sont jugées „non pertinentes“ ou „de moindre utilité“ pour une exploitation secondaire. (C’est là une évaluation difficile à faire, qui ne peut l’être qu’au cas par cas, et qui ne devrait pas être laissée à celui qui est à la source de ces données). Les directives des académies suisses et de l’OCDE insistent par ailleurs sur le caractère exhaustif de la documentation des données.

Les directives de l’OCDE [7] recommandent, dans le cas de la conservation à long terme des données en vue d’une exploitation secondaire, de remplacer la sélection ou l’évaluation préalable par un examen périodique de l’intérêt qu’elles présentent („assessments“) :

*«While publicly funded research data are subject to the default rule of openness ..., this does not mean that all such data should be preserved permanently. The data archiving community should carry out cost-benefit assessments periodically and constantly develop and refine retention protocols to ensure that those data sets with the greatest potential utility are preserved and made accessible. Use of accepted retention protocols and thorough documentation of data should help to*

---

<sup>11</sup> Par exemple : «It is the policy of the *American Economic Review* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the *Review*, prior to publication, the data, programs, and other details of the computations sufficient to permit replication.» [5]

*reduce unnecessary duplication of effort as well as to establish the necessary selectivity in preservation.» ([7], chap. K)*

*«[Long-term retention] ... can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.» ([7], chap. M)*

Ce qu'il faut noter dans ces recommandations de l'OCDE, c'est que – contrairement aux directives de l'ETH Zurich par exemple – elles ne délèguent pas la responsabilité de la décision de conserver ni de la conservation elle-même aux seules personnes qui sont à l'origine de ces données (les responsables du projet), mais la confie en quelque sorte à une „archiving community“ (les organes qui financent et portent le projet), ce qui en fait une tâche institutionnelle.

#### 3.1.4 Etat des mesures d'archivage

La situation de l'archivage des données scientifiques est très hétérogène. Lorsqu'elles l'ont fait, les archives d'Etat ne se sont intéressées que très marginalement aux données scientifiques jusqu'ici. Et c'est finalement là où l'Etat lui-même a lancé de grandes opérations de recensement ou de mesures que quelques services d'archives nationaux ont commencé, durant les décennies 1970 – 1980, à en archiver les données primaires (épurées).

Depuis les années 1960 sont apparus plusieurs centres d'archivage qui avaient pour premier champ d'action les sciences sociales. Puis apparurent d'autres centres d'archivage pour le domaine des sciences naturelles. Ces archives spécifiques aux différentes disciplines ont toutes été rattachées tôt ou tard à des académies ou à des organismes scientifiques nationaux et financées au niveau du pays, certaines d'entre elles s'étant même partiellement associées à des réseaux internationaux [16].

Un tel tableau hétérogène vaut aussi pour l'ensemble de la Suisse. Certains domaines, fortement connectés aux réseaux internationaux et très impliqués dans les échanges et l'exploitation secondaire des données, se sont auto-organisés et obtiennent un peu d'argent public pour cela. C'est ainsi que certaines disciplines de sciences naturelles livrent régulièrement leurs données à des centres internationaux appartenant à leur domaine<sup>12</sup>. Et les sciences sociales disposent en Suisse, avec SIDOS<sup>13</sup>, de leur propre centre d'archivage depuis 1992.

<sup>12</sup> La Suisse aussi finance de tels centres de données. Communiqué de presse du Département fédéral de l'intérieur du 6.6.2008 : «Un inventaire établi en 2007 par MétéoSuisse avec des institutions partenaires a cependant montré que le financement de plusieurs séries de mesures essentielles n'était pas assuré. Cela concerne les observations du dioxyde de carbone, de la couverture de glace des lacs, de l'équivalent en eau de la neige, des glaciers, du permafrost et de la phénologie, ainsi que les deux centres de données internationaux « World Glacier Monitoring Service » rattaché à l'Université de Zurich et « Euro-Climhist » de l'Université de Berne. L'arrêté pris aujourd'hui par le Conseil fédéral assure durablement la poursuite de ces séries de mesures et le maintien des centres de données, ce avec 0,9 million de francs en 2010 et 1,6 million de francs par an dès 2011..»

<sup>13</sup> Le «Schweizerischer Informations- und Datenarchivdienst für die Sozialwissenschaften» (SIDOS) est une fondation de l'Académie suisse des sciences humaines (ASSH). Les missions principales du SIDOS, fondé en 1992, sont l'archivage des données ainsi que l'établissement d'un inventaire de la recherche. Depuis le 1<sup>er</sup> janvier 2008, SIDOS fait partie de FORS (Schweizer Stiftung für die Forschung in den Sozialwissenschaften) et est hébergé par l'Université de Lausanne.

Les géosciences et la climatologie sont, avec les sciences sociales, les disciplines les mieux organisées en matière de conservation et de facilité d'accès aux données primaires<sup>14</sup>. Mais il faut bien souligner que l'on a affaire ici d'abord à des initiatives individuelles, qui concernent presque exclusivement des données primaires non reproductibles pour une exploitation secondaire, et qui ne prennent en compte que certains lots de données bien choisis. On ne saurait prétendre qu'il existe une stratégie de conservation des données scientifiques à moyen ou long terme et valable pour toutes les disciplines ni au niveau international, ni à l'échelle de la Suisse.

L'énorme diversité des formats et des structures de données, déjà perceptible à l'intérieur d'une même discipline, représente le principal problème que doit affronter un centre national et multidisciplinaire d'archivage de données primaires. Cette diversité, ainsi que les jargons propres à chaque domaine, ne permettent de réaliser un catalogage homogène (classement et signalement) qu'à un niveau très rudimentaire.

La situation est encore pire en ce qui concerne le format des données : une enquête menée en Allemagne [4] a montré que, sur l'ensemble des données produites, 97,8% étaient stockées dans un format binaire propre à la branche, 0,3% en XML et environ 1,9% dans des formats textes reconnus. Il est vrai que la plupart des formats binaires propres à une branche sont généralement bien documentés. Il n'en reste pas moins qu'une éventuelle conversion des données binaires dans des formats standard<sup>15</sup> – comme cela se pratique pour la conservation à long terme de documentations textuelles – reste illusoire, et n'est même pas envisagée. D'abord parce qu'il n'existe que très peu de formats standard universels. Ensuite parce qu'une exploitation secondaire ou une réutilisation des données après plusieurs années ne peut se faire que dans ces formats originaux spécifiques.

Et enfin parce que les directives sur l'intégrité scientifique mentionnées aux sections 3.1.3.1 et 3.1.3.2 font référence aux données originales qui fondent les résultats de la recherche. (La directive de l'ETH [14] évoque les «*données originales résultant directement des tests ou acquises par toute autre méthode, non retravaillées*», celle des Académies suisses [10] parlent de «*données brutes*».) Une conversion des formats binaires spécifiques en formats standard devrait dès lors intervenir sans aucune perte d'information (quant à la valeur, au type ou à la structure de données), ce que l'on a peine à concevoir.

On peut dire en résumé que des solutions d'archivages multidisciplinaires, qui doivent agréger des données provenant de multiples sources hétérogènes, ne sont envisageables qu'à un niveau très général. Il s'agira généralement de simples dépôts pour des «paquets de données» contenant des fichiers en formats binaires propres à chaque discipline, et que le centre d'archivage ne signalera et ne cataloguera que sous forme de «paquets». Un signalement plus précis ne lui serait possible qu'à l'aide de la documentation préparée et incluse dans les paquets par les producteurs de données. De même, le centre d'archivage ne peut établir de stratégie d'archivage pour la migration au cours du temps des données spécifiques en format binaire contenues dans

---

<sup>14</sup> Presque tous les pays européens disposent d'un centre d'archivage pour les données des sciences sociales. 20 de ces centres se sont regroupés à ce jour sous l'organe faitier Council of European Social Science Data Archives (CESSDA, <http://www.cessda.org>) et cataloguent leurs données en coopération. Pour la climatologie et les géosciences, c'est le Council for Science (ICSU, <http://www.ngdc.noaa.gov/wdc/>) qui joue un rôle central.

<sup>15</sup> Par exemple dans le "Portable Document Format for Archiving" (PDF/A), ISO 19005-1:2005

les paquets, qui peu à peu deviennent obsolètes. C'est là une tâche que seul celui qui est à la source des données peut assumer, mais en tous cas pas le centre d'archivage.

Il n'en reste pas moins que la conservation à moyen et long terme des données issues de la recherche va gagner en importance à l'avenir, sur les plans national et international, au motif de l'„intégrité scientifique“ et de la „bonne pratique de la science“ (voir à ce sujet le chapitre 8 : Environnement européen) ; ceci dans le but de conserver dans une forme bien documentée sur le long terme les données primaires issues de recherches financées par des fonds publics et de les mettre à disposition de tiers pour une exploitation secondaire. On prend de plus en plus conscience aussi qu'il faudra pour cela mieux tirer parti des moyens financiers parfois importants engagés pour l'acquisition des données.<sup>16</sup>

Ainsi la règle habituelle des „10 ans“ de conservation qui prévaut dans les pays européens devrait devenir la norme dans les pratiques suisses, délai analogue à celui qui est exigé par le Code des obligations dans l'ordonnance sur les comptabilités (SR 221.431).

On va d'ailleurs vite s'apercevoir que la délégation de responsabilité de la conservation aux chercheurs eux-mêmes ou aux chefs de projets telle qu'elle est prévue par les directives n'est en pratique pas applicable : d'une part il est irréaliste de vouloir confier à des chercheurs soumis à une forte mobilité une responsabilité à long terme sur des lots de données localisés en différents endroits. D'autre part, une conservation pérenne sécurisée exige des ressources techniques et financières qui ne sont pas sous le contrôle des chercheurs ni ne peuvent être garanties sur la durée.

Les recommandations de l'OCDE (voir section 3.1.3.1) proposent une voie plus réaliste, qui ne délègue pas au chercheur individuel la responsabilité de la conservation et de la mise à disposition à long terme de ses données, mais considère qu'il s'agit là d'une responsabilité institutionnelle qui doit être prise en charge de manière tangible par les organes officiels d'encouragement et de financement de la recherche.

Pour les données primaires issues de recherches financées publiquement, l'OCDE prône de plus le principe de l'„Open Access“ qui s'appuie sur des serveurs institutionnels et des procédures de sélection destinés à conserver les données pertinentes pour une réutilisation ultérieure.

Dans le cadre de recherches financées par des fonds publics, rien ne s'oppose, ni dans les lois sur les archives ni dans les directives sur l'intégrité scientifique, à ce que les chercheurs ne transmettent leurs données à conserver à un tiers, au terme du projet, pour que celui-ci se charge à sa place de les stocker à long terme de manière sécurisée dans une forme garantie authentique. C'est même une solution à recommander si le tiers en question est une institution d'Etat ou une institution soutenue et contrôlée par les organes d'encouragement de la recherche.

---

<sup>16</sup> Voir aussi sur ce sujet [16]

## 3.2 Les publications électroniques

### 3.2.1 Brève description

On entend sous ce terme de publications électroniques essentiellement les revues scientifiques (*eJournals*) et les livres (*eBooks*) numériques, les bases de données, de même que les publications des hautes écoles (thèses, travaux de diplôme) ainsi que la littérature grise sous forme électronique (par ex. les *preprints*).

On distingue en fait trois modes d'acquisition de publications numériques :

- a) Acquisition physique de l'objet numérique pour stockage et utilisation en local sur le serveur de la bibliothèque.
- b) Acquisition uniquement du droit d'utilisation (licence) et d'accès via internet au produit qui reste stocké sur le serveur de l'éditeur.
- c) Comme b), mais avec en plus l'acquisition physique de l'objet numérique aux seules fins de conservation par la bibliothèque (sans accès possible pour les usagers).

L'accès aux documents numériques sur le site de l'éditeur ou sur celui de la bibliothèque est généralement restreint aux membres de l'université en question par les dispositions contractuelles. De même le recours à l'exemplaire de sécurité stocké à demeure (variante c) est contractuellement fortement limité.

Dans la phase d'analyse des besoins en vue d'un archivage de documents numériques (par la bibliothèque), mais aussi lors de la phase d'analyse des risques d'une solution potentielle d'archivage, on veillera à considérer si la bibliothèque possède par ailleurs une version imprimée des publications ou se contente des versions numériques (ce qui est de plus en plus le cas pour les journaux).

On tiendra compte également du fait que, à une moindre échelle, des années antérieures de publications qui n'étaient disponibles jusqu'ici que sous forme papier font l'objet de campagnes de rétrounumérisation.

### 3.2.2 Utilisation possible des données conservées ou archivées

L'objectif premier d'une conservation des publications électroniques est de garantir sur la durée la disponibilité des contenus numériques, surtout lorsque la bibliothèque ne possède pas de version imprimée. L'accès peut se voir compromis de plusieurs façons :

- La résiliation d'une licence entraîne en général la suppression des accès aux documents électroniques qui se trouvent sur le serveur de l'éditeur.
- Il arrive que d'anciens numéros de revues disparaissent du serveur de l'éditeur.
- La disparition d'un titre ou d'une maison d'édition tout entière peuvent compromettre l'accès aux documents pris en abonnement sur le serveur de l'éditeur.
- Lors d'une adaptation de son environnement technologique, l'éditeur peut être amené à ne pas migrer sur la nouvelle plate-forme des numéros peu demandés, et met ainsi fin à son offre.



Dans tous ces cas de figure, on ne peut assurer que la bibliothèque concernée pourra continuer à mettre ses publications électroniques à disposition de ses utilisateurs et, partant, poursuivre sa mission d'approvisionnement pour la communauté scientifique.

Nous approfondirons cette problématique de fond dans le cadre du modèle pour les données secondaires, au chapitre 6.

Si, dans le souci d'honorer le contrat d'approvisionnement à long terme, on veut protéger l'investissement consenti au moment de l'acquisition des licences, il ne reste que la variante c) présentée au point 3.2.1 : la bibliothèque doit acquérir une copie physique pour l'archiver elle-même, de façon à pouvoir utiliser cette copie de sauvegarde le moment venu si l'accès via le serveur de l'éditeur n'est plus possible.

Il y a là un gros potentiel de synergies en matière d'archivage :

- Cela n'a pas grand sens que chaque bibliothèque mette en œuvre sa propre archive institutionnelle pour les publications électroniques acquises et ses rétronumérisations. Il est relativement facile de centraliser cet archivage dans une solution commune à l'échelon du pays. Un seul exemplaire conservé dans des conditions optimales suffit pour un nombre relativement grand d'institutions. Et il n'y a pas de problème juridique aussi longtemps qu'il ne s'agit que d'un exemplaire de sauvegarde, sans possibilité d'accès pour les usagers de la bibliothèque.
- On peut, par la même occasion, renoncer à l'acquisition d'une version imprimée, ce qui économise des frais de traitement et de magasinage pour la bibliothèque.

Les économies qui résulteraient des synergies peuvent en fait s'avérer supérieures aux investissements nécessaires à un archivage sécurisé approprié.

### 3.2.3 Besoins et taux d'utilisation

Le recours à ces copies d'archive devrait rester très faible, car cet archivage ne vise qu'à garantir la pérennité de la copie de sécurité, alors que les accès se feront uniquement sur les serveurs de l'éditeur ou de la bibliothèque. Mais le besoin d'une solution d'archivage pérenne est en revanche incontestable au vu de l'exposé des risques esquissé au point 3.2.2. C'est cette motivation qui pousse à la mise en œuvre de solutions et de stratégies destinées à garantir à long terme l'accès et l'usage des documents numériques, telles qu'elles ont été élaborées pour la période 2005-2008 dans le cadre du projet «E-Archiving»<sup>17</sup> du Consortium des bibliothèques universitaires. La question est simplement de savoir si les solutions d'archivage pérenne doivent être implémentées de manière centralisée ou non.

### 3.2.4 Etat des mesures d'archivage

Il faut distinguer parmi les *eJournals*, *eBooks* et bases de données ceux qui sont édités en Suisse et ceux qui sont acquis ou pris sous licence par les universités auprès de fournisseurs étrangers. Si, pour les premiers, c'est la mission de collecte de la Bibliothèque nationale (*helvetica*) et des

---

<sup>17</sup> <http://lib.consortium.ch>. Le rapport final n'est pas encore accessible publiquement, raison pour laquelle nous nous référons aux résultats d'une étude préliminaire [17] de 2005.



bibliothèques cantonales (*cantonalia*) qui entre en jeu, ce sont en revanche les bibliothèques universitaires qui doivent pourvoir à la conservation des publications provenant d'éditeurs étrangers.

Dans le projet qui nous concerne, ce sont – selon les termes du mandat – les publications du second volet qui doivent être considérées, à savoir celles que les bibliothèques acquièrent à l'étranger et mettent à la disposition de leurs usagers.

La conservation et l'exploitation de publications électroniques qui appartiennent aux catégories des *helvetica*, des *cantonalia* et des écrits académiques ne sont pas tout à fait résolues, mais les domaines de compétence et le cadre organisationnel sont clairs. Ainsi la Bibliothèque nationale prépare une solution *ad hoc* pour les *helvetica* d'entente avec les bibliothèques cantonales et universitaires, dans le cadre du projet «e-Helvetica»<sup>18</sup>.

On ne peut pas dire qu'il existe aujourd'hui en Suisse un archivage satisfaisant des publications électroniques acquises à l'étranger. Les exemplaires physiques des documents numériques acquis sont stockés la plupart du temps dans chaque institution de manière plus ou moins contrôlée sur des supports exposés (CD ou bandes magnétiques). Et on ne fait guère usage du droit d'obtenir une copie physique en complément de l'accès sur le serveur de l'éditeur, faute d'une solution d'archivage satisfaisante.

### 3.3 Supports numériques pour l'enseignement

#### 3.3.1 Brève description

Les supports numériques pour l'enseignement ou l'apprentissage se présentent aujourd'hui sous de multiples formes, la plupart du temps en multimédia. Il peut s'agir de didacticiels complets d'enseignement virtuel, ou seulement de supports de cours, ou encore de préparations audiovisuelles qui couvrent tout un enseignement.

#### 3.3.2 Utilisation possible des données conservées ou archivées

La préparation d'un matériel d'enseignement numérique exigeant passablement de soin, il vaut la peine de protéger cet investissement pour pouvoir le réutiliser. On pense notamment à deux cas concrets :

- Une réutilisation dans une forme inchangée, ou comme partie intégrante d'un nouveau matériel de cours
- Pour l'histoire des sciences et de l'enseignement

#### 3.3.3 Besoins et degré d'utilisation

Pour autant qu'ils soient facilement accessibles, ces matériels devraient faire l'objet d'une assez forte demande de réutilisation. Mais comme aussi bien les matières elles-mêmes que les manières

---

<sup>18</sup> <http://www.e-helvetica.admin.ch>

res d'enseigner changent relativement fréquemment, l'utilisation secondaire de cette documentation devrait rapidement décroître après deux ou trois ans et elle ne présenterait plus alors qu'un intérêt historique.

#### 3.3.4 Etat des mesures d'archivage

Un véritable archivage organisé de supports de cours électroniques n'existe pas en Suisse. Cependant aussi bien l'ETH que la fondation SWITCH ont déclaré avoir entrepris de cataloguer et d'archiver en commun des supports de cours électroniques pour en faciliter l'échange et la réutilisation.

Un archivage durable de ces supports se révèle toutefois très coûteux en raison de leur hétérogénéité et du caractère multimédia complexe de ces objets, même en recourant à un standard tel que SCORM<sup>19</sup>. Comme ils ne présenteront à terme plus qu'un intérêt historique, on se limitera à n'en archiver qu'un petit échantillon.

## 3.4 Documents rétronumérisés

### 3.4.1 Brève description

En Suisse comme ailleurs, de nombreux projets s'attachent à numériser rétrospectivement des documents papier ou des supports audiovisuels analogiques, ceci afin de ...

- diffuser plus facilement ces documents d'archives ;
- préserver les originaux des dégâts dus à leur manipulation ;
- remplacer les originaux par des copies numériques, lorsque la conservation de l'original n'est plus possible ou devient trop coûteuse.

Comme pour les publications électroniques (voir section 3.2), on consulte généralement ces documents rétronumérisés via des serveurs dédiés et dans un format allégé.

Les copies „master“ en haute résolution qui résultent directement de la numérisation sont en fait des données brutes, pas ou très peu retravaillées, qui deviennent les exemplaires d'archives après avoir servi à produire les diverses copies destinées à l'utilisation courante. Ces copies numériques „master“ sont aujourd'hui de plus en plus régulièrement générées dans un format compatible avec les exigences de l'archivage.

En raison du haut degré de résolution retenu pour la numérisation, la quantité de données produite par les projets achevés ou en phase de démarrage est considérable et atteint rapidement plusieurs terabytes pour chacun d'eux.

---

<sup>19</sup> SCORM (Sharable Content Object Reference Model) est un modèle de référence pour l'échange de matériel électronique du domaine de l'enseignement proposé par l'Advanced Distributed Learning Initiative du Département américain de la Défense.

### 3.4.2 Utilisation possible des données conservées ou archivées

On conserve les copies numériques „master“ pour les raisons suivantes :

- Elles représentent le matériel idéal pour générer toutes sortes de copies d'exploitation en fonction des besoins, puisqu'elles sont les plus proches de l'original.
- Ces copies numériques deviennent aussi de véritables *masters* d'archivage lorsque le document original se dégrade rapidement et ne peut plus être conservé, si ce n'est à grands frais (copies numériques de substitution).

Mais on tend maintenant, chaque fois que les documents à numériser peuvent encore être conservés dans de bonnes conditions, à privilégier une nouvelle numérisation plutôt que de s'acharner à archiver les copies numériques „master“, ceci pour bénéficier des progrès marquants que connaissent les techniques de numérisation en termes de qualité et d'efficacité.

### 3.4.3 Besoins et degré d'utilisation

L'utilisation de ces copies d'archive devrait rester modeste dans la mesure où l'exploitation effective de ces copies numériques ne peut se faire qu'à partir de systèmes spécifiques.

Mais le besoin subsiste de conserver avec soin ces copies „master“, besoin qui pourrait être considérable au vu du grand nombre de projets de numérisation. Il n'est toutefois pas nécessaire de viser une solution nationale pour cette catégorie de documents.

### 3.4.4 Etat des mesures d'archivage

Une vision d'ensemble précise de la situation d'archivage de ces objets est impossible. Celle-ci diffère pour chaque institution, mais on peut dire que, dans la majorité des cas, les copies „master“ ne sont pas archivées de manière suffisamment sécurisée. Ce qui manque surtout, c'est une capacité de stockage en suffisance qui soit disponible sur des serveurs gérés de manière professionnelle.

## 4 Protagonistes

Comme déjà mentionné dans l'introduction (chapitre 2) et dans le chapitre 3, beaucoup d'institutions et d'intervenants affichant des prétentions et intérêts divergents sont concernés par une solution centralisée pour l'archivage à long terme de données primaires et secondaires. C'est pourquoi il est primordial, dans l'optique d'une solution réaliste, de bien cerner les protagonistes potentiels lorsqu'on tente d'identifier les organes de soutien susceptibles d'intervenir et les modalités de financement. C'est en fonction des objectifs stratégiques du modèle qu'on pourra préciser, parmi tous les intervenants possibles, ceux qui en fin de compte seront les protagonistes importants d'une solution, qu'ils la pilotent ou en soient les bénéficiaires.

Et réciproquement la faisabilité du projet dépendra de la manière dont les protagonistes se positionneront par rapport aux propositions faites. Il n'a malheureusement pas été possible, en raison des ressources affectées à cette étude, de procéder à une large consultation des parties concernées.

Nous avons néanmoins essayé de définir qui sont les principaux intéressés, avec leurs besoins et leurs attentes, sur la base de notre connaissance du domaine, du dépouillement des prescriptions légales et de la littérature spécialisée, ainsi que des résultats d'un séminaire<sup>20</sup> [18] conduit avec des représentants des différents groupes. Le tableau qui suit les passe en revue et met en regard leurs intérêts et attentes respectifs ainsi que d'autres éléments dont on doit tenir compte.

Afin de simplifier la présentation, on ne fera pas de différence ici (au contraire du chapitre 3) entre les diverses catégories de documents. Mais on gardera à l'esprit que les attentes des parties concernées seront différentes selon qu'on aura affaire à des données scientifiques, à des publications ou à des documents rétronumérisés (voir chapitre 3).

Comme le montre le tableau ci-dessous, certains protagonistes peuvent jouer plusieurs rôles dans la mise en place d'une centrale d'archivage à long terme de données primaires et secondaires, dans la mesure où divers acteurs peuvent intervenir aussi bien comme clients (comme mandant ou utilisateur final) que comme organes de soutien ou prestataires de service.

---

<sup>20</sup> Workshop II, 26.08.2008, ETH-Bibliothek à Zurich. Participants : Kay Bieri (Physikalisches Institut Uni Bern, Climate and Environmental Physics / Climat unibe), Martin Borer (ETH-Informatikdienste), Pascal Boutsouci (ETH-Bibliothek), Urs Brander (Paul Scherrer Institut / PSI), Rolf Brugger (SWITCH), Niklaus Bütikofer (ikeep AG), Jean-Marc Comment (Bundesarchiv / BAR), Michael Ehrismann (Konsortium der Schweizer Hochschulbibliotheken), Michael Gasser (ETH-Archiv, ETH-Bibliothek), Angela Gastl (Projekte Digitale Bibliothek, ETH-Bibliothek), Susanne Grulich Zier (ETH-Archiv, ETH-Bibliothek), Patrick Hinni (Nationalbibliothek / SNB), Martin Kaiser (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen / KOST), Peter Keller-Marxer (ikeep AG), Rudolf Mumenthaler (ETH-Bibliothek, Projekt e-rara), Christian Rohrer (SWITCH), Marco Schmidt (Functional Genomics Center Zurich, ETHZ und Uni ZH), Nicole Simonet (Konsortium der Schweizer Hochschulbibliotheken), Matthias Töwe (Konferenz der Universitätsbibliotheken der Schweiz / KUB). Rapport du séminaire : [18].

Stakeholder	En bref	Intérêts/Attentes	Points à éclaircir
<b>BIBLIOTHEQUES</b>			
<b>Bibliothèque nationale</b>	Service central de la Confédération pour la collecte et l'archivage à long terme des publications en rapport avec la Suisse (Helvetica).	Intérêt pour une coopération optimale au niveau national pour l'archivage des publications suisses.  Propose l'archivage de lots de données à prix coûtant (principe du coffre bancaire).	
<b>CBU (Conférence des bibliothèques universitaires)</b> <b>Bibliothèques universitaires</b> <b>Bibliothèques des HES</b>	La CBU se charge de plusieurs missions de coordination pour les bibliothèques universitaires. Elle pilote notamment le Consortium et les projets E-archiving et E-lib.ch.	La CBU a mandaté la présente étude sur l'archivage à long terme des données scientifiques, qui est une composante du projet national E-lib.ch.  Solution centralisée pour un archivage à long terme des revues électroniques (eJournals) de toutes les universités et HES	Délimitation des compétences et coopération entre bibliothèques et archives au sujet des données primaires.
<b>ARCHIVES</b>			
<b>Archives fédérales (AFS)</b>	Service d'archivage central de la Confédération pour le Conseil fédéral, le Parlement et l'Administration fédérale, au sens de la Loi fédérale sur l'archivage (LAr).	Contrôle du respect de la législation fédérale (LAr) par l'ETH et les instituts de recherche fédéraux (voir ci-dessous).  Propose l'archivage à prix coûtant pour des documentations de tiers, à condition que celles-ci soient conformes aux exigences des Archives fédérales en matière de format de données.	
<b>Centres autonomes d'archivage au sens de la Loi fédérale sur les archives</b>	On pense ici à l'ETH, à l'EMPA, à l'EAWAG, au WSL, au PSI. Ils archivent leurs données de manière autonome d'après les principes de la LAr et en accord avec les Archives fédérales.	Possibilité de conserver les données de la recherche pour environ 10 ans (intégrité de la science). Archivage conforme à la Loi fédérale. Le PSI conserve de manière plus ou moins systématique pour environ 10 ans les données qu'il produit.	

#### 4 Protagonistes

<b>Archives cantonales</b>	Service d'archivage central des gouvernements, parlements et administrations de chaque canton.	Contrôle, et parfois prise en charge, de l'archivage des dossiers des universités et hautes écoles spécialisées. Les archives cantonales sont souveraines pour apprécier la valeur d'archivage.	Les données primaires des universités et des HES ne sont actuellement ni remises aux archives, ni évaluées ou conservées par celles-ci.
<b>CECO (Centre de coordination pour la coordination de l'archivage à long terme des documents électroniques)</b>	Centre supporté par 18 cantons, le Liechtenstein et la Confédération. Il conduit des projets en matière d'archivage électronique, édicte des règles à suivre et observe les développements en matière d'archivage numérique.	Désir de proposer aux archives cantonales des prestations d'archivage professionnelles, surtout en matière de stockage et de „préservation bitstream“, par exemple pour les masters de documents rétroumérés par les archives (image, son, film).	Un engagement à long terme de tous les partenaires du CECO semble peu probable. Le CECO n'a lui-même aucune compétence décisionnelle.
<b>Centres d'archives des universités et HES</b>	Remplissent les missions d'archivage auprès de chaque institution, dans le respect des législations cantonales, et sous le contrôle des archives cantonales.	Manque de procédures et d'infrastructures pour l'archivage numérique.  Procédures de sauvegarde et d'évaluation (ces archives sont en principe les premiers interlocuteurs des groupes de recherche)	Tout comme les archives cantonales auxquelles ils sont subordonnés, ces centres d'archives manquent de ressources financières pour l'archivage numérique.  L'archivage de données scientifiques exige une connaissance très spécialisée des différents domaines.
<b>Centres d'archives spécialisés (pour des domaines spécifiques)</b>	Centres d'archives autonomes institués et supportés par des domaines de recherche particuliers (comme par exemple SIDOS pour les sciences sociales).	Probablement aucun intérêt pour une solution centralisée.	Contexte très hétérogène, avec de multiples solutions spécifiques. Les données sont généralement encodées selon des règles propres à chaque domaine.
<b>UNIVERSITES, HAUTES ECOLES ET INSTITUTS DE RECHERCHE</b>			
<b>Direction des universités et instituts de recherche</b>		Respect des règles sur l'intégrité de la science et sur la traçabilité de la recherche.  Sauvegarde de l'investissement : réexploitation des données de la science et sauvegarde des publications électroniques prises sous licence (ou rétroumérées).	

4 Protagonistes

<p><b>Groupes de recherche</b></p>	<p>Principal groupe d'intérêt : ensemble des chercheurs et projets de recherche financés ou soutenus publiquement, dans tous les domaines et de toutes institutions (universités, HES, ETH, instituts de recherche fédéraux et autres).</p>	<p>Les prescriptions en matière d'intégrité et les règlements d'application des organes de soutien à la recherche délèguent la responsabilité de la sauvegarde des données primaires aux chercheurs, en tant que chefs de projets. Mais en général ceux-ci ne disposent ni du savoir-faire, ni des infrastructures, ni des moyens financiers pour cela. Ils ont dès lors une préférence pour des solutions simples, qui leur apportent un soutien efficace en matière d'archivage professionnel.</p> <p>Intérêt des chercheurs pour une réutilisation de leurs données primaires ou de celles d'autrui.</p> <p>Intérêt pour un référencement de leurs propres données ou de celles d'autrui.</p>	<p>Pour avoir des chances de succès, un service central d'archivage aux fins de l'intégrité doit être simple d'emploi pour les chercheurs et ne représenter aucune contrainte ni conséquences financières pour lui, c'est-à-dire ne doit pas émarginer au budget de recherche.</p>
<p><b>SWITCH</b></p>	<p>Organisation sans but lucratif de la communauté scientifique suisse, qui maintient essentiellement des infrastructures de réseau de télécommunications académique.</p>	<p>Appuie les universités dans la production, la gestion et la mise en ligne de produits numériques d'enseignement (eLearning).</p> <p>Partenaire potentiel pour l'infrastructure d'archivage.</p>	
<p><b>ORGANES D'ENCOURAGEMENT ET DE POLITIQUE DE LA SCIENCE</b></p>			
<p><b>Fonds national (FNS)</b> <b>Académies suisses</b> <b>Secrétariat d'Etat à l'éducation et à la recherche (SER)</b> <b>Office fédéral de la formation professionnelle et de la technologie(OFFT)</b></p>	<p>Le Fonds national et les académies doivent enjoindre depuis 2008 les bénéficiaires de financements de respecter la bonne pratique scientifique (art. 11a de la Loi sur la recherche). Cette règle a été reprise par de nombreuses universités et hautes écoles.</p>	<p>Ces organes ont intérêt à ce que les directives qu'ils ont édictées sur la conservation des données primaires aux fins d'intégrité de la recherche soient remplies par les chercheurs eux-mêmes de manière aussi complète que possible.</p> <p>Sauvegarde de l'investissement : réexploitation des données de la recherche.</p>	<p>Même remarque que pour les groupes de recherche.</p>

Les **clients** potentiels d'un service centralisé d'archivage sont, pour les données primaires, en première ligne les chercheurs, les directeurs de projets et les groupes de recherche, et, pour les publications électroniques, les bibliothèques. L'un et l'autre en tant que clients sont à la fois mandant et utilisateur, dans la mesure où ils transmettent des données pour un archivage à long terme et exploitent eux-mêmes les données archivées.

Les **organismes porteurs** potentiels, lorsqu'on a affaire à des impératifs d'archivage prescrits par la loi, sont les services qu'elle désigne (voir section 4.1 ci-dessous), et en tous cas les bibliothèques pour les publications électroniques. D'éventuelles restrictions dans les contrats de licence des éditeurs pourraient à la rigueur faire obstacle à une centralisation de l'archivage au sein d'une seule bibliothèque universitaire.

Les organes d'encouragement comme les universités et les hautes écoles, qui ont récemment édicté de nouvelles directives sur la conservation des données primaires dans le cadre de l'„intégrité de la science“ (voir sect. 3.1.3), auront tout intérêt à ce que ces directives puissent être respectées de manière aussi scrupuleuse et efficace que possible par les chercheurs.

Comme **fournisseur de prestations** potentiel, on ne voit personne parmi les protagonistes énumérés qui soit à même d'offrir aujourd'hui une solution d'archivage à long terme pour les données primaires et secondaires. Mais plusieurs institutions peuvent apporter quelques briques qui pourraient contribuer à la construction d'un service centralisé d'archivage.

## 4.1 Cadre juridique

Dans les sections introductives 2.2.3 et 2.3, on a discuté des exigences fondamentalement différentes que manifestent les chercheurs et les organes liés à la recherche (devoir de conservation pour les besoins de l'intégrité de la science), et les centres d'archives d'Etat (lois sur les archives). Pour les données primaires, ces exigences qui visent soit une conservation limitée dans le temps soit un archivage à long terme ne peuvent que partiellement se compléter (sect. 2.3). Raison pour laquelle nous présenterons, dans le chapitre 5, deux variantes assez différentes pour les données primaires.

Toutes les hautes écoles et tous les instituts qui dépendent de l'Etat disposent d'une base juridique et sont soumis à la législation sur l'archivage. Les hautes écoles et instituts de recherche de la Confédération sont soumis à la Loi fédérale sur l'archivage du 26.6.1998 (LAR) [2], et les universités et HES qui dépendent des cantons sont soumises aux législations cantonales respectives.

Ceux qui produisent des données qui ne méritent pas l'archivage au sens des lois d'Etat (Confédération ou cantons) peuvent de toute façon organiser une conservation selon leur propres besoins ou en confier le soin à une institution quelconque, pour autant que les prescriptions des législations concernées sur l'archivage et la protection des données soient respectées. Il devrait s'agir dans l'idéal d'un service digne de confiance et de préférence soutenu par l'Etat.

Les données primaires considérées comme dignes d'être archivées au sens des lois gouvernementales sur l'archivage doivent être archivées conformément aux directives des centres compétents et sous leur contrôle.



### 4.1.1 La Loi fédérale sur l'archivage

La Loi fédérale sur l'archivage [2] (LAr) distingue trois populations :

- Les **services qui ont l'obligation de déposer**, qui doivent proposer aux archives compétentes toute documentation dont ils n'ont plus un besoin permanent. Les centres d'archives décident de l'intérêt de la documentation apportée, prennent en charge la partie qui le mérite, la conservent durablement et la transmettent à leurs usagers contre rétribution. C'est aux services déposants qu'incombe la nécessaire mise en forme des données à déposer pour un archivage de longue durée, préparation qui doit répondre aux critères des Archives fédérales.
- Les **centres d'archivage autonomes** au sens de la LAr sont les centres mentionnés dans l'ordonnance sur la LAr [19] qui ne remettent pas leur documentation aux Archives fédérales, mais qui remplissent eux-mêmes et à leurs propres frais leurs obligations qui découlent de la loi fédérale. Les modalités sont réglées bilatéralement entre les Archives fédérales et eux.
- Les services ou les personnes qui assument **des missions fédérales dans le cadre d'un contrat de droit privé** (VBGA art. 2 par. 3 et art. 9) : la partie prenante au contrat doit régler les questions d'archivage d'entente avec les Archives fédérales.

D'après les termes du mandat, ce sont les hautes écoles et les instituts de recherche qui sont au centre de la présente étude. Les deux hautes écoles fédérales (ETH Zurich et EPF Lausanne) et les instituts de la Confédération font partie des centres d'archivage autonomes au sens de la LAr, et sont énumérés de manière exhaustive dans l'ordonnance VBGA, annexe 2.a :<sup>21</sup>

- Institut de Recherche de l'Eau du Domaine des EPF (EAWAG)
- Institut fédéral de recherche sur la forêt, la neige et le paysage (WSL)
- Laboratoire fédéral d'essai des matériaux et de recherche (EMPA)
- Ecoles polytechniques fédérales (Lausanne et Zurich)
- Paul Scherrer Institut (PSI)
- Conseil des Ecoles polytechniques fédérales

L'ETH Zurich a conclu depuis un certain temps déjà les accords prévus par la loi avec les Archives fédérales. On peut en déduire que les autres conventions qui sont encore pendantes contiennent fondamentalement les mêmes dispositions. Celles-ci permettraient aux centres d'archivage des hautes écoles et des instituts fédéraux de collaborer en matière d'archivage, et de mettre en œuvre une archive numérique commune. Pour autant que les matériels archivés restent la propriété des centres et sous leur contrôle, aucune disposition juridique ne pourrait s'opposer à une telle manière de faire.

Mais on sait que, dans le cadre des hautes écoles et des instituts de la Confédération, il existe un bon nombre de cadres contractuels de nature privée (par exemple pour des projets financés par des tiers). Il n'a pas été possible, dans le cadre de cette étude, d'évaluer dans quelle mesure

---

<sup>21</sup> On ne mentionne ici que les hautes écoles et les instituts de recherche. La liste de l'ordonnance LAr contient encore d'autres institutions qu'il n'est pas utile d'évoquer ici.

l'archivage est explicitement réglé dans ces dispositions contractuelles. Mais on peut supposer que ce n'est le cas que pour un nombre infime de contrats de recherche.

Le droit fédéral est applicable à l'ensemble des dossiers des services qui relèvent de la Confédération. La définition légale de ‚dossiers‘ est ici très générale :

- *«Par documents, on entend toutes les informations enregistrées sur quelque support que ce soit, qui ont été reçues ou produites dans le cadre de l'accomplissement de tâches publiques de la Confédération, ainsi que tous les instruments de recherche et toutes les données complémentaires qui sont nécessaires à la compréhension et à l'utilisation de ces informations.» (LAR art. 3 al. 1) [2]*

Cette définition ne figure pas seulement dans le texte de loi, mais aussi dans le message du Conseil fédéral sur la loi sur l'archivage du 26.2.1997 [20]:

- *«Pour les services fédéraux dont la tâche consiste à collectionner et à archiver des documents déterminés (par ex. les nombreuses archives du domaine des EPF, la Bibliothèque nationale suisse, etc.), la Loi sur l'archivage ne s'appliquera évidemment qu'aux documents relatifs à leur gestion et non pas à leurs collections »*

Ce concept de documents ne comprend donc pas les collections de livres et de revues, de données et de papiers d'archives collectées par des tiers. Les collections de livres et de revues acquises ou constituées par les bibliothèques de l'ETH ou les instituts satellites ne s'inscrivent donc pas dans le périmètre de la législation sur l'archivage.

Les activités de recherche et développement sont en revanche au cœur de l'activité des universités et instituts de recherche. Comme de la documentation produite par des tiers doit fréquemment être incorporée par ces institutions à côté des travaux de recherche, il subsiste une zone d'ombre pour laquelle on a besoin d'un critère précis pour définir si cette documentation provenant de tiers fait partie des collections générales ou accompagne simplement un dossier de projet. Dans les discussions techniques, on invoque le critère d'importance pour la conduite du projet : les dossiers (y compris les données) provenant de tiers sont considérés comme partie intégrante du projet lorsqu'ils sont essentiels pour sa poursuite et son achèvement, et qu'on ne peut y avoir accès par un autre canal (comme par des publications).

La Loi sur l'archivage de la Confédération (LAR) ébauche dans les grandes lignes les critères d'après lesquels on doit juger de la nécessité d'archiver :

- *Tous les documents de la Confédération qui ont une valeur juridique, politique, économique, historique, sociale ou culturelle sont archivés. (LAR Art. 2 al. 1) [2]*
- *Ont une valeur archivistique les documents qui ont une importance juridique ou administrative ou qui ont une grande valeur d'information. (LAR Art. 3 al. 3)*

Les accords mentionnés plus haut entre les Archives fédérales et les centres d'archivage autonomes permettent à ceux-ci d'édicter et de mettre en œuvre de leur côté un concept d'évaluation. L'ETH Zurich a précisé son propre concept sous la forme de directives pour l'évaluation de la documentation à archiver par l'ETH, datées des 29.1./23.4.02 [21]. Mais ces directives ne concernent pas explicitement la documentation issue de la recherche, de sorte qu'il n'existe pas de décisions préalables auxquelles se référer dans le cadre de cette étude centrée avant tout sur les données primaires.

Toute documentation qui n'aura pas été retenue par les organes compétents (centres d'archives autonomes ou Archives fédérales) ne sera pas considérée comme un bien à archiver et donc plus soumise à la LAr.<sup>22</sup> Leurs propriétaires sont alors libres de la conserver aussi longtemps qu'il le faut, soit aussi longtemps que l'exigent les délais usuels ou d'autres cadres juridiques que la LAr.

#### 4.1.2 Le droit cantonal

Les universités cantonales sont soumises aux législations sur l'archivage de leurs cantons respectifs. Elles entretiennent généralement leur propre service d'archivage dans le cadre des prescriptions légales, services qui la plupart du temps ne servent que de station intermédiaire vers les archives cantonales compétentes.<sup>23</sup>

Les universités sont soumises à un devoir de déposer qui, selon les cantons, est plus ou moins explicitement formulé, et qui attribue un plus ou moins grand pouvoir de décision en matière d'évaluation aux archives cantonales respectives.

Les HES sont des organismes de droit public cantonaux ou supracantonaux, dotés d'une personnalité juridique propre, et qui s'organisent spontanément dans les limites des lois et contrats respectifs qui les fondent. Pour les écoles instituées par un concordat intercantonal, il n'est pas toujours facile de voir, de l'extérieur, quelles sont les lois cantonales sur l'archivage qui prévalent à tel ou tel endroit. D'après ce qu'on a pu voir au cours de cette étude, les contrats de concordat n'évoquent jamais la question de l'archivage.

Les dernières lois cantonales sur l'archivage empruntent toutes le concept de 'documentation' au droit fédéral, en lui conférant un sens identique et donc sans en compléter la définition.

D'après les recherches menées dans le cadre de cette étude, il n'est nulle part explicitement exclu de recourir aux prestations de tiers pour l'archivage de documents numériques.<sup>24</sup> Le projet de nouvelle loi bernoise sur l'archivage [22] prévoit même explicitement, à l'art. 7, al.2, la possibilité d'avoir recours aux services d'entreprises spécialisées pour l'archivage de fonds d'archives numériques.

#### 4.1.3 La législation sur les bibliothèques

On trouve généralement, dans les lois qui fondent les bibliothèques, la mention d'une mission de collecte et d'approvisionnement en documentation. Qu'il s'agisse de la Bibliothèque nationale, des bibliothèques cantonales ou des bibliothèques universitaires, c'est là une mission qui n'est pas limitée dans le temps. C'est pourquoi on peut considérer ces établissements comme de véritables 'bibliothèques de conservation'. La possibilité de se défaire du matériel acquis n'est pas mentionnée dans leurs cadres juridiques, mais des éliminations ont bel et bien lieu dans la pratique et sont laissées à l'appréciation de chaque institution. Les *helvetica* et les *cantonalia* constituent cependant des collections qui doivent être conservées pour toujours.

---

<sup>22</sup> Les dispositions de la LAr sur l'accès en ligne (par. 3), sur l'utilisation commerciale (art. 19) et sur le caractère inaliénable (art. 20) ne concernent que les biens à archiver.

<sup>23</sup> Comme par exemple dans les universités de Bâle, Berne ou Zurich.

<sup>24</sup> Une conservation à l'étranger devrait de toute façon faire l'objet d'un examen juridique plus approfondi.

## 5 Modèles pour les données primaires

La Conférence des bibliothèques universitaires suisses (CBU) vise, dans le projet qui nous occupe ici, «l'élaboration d'un modèle pour un archivage pérenne centralisé des données numériques primaires et secondaires issues de la recherche en Suisse» (intitulé du projet<sup>25</sup>).

Nous proposons dans les chapitres 5.1 et 5.2 ci-dessous deux variantes de modèle pour les données primaires, en nous appuyant sur l'énoncé des objectifs du projet, sur nos analyses des chapitres 3 (Besoins et périmètre), 4 (Protagonistes) et 8 (Environnement européen) ainsi que sur l'exposé de la thématique centrale au paragraphe introductif 2.3.

### 5.1 Modèle «Conservation avec autocontrôle »

Ce modèle repose sur le principe du «self deposit» et découle du fait que les chercheurs doivent respecter la directive nouvelle en Suisse sur l'„intégrité de la science“. Le service national, désigné ci-après par «le service d'archivage pour les données primaires», ne prévoit aucune réglementation et délègue par conséquent la responsabilité et l'initiative de l'archivage aux chercheurs et instituts de recherche eux-mêmes. C'est à eux qu'il appartient de définir les critères relatifs à la durée d'archivage, au périmètre de collecte et au format de stockage qu'il faudra appliquer à des données primaires qui doivent être conservées aux fins de l'intégrité ou/et en vue d'une exploitation secondaire ultérieure.

Fonctionnellement, cette variante s'inspire des RFC 4810 «Long-Term Archive Service Requirements»<sup>26</sup> [23] du «Long-Term Archive and Notary Services Charter» (LTANS), mais n'est pas totalement conforme à ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS).

Le service d'archivage que nous proposons dans ce modèle permet aux chercheurs et aux chefs de projet de remplir leur devoir de conservation des données primaires aux fins de l'intégrité, et propose des prestations supplémentaires telles que le conseil, la possibilité de citer les données ou leur catalogage éventuel sur des serveurs publics. Il est impératif que ce service n'émerge pas aux budgets des chercheurs ou des projets.

Ce service d'archivage remplit les fonctions de base suivantes :

---

<sup>25</sup> On a évoqué au chap. 2.1 la différence de signification entre les termes *archivage* et *conservation*, resp. *archivage* à long terme et *conservation* à long terme, tels qu'ils apparaissent dans les textes législatifs et dans le vocabulaire des archivistes.

<sup>26</sup> <http://www.ietf.org/rfc/rfc4810.txt>

- Sur demande, un professionnel d'un réseau de compétences apporte conseil et support pour la préparation des données.
- Les données déposées sont stockées de manière sécurisée et protégées contre tout accès non autorisé.
- Une «bitstream preservation» pour les données déposées.
- Vérification à valeur de preuve de l'intégrité (inaltération) des données par un système d'horodatage certifié (cachet temporel) conformément à la loi sur la signature (ZertES).
- Référencement, dans les publications, des données déposées à l'aide d'un Digital Object Identifiers (DOI) défini et géré par le service d'archivage.
- Les chercheurs peuvent octroyer à un tiers un accès unique en lecture aux données qu'ils auront déposées (en fonction de la quantité de données en jeu, soit directement en ligne, soit par mandat écrit au centre d'archivage le priant de remettre ces données à telle ou telle personne).
- Les chercheurs peuvent s'ils le veulent signaler les données qu'ils auront déposées sur un catalogue en accès public.

Le service d'archivage offre en outre aux organes de la recherche la possibilité en tous temps, mais au plus tard à l'échéance du délai de conservation prévu, de soumettre les données à un expert pour une évaluation de leur intérêt et, le cas échéant, de prolonger le délai de conservation ou de faire remettre les données encore intéressantes à un service d'archivage à long terme. Cette variante respecte ainsi les recommandations de l'OCDE (voir section 3.1.3.4) qui préconisent des réexamens périodiques plutôt qu'une sélection hâtive et prématurée des données à conserver.

### 5.1.1 Objectifs stratégiques

Le modèle doit permettre d'atteindre les six objectifs stratégiques suivants :

- G1. **National** : le service d'archivage doit être au service des chercheurs de toutes les institutions de recherche en Suisse : à côté du domaine des ETH, notamment les universités cantonales et les HES, ainsi que les instituts de recherche financés par des fonds publics.
- G2. **Pluridisciplinaire** : le service d'archivage doit être au service des chercheurs de tous les domaines de la recherche, et donc pas seulement au service de quelques disciplines ou institutions qui auraient des exigences particulières.
- G3. **Contributif** : le service d'archivage doit contribuer efficacement à la mise en œuvre pratique des directives „intégrité de la science“ prônées par les organes de soutien, les universités et l'art. 11a de la Loi sur la recherche ; il doit permettre le référencement des données primaires dans les publications, faciliter la réutilisation de ces données et garantir l'accès à celles qui ont gardé leur valeur.
- G4. **Politique de la recherche** : le service d'archivage doit faire en sorte que la responsabilité de conserver les données scientifiques de manière appropriée ne soit pas uniquement l'affaire des chercheurs individuels, mais puisse être considérée, dans le sens des recommandations de l'OCDE [7], comme une tâche incontournable de l'infrastructure nationale de promotion de la recherche et de la culture.

- G5. **Transparence** : par souci de transparence, et en raison des coûts induits et de la complexité rencontrée sur les plans organisationnel et technique, le service d'archivage doit être assumé par une seule institution qualifiée, reconnue comme digne de confiance par les organes porteurs et de ce fait mandatée et soutenue financièrement à cette fin.
- G6. **Non obligatoire** : les organes porteurs ne doivent en aucun cas contraindre à utiliser le service d'archivage, ce qui signifie que les chercheurs et institutions doivent pouvoir y recourir sur une base absolument volontaire.

### 5.1.2 Exposé des motifs

La problématique de la conservation des données primaires issues de la recherche a été discutée dans les chapitres 2.2.3 et 3.1 (voir aussi l'environnement européen aux chapitres 8.4 et 8.5).

Le développement de stratégies, de projets et de services dédiés à l'archivage de données primaires, considérés comme composants essentiels de l'infrastructure de recherche d'un pays, a fait l'objet de recommandations récentes de l'OCDE [7] (voir les sections 3.1.3.1 et 3.1.3.4). A cela s'ajoute que, depuis peu en Suisse, on prône la conservation systématique des données primaires pour un laps de temps „approprié“ au domaine de recherche (en général pour 10 ans à l'étranger), ceci au motif de l'„intégrité de la recherche“ ou de la „bonne pratique de la science“ et en vue de la réutilisation des données. Les académies scientifiques suisses et le Fonds national (se fondant sur le nouvel article 11a de la Loi sur la recherche) ainsi que plusieurs universités et hautes écoles ont édicté des directives allant dans le même sens (voir chap. 3.1.3.2).

Ces directives délèguent aux seuls chercheurs et chefs de projet la responsabilité d'une conservation appropriée. Vu la volatilité des projets et la grande mobilité des chercheurs, assumer au fil du temps la responsabilité de conserver correctement de multiples lots de données répartis en divers endroits représente une mission quasi impossible pour le chercheur.

D'autant plus que l'effort à consentir pour conserver ou archiver de manière durable ou permanente un lot de données primaires ne se résume pas à un investissement unique, mais entraîne des coûts d'exploitation récurrents.

Sous ces deux aspects – responsabilité et coûts – on voit que, dans la pratique, la conservation ne peut être assumée par le seul chercheur mais doit, dans le meilleur des cas, être assumée par l'institution auprès de laquelle les données primaires ont été générées. Celle-ci devrait donc garantir, sous mandat du chercheur, que les données importantes sont conservées de manière sûre et restent exploitables même après que l'intéressé a quitté l'institution et que son compte d'utilisateur a été fermé. Mis à part le fait qu'elles offrent de la mémoire de stockage en suffisance, les institutions de recherche ne sont aujourd'hui quasi pas préparées à cela, ni techniquement, ni sur le plan organisationnel.

Ce n'est que tout récemment que l'impératif de conserver intégralement les données primaires et la situation insatisfaisante qui en résulte pour les chercheurs ont été pris pour thème par l'OCDE [7] et par la European Science Foundation [6] et considérés avec attention sur les plans national et interinstitutionnel. Et ce sont entre autres l'«Alliance for Permanent Access to the Records of Science in Europe» (Alliance PARSE) [24] (voir chap. 8.4) fondée en 2006 et rassemblant des membres à haute visibilité, de même qu'en Allemagne l'«Allianz-Initiative ,Digitale

Information’» [25], qui réunit neuf sociétés de recherche allemandes conscientes d’„un urgent besoin d’action” dans tous les secteurs de la recherche (voir chap. 8.5), qui assument le leadership en Europe. La Suisse n’est pas représentée dans l’alliance PARSE. Tout comme la place de la recherche suisse, avec 18 projets financés à hauteur de 70 millions par la Commission européenne dans le cadre du programme-cadre FP6 «IST-2005-2.5.10: Access to and preservation of cultural and scientific resources» (voir chap. 8) était sensiblement plus faiblement représentée<sup>27</sup> que, par exemple, l’Autriche.

Bien que ce soient quasi les mêmes critères de conservation des données primaires qui prévalent en Suisse et dans les pays européens, et qu’on y soit confronté aux mêmes problèmes organisationnels, techniques et financiers, on ne peut pas dire que la Suisse se soit associée de manière substantielle au débat international sur la recherche de solutions concernant les données primaires, ni qu’elle en profite beaucoup.

Le présent projet de la CBU entend donc engager la discussion sur le plan national d’une part, et d’autre part esquisser une première solution pragmatique qui puisse profiter à tous les chercheurs, et particulièrement aux domaines de recherche plus modestes qui sont mal armés en matière de conservation de données primaires.

Que ce soient les bibliothèques universitaires qui prennent cette initiative ne doit pas étonner, puisque c’est avant tout à l’intention des chercheurs qu’elles remplissent leurs missions „traditionnelles”, et on voit d’ailleurs qu’à l’étranger l’initiative vient d’abord des bibliothèques et non pas des archives.

### 5.1.3 Objectifs opérationnels

Dans le cadre des objectifs stratégiques susmentionnés, nous proposons les objectifs opérationnels suivants pour les données primaires :

- P1. **Facilité d’usage** : le service d’archivage pour les données primaires doit être utilisable de manière simple et directe par les chercheurs individuels (et surtout les doctorants), quelle que soit l’ampleur du projet<sup>28</sup>.
- P2. **Neutralité des coûts** : le service d’archivage des données primaires doit, par principe, être accessible sans frais pour les chercheurs individuels, et sans coûts latéraux pour les projets de recherche<sup>29</sup>.
- P3. **Standards** : le service d’archivage doit suivre les RFC 4810 «Long-Term Archive Service Requirements»<sup>30</sup> [23] der «Long-Term Archive and Notary Services Charter»<sup>31</sup> (LTANS).

<sup>27</sup> Participation des Archives fédérales à PLANETS (Preservation and long-term access to our cultural and scientific heritage), de la SRG/SSR à MEMORIES (Design of an audio semantic indexation system allowing information retrieval for the access to archive content), des Archives Internationales de Musique Populaire à ETHNOARC (Linked European archives for Ethno-musicological research) et de l’Université de Genève à MULTIMATCH (Multilingual/Multimedia access to cultural heritage).

<sup>28</sup> On pense d’abord ici à l’ampleur organisationnelle et financière. Il faut bien sûr placer des limites raisonnables vers *le haut* en fonction de la quantité de données à traiter. Ainsi il ne doit ni ne peut être question de desservir certains „Data Hot Spots” qui produisent des douzaines de terabytes par année. De tels projets disposent de toute façon de leur infrastructure propre pour stocker et archiver leurs données (exemple : le PSI)

<sup>29</sup> On peut imaginer d’établir une règle qui imposerait, pour des projets d’un certain type ou d’une certaine taille, de prévoir les coûts d’un archivage à long terme des données primaires dès la phase de demande de crédits.



- P4. **Inscription** : les chercheurs qui désirent utiliser le service d'archivage lancent une requête de dépôt via leur institut qui chapeaute le projet (en règle générale, c'est lui qui sera le détenteur des données). Les données utilisateur requises pour l'accès seront celles qui sont actuellement définies dans le système «Authentifizierungs- und Autorisierungs-Infrastruktur (AAI)»<sup>32</sup> de SWITCH, pour autant que l'institut y soit affilié. Si le chercheur quitte la Suisse, l'accès doit être réglé de cas en cas.
- P5. **Self-deposit** : le service d'archivage offre aux chercheurs au minimum la possibilité de transmettre leurs données en format original de manière autonome et efficace, via une liaison Internet sécurisée ou sur support amovible, sans qu'il soit nécessaire d'adapter les données ou d'intervenir sur leur format. Il incombe au chercheur de fournir une documentation circonstanciée des données et de choisir un format de données durable, conforme aux *best practices* de la branche.
- P6. **Evaluations périodiques** : le service d'archivage permet aux institutions de recherche concernées de faire évaluer par un spécialiste la valeur à long terme des données déposées, avant le dépôt initial ou tout au long du délai de conservation.
- P7. **Fonction de conseil** : selon P5, le *self-deposit* s'effectue dans le format et d'après la documentation des données originales choisis par le chercheur. Il peut arriver que l'un et l'autre se révèlent inadéquats ou insuffisants pour une conservation à long terme ou une interprétation correcte. Le service d'archivage offre alors au chercheur la possibilité de se faire conseiller en matière de documentation et de format durable par un expert de la conservation de données de sa branche.
- P8. **Bitstream preservation** : durant le délai de conservation convenu, le service d'archivage, sur mandat du chercheur, garantit au moins la disponibilité et la préservation des chaînes de bits qui constituent les données qui lui ont été confiées. Cela inclut la protection contre les accès non autorisés et les catastrophes, ainsi que le stockage sécurisé.
- P9. **Intégrité** : le service d'archivage permet au chercheur de prouver légitimement, et en tout temps à partir du moment du dépôt, le caractère authentique (c'est-à-dire l'inaltération) des données qu'il aura déposées. A cet effet, le service d'archivage attribue automatiquement, au moment du dépôt, un cachet d'horodatage certifié conforme à la loi fédérale sur la signature électronique ZertES<sup>33</sup> [26]. Il veille en outre au renouvellement de ce cachet en temps voulu lorsque les procédés cryptographiques utilisés en Suisse ne paraissent plus assez sûrs, et fait alors remonter la preuve d'inaltération au moment du premier cachet.
- P10. **Responsabilité individuelle** : le service d'archivage permet au chercheur, sans qu'il ait à donner de justification, d'effacer les données qu'il aura déposées, ou de les remplacer par une nouvelle version. Le service ne conserve pas de copie des données effacées, mais dresse un protocole de cet effacement qu'il conserve.

---

<sup>30</sup> <http://www.ietf.org/rfc/rfc4810.txt>

<sup>31</sup> <http://www.ietf.org/html.charters/ltans-charter.html>

<sup>32</sup> <http://www.switch.ch/aai/>

<sup>33</sup> [http://www.admin.ch/ch/d/sr/c943\\_03.html](http://www.admin.ch/ch/d/sr/c943_03.html)



- P11. **Pas de transfert de droits** : le service d'archivage n'endosse aucun droit de propriété ou d'utilisation sur les données déposées. Tous les droits de propriété, d'utilisation, d'auteur, de copyright, etc., pour autant qu'il y en ait, restent acquis au chercheur ou à l'institution qui a financé la recherche. Le dépôt s'effectue en quelque sorte à titre „fiduciaire“<sup>34</sup>.
- P12. **Réserve de responsabilité** : le service d'archivage remplit sa mission au plus près de sa conscience en se conformant à des standards exigeants, mais, en cas de perte des données déposées, il ne peut être tenu responsable des dommages occasionnés à des usagers du service ou à d'autres tiers. De même il ne peut être tenu responsable du contenu des données elles-mêmes.
- P13. **Signalement** : chaque dépôt est documenté par le chercheur sur la base d'un formulaire standard, qui, en plus d'une description et d'une documentation du contenu des données, contient aussi d'autres informations de type administratif (coordonnées de contact, propriété des données, délai de conservation, droits d'auteur et droits d'usage, restrictions d'accès, etc.)
- P14. **Référencement** : le service d'archivage permet au chercheur, sitôt après le dépôt, de faire référence (citation) aux données qu'il aura déposées dans des publications scientifiques à l'aide d'un Digital Object Identifiers (DOI)<sup>35</sup>, ce qui permet aussitôt à des tiers d'avoir accès via Internet à une description bibliographique (selon P13) et de trouver les coordonnées d'une personne de contact pour le projet en question.
- P15. **Garantie d'accès** : le service d'archivage permet au chercheur, à tout moment durant le délai de conservation, de recevoir une copie des données qu'il a déposées, soit via Internet, soit sur un support externe selon la quantité de données en jeu.
- P16. **Disponibilité pour l'institution** : la règle veut que les données appartiennent à l'institution qui a financé la recherche. Le service d'archivage garantit en bonne et due forme que les institutions affiliées (voir P3) peuvent obtenir en tout temps, et aussi longtemps qu'elles sont désignées comme propriétaires de ces données dans le dépôt, un inventaire des données qui les concernent et se voir accorder, à des conditions bien précises, un accès en lecture à celles-ci (par exemple lorsque le chercheur qui les a déposées n'est plus atteignable).
- P17. **Exploitation secondaire** : le service d'archivage permet aux chercheurs d'accorder à un tiers un accès ponctuel en lecture aux données qu'ils auront déposées (soit, selon la quantité de données en jeu, par l'attribution d'un mot de passe Internet „one-time“, soit en demandant par écrit au service de livrer les données à la personne concernée).
- P18. **Pas de contrainte en faveur de l'Open Access** : le service d'archivage n'exerce aucune discrimination à l'égard des chercheurs qui ne veulent ou ne peuvent diffuser

---

<sup>34</sup> Ceci peut avoir des conséquences importantes, par ex. dans le cas d'une procédure disciplinaire entamée par une institution contre un chercheur pour infraction à la „bonne pratique de la science“. Le service d'archivage ne remet les données qu'avec l'approbation du propriétaire effectif des données (qui dans la majeure partie des cas est l'institution elle-même qui a engagé la procédure).

<sup>35</sup> On peut recourir ici aux solutions du projet conduit par l'ETH «DOI-CH» (<http://www.e-lib.ch/doi.html>), qui édifie jusqu'en 2011, à l'intention de toutes les universités suisses et de leurs bibliothèques, une infrastructure centralisée pour l'attribution, l'enregistrement et la gestion de Digital Object Identifiers (DOIs) pour les données primaires et secondaires.

leurs données librement selon les principes de l'Open Access (par exemple lorsque des droits d'auteur de tiers qui ont participé au financement avec des fonds privés sont en jeu). Mais le service leur permet aussi de rendre leurs données accessibles pour des tiers par un téléchargement en ligne ou un envoi sur support amovible, de manière totalement ouverte ou sélective (c.à.d. avec mot de passe pour certaines personnes seulement) (voir P17).

- P19. **Catalogage sur demande** : le service d'archivage laisse au chercheur le choix de décider s'ils veulent qu'une description bibliographique des données qu'il aura déposées (description déjà utilisée pour le référencement, voir P14 ci-dessus) apparaisse dans un catalogue en ligne. Un tel catalogue est géré par le service d'archivage, permet des accès filtrés par domaine ou par institution, et contient des mots-clés pour faciliter la recherche. Un accès en lecture aux données référencées par une entrée catalographique n'est possible que lorsque le chercheur a libéré cet accès au moment du dépôt de ses données (voir P18).
- P20. **Délai de conservation maximum** : le service d'archivage laisse le chercheur ou son institut définir lui-même le délai durant lequel le service d'archivage doit conserver ses données, en lui laissant à tout moment la possibilité de le prolonger ou de le raccourcir, ceci sans excéder une période de 15 ans. Selon préavis convenu, le service avertit par lettre recommandée ou par courriel la personne de contact désignée de la prochaine expiration du délai en la priant de se déterminer sur les options possibles.
- P21. **Effacement automatique** : le service d'archivage se charge d'effacer automatiquement les données après expiration du délai de conservation, pour autant que la personne de contact susmentionnée n'ait pas décidé d'une prolongation. L'effacement des données est consigné dans un protocole, et la description bibliographique (P14) ainsi que l'éventuelle entrée catalographique (P19) sont maintenues.
- P22. **Compétence d'évaluation** : la décision définitive quant à une prolongation du délai de conservation ou au passage à un archivage de longue durée appartient à l'institut qui aura mandaté le dépôt des données et qui aura été informé comme indiqué aux P20 et P21.
- P23. **Archivage pérenne** : le passage à un archivage à long terme ne survient qu'après une expertise rigoureuse (P6 et P22). S'il s'avère que les données doivent être archivées durablement par le service d'archivage, cela est consigné par le service, qui supprime de ce fait pour le chercheur l'accès aux données dont il n'a désormais plus la responsabilité pour transférer le tout à son institut d'attache. Celui-ci se détermine alors sur une éventuelle conversion des données dans un format qui permette une conservation et une relecture à long terme, ou sur un transfert des données à une archive d'Etat. Au gré de l'institut, le centre d'archive peut limiter son intervention à une fonction de conseil ou se charger de la réalisation pratique des travaux, qui sont facturés au prix coûtant. En revanche, l'archivage pérenne qu'assure le service n'est pas facturé, ce qui a pour corollaire qu'il n'est alors plus tenu de garantir que leur format et la documentation qui les accompagne font que les données restent intelligibles à long terme.
- P24. **Réseau de compétences** : pour atteindre tous ces objectifs, le service d'archivage établit un réseau de compétences composé de personnes issues des mondes de la re-

cherche et des archives. Il s'agit d'une part de personnes qui soient à même d'évaluer l'importance des données déposées pour une exploitation secondaire (selon P6) et qui, dans la mesure du possible, sont mandatées par les organes de la recherche. Et d'autre part de personnes qui ont déjà de l'expérience dans la conservation ou l'archivage à long terme de données primaires, ou qui veulent l'acquérir pour se mettre au service du centre d'archivage en tant qu'expert conseil (selon P7 et P23).

Bien que le modèle que nous proposons ici vise d'abord à permettre au chercheur, via un service d'archivage centralisé, de satisfaire de manière aussi simple et efficace que possible aux directives sur la conservation des données aux fins de l'„intégrité de la recherche“, et ceci sur une base volontaire et en faisant appel à la responsabilité individuelle, il offre néanmoins des points de contact et de synergie avec d'autres types d'attentes. Citons parmi ces dernières le référencement des données, une exploitation simplifiée de celles-ci, l'identification et la conservation de données qui gardent une valeur importante, ainsi que la coopération potentielle avec les archives de la Confédération et des cantons.

#### 5.1.3.1 Coûts et financement

Nous considérons comme relativement modiques les coûts de réalisation d'un modèle établi sur les principes énumérés au point 5.1.2. Les coûts d'exploitation devraient s'avérer plutôt modestes une fois réalisés les investissements pour la spécification, l'acquisition ou le développement du logiciel nécessaire et la mise en place de l'infrastructure. Un service d'archivage qui répond aux points P1 – P24 peut être en grande partie automatisé sans recourir à du personnel scientifique ou spécialisé en archivistique, et le réseau de compétence (P24) fonctionnerait selon le principe de milice.

Sur le long terme, ce sont pour l'essentiel des coûts d'exploitation pour l'infrastructure matérielle et logicielle ainsi que pour la maintenance des capacités (croissantes) de stockage qui subsisteront. Une enquête conduite en 2004 sur la situation en Allemagne [19] indique un volume global de 100 GB par année pour les branches de sciences humaines (langues, sociologie), et, pour la quantité totale de données primaires importantes à conserver pour les sciences naturelles et autres domaines, un volume de 1000 – 2000 TB par an „avec une forte tendance à la hausse“.

Il n'existe pas d'enquête analogue pour la Suisse. En tenant compte de l'ancienneté de l'enquête allemande, la règle du pouce permet d'évaluer à environ 50-300 TB par année la capacité nécessaire pour la conservation des données primaires importantes, sans retenir certains „Data Hot Spots“ comme l'accélérateur du PSI ou certaines installations de traitement numérique d'images.

S'ouvre ici l'opportunité de gagner comme partenaires et protagonistes trois centres de calcul (par exemple ceux de l'ETHZ et du PSI ainsi que le Swiss National Supercomputing Centre), qui mettraient d'abord à disposition la capacité de stockage nécessaire, et qui seraient ensuite rassemblés en un réseau de stockage administré de manière centralisée<sup>36</sup>, dont SWITCH comme autre partenaire pourrait assurer l'interconnexion.

---

<sup>36</sup> Il existe pour cela des solutions standard, dont certaines gratuites pour les établissements d'enseignement, comme par ex. le «Storage Resource Broker SRB» du San Diego Supercomputer Center ([http://www.sdsc.edu/srb/index.php/Main\\_Page](http://www.sdsc.edu/srb/index.php/Main_Page))

Le financement d'un modèle qui répond aux points P1 – P24 comprend des frais d'investissements (étude de concept, projet pilote, acquisition ou développement de logiciel, réalisation, mise en œuvre) et des coûts d'exploitation récurrents. Nous évaluons les coûts d'investissement à environ 3 millions de francs, sans compter l'implémentation de l'infrastructure initiale de stockage et de communication qui devrait être mise à disposition gratuitement par les quatre partenaires susmentionnés. Pour nous, ce montant devrait être vu comme un crédit unique lié à un projet alloué par la Confédération et les cantons, par les organes d'encouragement de la recherche ainsi que par des fondations et, si possible, des entreprises privées de la branche IT (sponsoring).

Pour la couverture des frais courants – les prestations du centre d'archivage doivent, d'après P2, rester par principe sans charge pour les chercheurs ou les projets – ce sont à notre sens les organes d'encouragement et les universités, c.à.d. les cantons universitaires, ainsi que la Confédération<sup>37</sup> pour le domaine des ETH, qui devraient intervenir. La clé de répartition pourrait prendre comme critère la statistique des quantités de données déposées par les chercheurs des différentes institutions, ou des différents domaines de recherche.

Comme déjà mentionné, les coûts d'exploitation externalisés devraient rester modiques, puisque notamment les frais de stockage et de communication seraient directement pris en charge par les partenaires universitaires, sans rémunération.

#### 5.1.4 Explications

Le modèle tel que défini aux points P1 – P24 souffre toutefois de quelques limitations importantes. Tout d'abord, il s'agit d'un pur *depository*, qu'on pourrait appeler „Swiss Scientific Data Deposit“, et pas d'un *repository*. *Depository* signifie littéralement „étagère à livres“ ou „armoire à dossiers“, alors que *repository* signifie „endroit où l'on conserve“ ou „lieu de dépôt“.

Un *repository* est conçu de telle sorte que les données puissent être stockées thématiquement en fonction de la discipline concernée et de leur fréquence d'utilisation, que l'on puisse rechercher d'après leur contenu et qu'on puisse combiner plusieurs ensembles de données dans un même dépouillement. A l'opposé, un *depository* met l'accent sur la conservation/archivage et la disponibilité à long terme, sur une utilisation occasionnelle et globale des données, ainsi que sur un classement/signalement plutôt homogène et de caractère général. Il faut d'ailleurs d'abord exporter du *depository* les lots de données voulus avant de pouvoir y effectuer des recherches, de les combiner ou de les exploiter avec des outils auxiliaires.

L'«Alliance for Permanent Access to the Records of Science in Europe» (Alliance PARSE, voir chap. 8.4) part du principe, dans sa stratégie [24], tout comme l'«Allianz-Initiative „Digitale Information“»[25] des sociétés scientifiques allemandes (voir chap. 8.5), que les *repository* scientifiques (pour les données primaires ou les données de dépouillement) sont toujours spécifiques à une discipline et sont créés et gérés directement par les différents secteurs de la recherche. Il s'ensuit qu'une stratégie nationale ou internationale doit d'abord pousser à la création de *repo-*

---

<sup>37</sup> Mentionnons au passage que le Conseil fédéral soutient, à hauteur de 0.9 millions CHF/an depuis 2010 et 1.6 million CHF/an depuis 2011, les deux centres de données „Euro-Climhist“ (Universität Bern) et „World Glacier Monitoring Service“ (Universität Zürich), qui sont entre autres aussi responsables de la conservation à long terme de leurs données. (voir à ce sujet la note 12 à la page 21.)

*sitory* de disciplines, à l'établissement de standards d'interopérabilité ainsi qu'à la coordination et à l'interconnexion des *repository* au sein de chaque domaine.

C'est à notre sens une manière correcte de voir les choses, raison pour laquelle ce n'est pas un *repository* que nous proposons, car il ne pourrait qu'être dédié à une seule discipline. Un *repository* unique pour toutes les branches signifierait qu'on cherche à tenir compte des exigences spécifiques de chaque discipline, ce qui d'abord serait techniquement irréaliste, et ensuite supposerait qu'on ait rassemblé au sein du service d'archivage l'ensemble des compétences propres à chaque domaine. Ce serait à notre sens une ambition déraisonnable pour un service qui doit offrir des prestations de base sur un plan national.

A propos du signalement à la demande dans le catalogue général du *depository* (P19), on pourrait imaginer que plusieurs chercheurs, voire un groupe de projet réparti en plusieurs endroits, décident d'utiliser cette fonctionnalité pour créer des *depository* „virtuels“ propres à leur domaine (mais qui ne permettraient pas de rechercher ou d'exploiter le contenu des données elles-mêmes), en employant systématiquement un même identifiant bibliographique. Mais ceci dépasse notre propos.

La seconde limitation du modèle est de nature beaucoup plus sérieuse et plus fondamentale : le principe du *self-deposit* (P5) veut que les chercheurs déposent leurs données dans le format qu'ils auront eux-mêmes choisi (pour les données et pour la documentation). Bien qu'on leur recommande les formats standard les plus appropriés pour une conservation à long terme, et qu'un support conseil leur soit offert (P7), on leur laisse l'initiative sur les deux plans. Ce qui, pour notre part, se justifie pour les raisons suivantes :

- Le devoir de conservation au titre de l'„intégrité de la recherche“ se rapporte aux données primaires (plus précisément, à l'ETH, aux «données originales qui n'ont pas été retravaillées» [14]) qui sont à la base des dépouillements, et donc des résultats publiés. Une conversion des données ou de leur format pourrait résulter en une perte d'information (par ex. type de données) ou une altération, ce qui pourrait avoir des conséquences délicates.
- Vouloir subordonner l'utilisation du service à une conversion préalable des données dans un format standard adapté au long terme exigerait un effort notable de la part du chercheur et diminuerait significativement son désir d'en faire usage.
- Pour beaucoup de disciplines il n'existe pas de format standard adapté à la conservation à long terme<sup>38</sup>.
- Si le service d'archivage se mettait en tête d'exiger plutôt que de recommander des formats standard appropriés, il se verrait contraint de tester la conformité des données reçues avec ceux-ci, ce qui entraînerait un travail considérable.

C'est la raison pour laquelle le point P3 ne prévoit „que“ la conformité avec RFC 4810 «Long-Term Archive Service Requirements» (2007, [23]), et pas avec ISO 14721:2003 «Open Archival

---

<sup>38</sup> Une enquête menée en Allemagne[4] a montré que, sur l'ensemble des données produites, 97,8% étaient stockées dans un format binaire propre à la branche, 0,3% en XML et environ 1,9% dans des formats textes reconnus. Il est vrai que la plupart des formats binaires propres à une branche sont généralement bien documentés. Il n'en reste pas moins qu'une éventuelle conversion des données binaires dans des formats standard – comme cela est usuel pour la conservation à long terme de documentations textuelles – reste illusoire.

Information System Reference Model» (OAIS) [27]. En effet, le standard OAIS exige que quatre critères fondamentaux („OAIS Responsibilities“) soient respectés pour qu’une solution puisse être déclarée conforme OAIS [27] :

- Au moment de la prise en charge des données, un service d’archivage OAIS endosse la pleine responsabilité en ce qui concerne leur disponibilité et leur utilisation à long terme pour ses clients. Cela inclut une conversion ou une migration périodique des informations dans de nouveaux formats lorsque les formats originaux menacent de devenir obsolètes. Le service d’archivage OAIS doit par conséquent être nanti du droit de modifier les caractéristiques structurelles des données de son propre gré.
- Le service d’archivage OAIS doit obtenir sur les données un droit de contrôle technique et organisationnel minimal qui lui permette de garantir le succès d’un archivage pérenne.
- Le service d’archivage OAIS doit être en mesure de pouvoir comprendre par lui-même les données déposées, même en l’absence des déposants initiaux.
- Le service d’archivage OAIS doit pouvoir négocier avec les déposants les critères de la prise en charge des données (y compris le format et la structuration des données ou de la documentation).

Si on se cantonne aux points P1 – P24, le service d’archivage ne peut répondre à aucune de ces exigences, et en fait on ne peut pas s’attendre à ce qu’elles soient effectivement remplies dans la réalité tant qu’on vise un service facile d’emploi, efficace pour les besoins des chercheurs, et qui réponde d’abord aux impératifs de la conservation aux fins de l’„intégrité de la recherche“.

### 5.1.5 Elargissement vers OAIS des objectifs opérationnels

A propos des limitations qui viennent d’être discutées - *depository* au lieu de *repository* et manque de conformité OAIS – le modèle peut être élargi dans une certaine mesure, ce qui toutefois ne nous semble valoir la peine que pour les données à conserver selon ce qui est évoqué au point P23 :

- P25. Responsabilité à long terme :** le service d’archivage peut, en plus de la „bitstream preservation“ prévue au point P8 et d’une activité de pur soutien telle que mentionnée au point P23, assumer une responsabilité plus ample pour les données dont on aura jugé qu’elles méritent d’être conservées durablement (P22). Ceci sur mandat de l’institut de recherche concerné, ou d’une archive d’Etat. Le service d’archivage et l’institut s’accordent alors sur une convention d’archivage, et préparent ensemble les données dans un format adéquat pour l’archivage de longue durée. Faute d’entente, le service d’archivage peut décliner la prise en charge des données. Ce genre de conventions n’est pas possible pour des chercheurs individuels ou des groupes de projet.
- P26. Participation aux frais :** l’institution déposante doit s’acquitter d’un montant proportionnel aux frais engendrés par la préparation des données ainsi que d’une contribution annuelle pour couvrir les frais de garde des données archivées au sens de P25 (et seulement pour celles-là). Si la contribution annuelle vient à tomber, l’institution doit reprendre ses données.

- P27. **«Preservation levels»** : le service d’archivage peut moduler son offre selon P25 en définissant plusieurs niveaux de prestation (Service Level Agreement ou Preservation Levels). Ceux-ci précisent les exigences à satisfaire de part et d’autre en ce qui concerne les modalités de conservation des données.
- P28. **Conformité OAIS** : les composantes organisationnelles et techniques du service d’archivage nécessaires pour remplir la mission selon P25 doivent satisfaire aux principes et exigences du standard international ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27] ainsi qu’au «nestor Kriterienkatalog vertrauenswürdige digitale Langzeitarchive»<sup>39</sup> [28] ou aux «Trustworthy Repositories Audit & Certification: Criteria and Checklist»<sup>40</sup> [29] des «Center for Research Libraries» (CRL). Dans la mesure du possible, le service d’archivage s’efforcera d’obtenir une certification dans ce domaine.
- P29. **Compétence professionnelle** : le service d’archivage est à même de représenter la Suisse dans les milieux internationaux qui œuvrent en faveur de l’archivage à long terme des données scientifiques primaires, de s’associer aux recherches en cours et de proposer des projets de son crû.
- P30. **Transfert de connaissances** : le service d’archivage veillera à promouvoir des transferts de connaissances avec les chercheurs de toutes disciplines qui développent ou entretiennent en Suisse leurs propres centres d’archivage, et à leur rappeler les critères d’exigence liés à l’archivage à long terme.

#### 5.1.5.1 Coûts et financement

Les prestations mentionnées sous P25 – P30 exigent sensiblement plus de personnel et de compétences professionnelles que les prestations évoquées sous P1 – P24 qui touchent essentiellement l’infrastructure. Il faut absolument que cette responsabilité endossée par le service d’archivage pour garantir un archivage pérenne optimal (P25) soit couverte financièrement sur la durée par une contribution régulière du déposant, sans quoi la mission en question risquerait d’être prise à la légère et l’évolution des coûts de la prestation dans le temps ne pourrait être contrôlée.

Il n’est pas possible de proposer une estimation des coûts pour l’élargissement des points P25 – P30 sans définir précisément au préalable les *preservation levels* de P27 et sans savoir quelle serait la demande pour un tel „archivage complet“.

#### 5.1.6 Vue d’ensemble

La figure 1 présente l’éventail des prestations du modèle selon les objectifs opérationnels décrits aux points 5.1.3 et 5.1.5.

<sup>39</sup> <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>

<sup>40</sup> <http://www.crl.edu/PDF/trac.pdf>



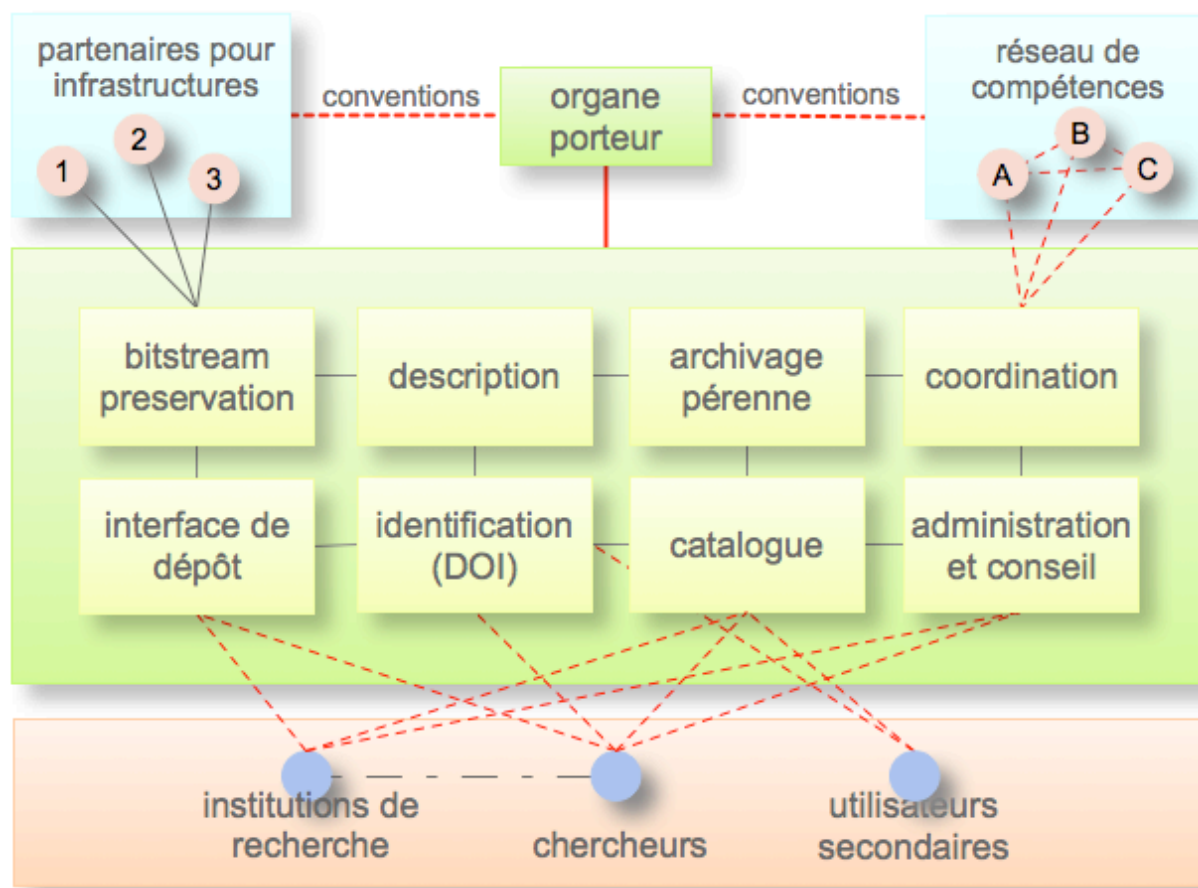


Figure 1 : éventail des prestations du modèle "Conservation avec autocontrôle"

- **Organe porteur** : sa mission est de définir et de contrôler le mandat de prestations, de garantir le financement du service d'archivage, et de conclure des conventions avec les partenaires pour les infrastructures et les experts du réseau de compétences. L'organe de tutelle doit avoir une base solide qui lui garantisse une certaine pérennité. Pour l'organisation formelle de cet organe, on peut penser par exemple à une fondation rattachée à la Confédération, aux universités ou aux organes d'encouragement de la recherche, ce qui lui garantirait un capital de confiance suffisant pour asseoir ses prestations dans la durée.
- **Partenaires pour l'infrastructure** : ce sont eux qui mettent à disposition les ressources en capacité de stockage et en télécommunication, gratuitement ou contre dédommagement.
- **«Bitstream preservation»** : gestion du réseau de serveurs et du système de stockage, stockage à long terme sécurisé, contrôle de qualité et suivi régulier de la validité des cachets temporels électroniques certifiés garantissant l'inaltération des données.
- **Interface pour le dépôt** : fonctionnalités pour la transmission des données et la gestion courante du dépôt.



- **Description** : description standardisée du dépôt de données, tant pour le contenu (par ex. via DublinCore) que pour les aspects administratifs (durée de conservation, propriété des données, droits d'auteur et droits d'exploitation, conditions d'accès, etc.)
- **Identification (DOI)** : génération de Digital Object Identifiers (DOI) pour l'identification précise et le référencement des données dans les publications. On peut reprendre à cet effet à partir de 2009 les solutions du projet «DOI-CH»<sup>41</sup>. C'est grâce à un service *ad hoc* de résolution des DOI que le service d'archivage peut donner à des tiers un accès Internet aux données publiques du dépôt.
- **Catalogue** : catalogue public dans lequel les chercheurs peuvent publier des entrées descriptives signalant leurs dépôts et, le cas échéant, offrir l'accès à certaines données.
- **Réseau de compétences** : personnes qui, dans leur domaine de recherche, disposent de l'expérience et des connaissances liées à la conservation à long terme des données scientifiques, et qui peuvent de ce fait conseiller d'autres chercheurs. Mais aussi celles qui sont à même d'évaluer l'importance à long terme des données primaires dans leur discipline et de réaliser des expertises dans ce sens.
- **Coordination** : coordination du réseau de compétences ; recherche, «fidélisation» et motivation des spécialistes du réseau de compétences ; mise à disposition d'experts pour les chercheurs en quête de conseils, ou pour les institutions qui désirent faire évaluer la valeur à long terme de leurs données ; organisation de cours de formation continue ; préparation d'un portail dédié à l'archivage pérenne, etc.
- **Administration et conseil** : contacts avec les déposants ; création et administration des dépôts ; conseils aux chercheurs sur les questions essentielles qui touchent à l'archivage à long terme.
- **Archivage pérenne** : réception et préparation des dépôts de données primaires de haute valeur. Collaboration avec les instituts de recherche, propriétaires des données, chercheurs et experts du réseau de compétences. Donc responsabilité entière de l'archivage pérenne des données.

La figure 2 qui suit illustre le déroulement procédural dans le modèle, conformément aux objectifs opérationnels décrits aux sections 5.1.3 et 5.1.5. On distingue quatre cas d'utilisation :

- **1. Dépôt (rouge)** : le chercheur (ou chef du projet) responsable des données sollicite, par l'entremise de son institut, la création d'un nouveau dépôt. Il complète un formulaire standardisé qui signale le dépôt en mentionnant la nature des données et leur provenance. Pour la préparation des données en vue de la conservation, le recours aux conseils d'un expert du réseau de compétences, voire à son assistance, peuvent être sollicités si le chercheur ou l'institut le désire. Le chercheur transmet alors les données qu'il a préparées (avec leur documentation) au service d'archivage (via Internet ou support amovible) et fixe le délai de conservation ainsi que les diverses modalités de communication (licence d'utilisation, droits d'auteur éventuels, etc.) et d'administration du dépôt. Le service

<sup>41</sup> Le projet «DOI-CH» (<http://www.e-lib.ch/doi.html>) développé par l'ETH entend créer, à l'intention de toutes les universités de Suisse et de leurs bibliothèques, une infrastructure centralisée pour l'attribution, l'enregistrement et la gestion de DOI pour les données primaires et secondaires.

d'archivage stocke alors les données dans l'archive de longue durée, les dote d'un cachet temporel certifié, leur attribue un Digital Object Identifier (DOI) pour le référencement et, sur demande, insère une description catalographique dans un catalogue en accès public.

- **2. Accès aux données** (*vert*) : par l'entremise d'un DOI cité dans une publication ou suite à une recherche en catalogue, quelqu'un aimerait accéder aux données pour une exploitation secondaire. Ceci peut se faire sans autre (par téléchargement en ligne ou envoi sur support externe) pour autant que le chercheur ait placé ses données en accès libre. Dans le cas contraire, le chercheur reçoit une demande par courriel (ou selon tout autre moyen spécifié dans l'entrée signalétique) et peut autoriser l'accès aux données par attribution d'un mot de passe valable une seule fois, ou en donnant mandat au service d'envoyer les données sur support amovible.
- **3. Inventaire de l'institution** (*vert*) : l'institution partenaire du service d'archivage peut en tout temps avoir accès à l'inventaire, c'est-à-dire aux entrées descriptives et aux informations sur l'état de ses dépôts pris en charge par le service (voir cas 1 ci-dessus). Elle peut également avoir accès aux données si elle est mentionnée comme étant le propriétaire.
- **4. Echéance du délai de conservation** (*bleu*) : Avant expiration du délai de conservation prévu, l'institution propriétaire des données décide, après entente avec le dépositaire, si les données doivent être détruites, s'il faut prolonger le dépôt ou si les données doivent être archivées durablement. Elle peut à cet effet faire évaluer l'intérêt des données pour une exploitation ultérieure par un expert du réseau de compétences. Si l'archivage à caractère définitif est choisi, le service d'archivage prépare alors les données pour un tel archivage, aux frais de l'institution propriétaire et avec l'appui du propriétaire du dépôt et de l'expert du réseau de compétences, et endosse enfin la responsabilité pour l'archivage pérenne.

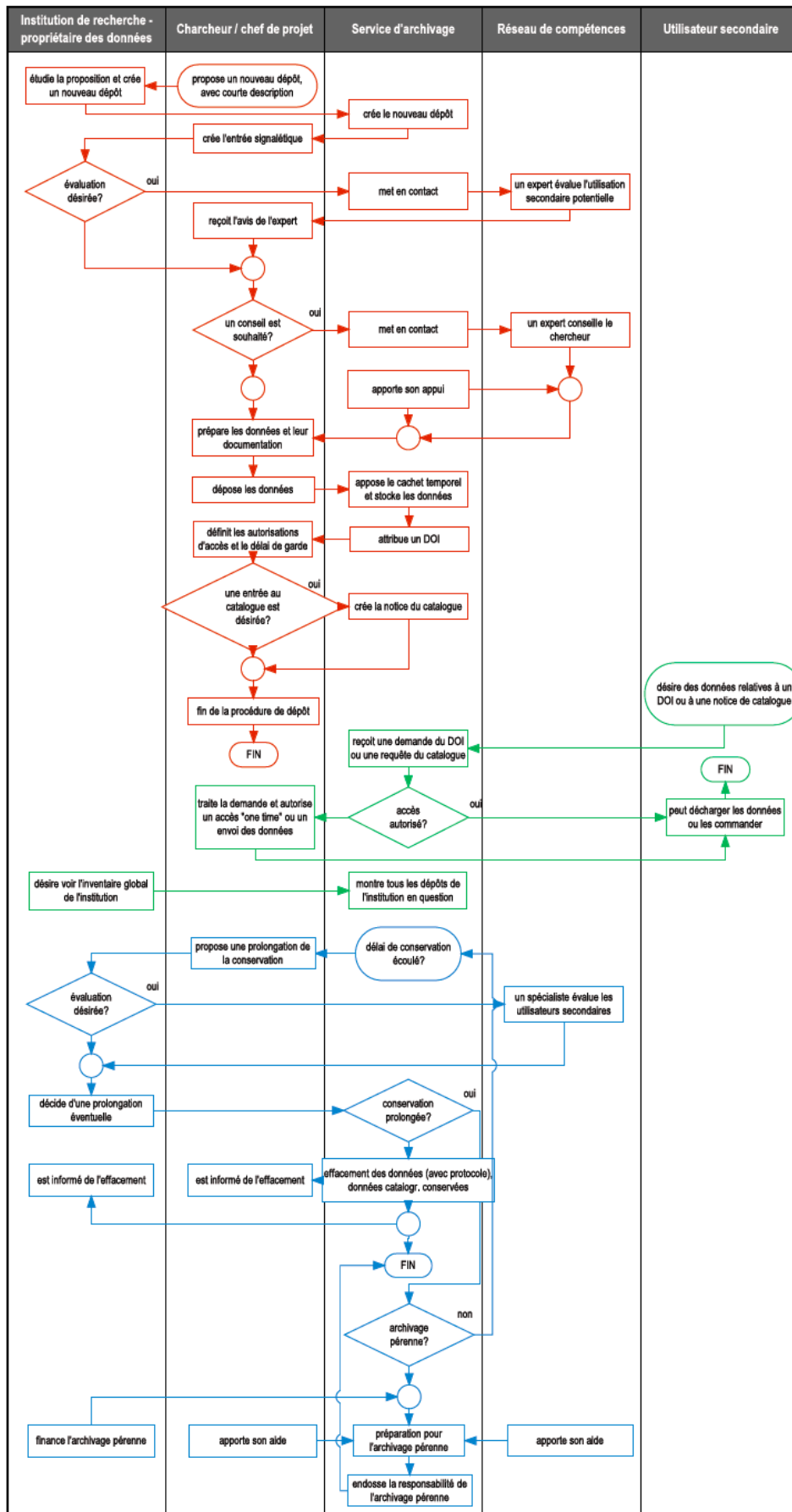


Figure 1 Déroulement procédural du modèle „Conservation avec autocontrôle“

## 5.2 Modèle «Archivage avec obligation de dépôt»

Au contraire du premier (voir chap. 5.1, Modèle «Conservation avec autocontrôle»), ce second modèle pour les données primaires fait appel à une approche de type "contrôle", qui s'appuie sur le devoir de dépôt généralisé pour la documentation scientifique exigé par les législations gouvernementales sur les archives et s'effectue via un service national d'archivage pour les données primaires. Dans ce modèle, toutes les données accompagnées de leur documentation sont systématiquement soumises à une évaluation quant à leur intérêt pour un archivage.

Le devoir de dépôt systématique ainsi que l'évaluation du point de vue archivistique prennent naissance au moment où les projets de recherche sont évalués : soit au moment de l'approbation et du financement du projet, ou au moment de l'approbation et de l'évaluation des résultats, ou encore au moment du *reporting*, c.à.d. de la présentation du rapport final. On attend des experts respectifs qu'ils évaluent la pertinence d'archiver (pour un délai limité ou de manière durable) en adoptant le point de vue de la discipline concernée. On demande aux chercheurs également d'apprécier la valeur à long terme des données qu'ils produisent, soit au moment de la soumission du projet, soit dans le rapport final.

Les propositions qui en résultent, de même que les éventuelles obligations d'archiver que le Fonds national aura pu imposer au moment du financement, sont systématiquement transmises au service d'archivage compétent, qui prononce définitivement après en avoir débattu avec les experts et les chercheurs. Certaines catégories de projets, comme par exemple les travaux d'étudiants jusqu'aux niveaux diplôme ou *master*, peuvent être dispensés du devoir d'archiver dans l'attente d'une évaluation ultérieure.

Les principaux critères qui plaident pour un archivage durable sont le potentiel que recèlent les données primaires pour une exploitation secondaire de même que le fait que l'intelligibilité des résultats doit être garantie, notamment lorsque la recherche a porté sur un domaine délicat du point de vue éthique ou politique (par ex. recherches sur l'humain), lorsqu'il y a un risque important que la recherche ait des conséquences décisives sur l'homme et son environnement (par ex. les disséminations expérimentales), ou lorsque les résultats peuvent influencer les processus de décision du législateur ou du politique. On décide dans ces cas-là d'archiver et de remettre aux archives compétentes toute la documentation qui se rapporte au projet.

Pour les données primaires qui n'auront pas été retenues comme candidates à un archivage durable par les services d'archives compétents, le service national offre néanmoins déjà une solution avec le premier modèle (chap. 5.1).

A côté du devoir de proposer et de l'évaluation obligatoire de la valeur archivistique des données, la préparation systématique des données en prévision de l'archivage présente une deuxième différence essentielle avec le premier modèle (chap. 5.1) : l'archivage au titre de la législation sur les archives se fait sur une base durable et doit satisfaire aux exigences des archives d'Etat en matière d'exploitation, de classement et de signalement, car ces services sont tenus, dans l'accomplissement de leur mission, de veiller à un archivage professionnel et irréprochable.

Ceci fait qu'on ne peut plus laisser au chercheur seul le soin de décider du format des données ou de la manière de les décrire, comme dans le premier modèle. Bien plus, il faut s'assurer que les formats employés sont compatibles avec un archivage durable, de façon à pouvoir être encore lus et compris après un laps de temps appréciable, ou, si nécessaire, convertis le moment venu dans de nouveaux formats. Cela signifie entre autres que ces formats doivent être entièrement documentés de manière transparente et ne pas être dépendants d'un logiciel particulier que les circonstances pourraient amener à ne plus être disponible après quelques années.

Il y a peu de chances que les données primaires répondent d'emblée à ces critères, et il faudra par conséquent les préparer pour un archivage durable et les convertir dans des formats standard ouverts. Ce sera probablement surtout pour des formats binaires complexes qu'une telle conversion conduira à des altérations des données originales ou à des pertes d'information (précision, type de données, structure, etc.), qui paraîtront certes acceptables pour l'archiviste, mais qui empêcheront de reproduire les résultats qui avaient été obtenus à partir des données exploitées dans leur état original.

Une enquête menée en Allemagne [4] a montré que, sur l'ensemble des données produites, 97,8% étaient stockées dans un format binaire propre à la branche, 0,3% en XML et environ 1,9% dans des formats textes reconnus. Un sondage de l'ETH Zurich [3] a révélé que 49% des interrogés estiment que, pour leur domaine, les données primaires qu'il faudrait archiver ne sont pas dans un format standard, mais «dans un format reconnu et bien documenté dans leur branche ou dans un cercle plus large».

La multiplicité des formats de données sous-jacents aux données primaires – formats liés à des logiciels propriétaires, formats spécifiques à des instruments de mesure ou formats conçus par les chercheurs eux-mêmes – est déjà beaucoup plus considérable à l'intérieur d'un même champ de recherche que dans l'ensemble générique des formats.

Il n'est donc pas possible, ni pour le service national d'archivage, ni pour les archives d'Etat, d'identifier et de tester pour chaque discipline les formats compatibles avec un archivage durable ni de définir des directives pour la conversion des multiples formats propriétaires en formats standard.

C'est la raison pour laquelle le modèle prévoit que les organes de support du service d'archivage, en collaboration avec les académies scientifiques des différentes branches, recrute et mandate des commissions *ad hoc* qui ont pour mandat de d'identifier et de définir quels sont les formats compatibles avec l'archivage. Les formats ainsi retenus par les commissions spécialisées sont alors reconnus par les services officiels et deviennent des prérequis obligatoires pour la préparation des données primaires à archiver.

Les données primaires dignes de conservation sont prises en charge et conservées définitivement par le service d'archivage national, sur mandat des différentes archives d'Etat et en collaboration avec les chercheurs, les instituts et les archives officielles compétentes, après avoir été préparées selon les standard des commissions spécialisées.

En fonction des dispositions légales, les coûts entraînés par la préparation des données sont pris en charge soit par l'instance soumise à l'obligation de déposer (l'institut de recherche), soit par l'archive d'Etat. S'il est déjà certain, au moment de l'approbation d'un projet, que les données

devront être archivées durablement, par exemple lorsque le Fonds national en fait une condition du subventionnement, ces coûts doivent alors être portés au budget du projet de recherche.

Si, au moment de leur préparation, des données primaires devaient être transformées (conversions) de façon telle qu'il ne soit plus possible pour les chercheurs de remplir désormais leur devoir de déposer au sens de l'„intégrité de la recherche“ (voir chap. 3.1.3), il faudrait alors les conserver parallèlement dans leur format original selon les modalités du premier modèle (voir chap. 5.1).

Pour des données primaires qui doivent être conservées à long terme, ce modèle (comme déjà l'élargissement du premier modèle dans la section 5.1.5) suit la norme ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27]. La responsabilité globale, sur les plans professionnel et technique, d'un archivage à long terme réellement garanti pour les données primaires qui doivent en faire l'objet, incombe à la fois au service national d'archivage et aux commissions d'archivage spécialisées de chaque discipline qui apportent le *knowhow* propre à leur domaine.

### 5.2.1 Objectifs stratégiques

Le modèle représenté dans ce sous-chapitre vise les objectifs stratégiques suivants :

- H1. **Systématique** : le service d'archivage repose sur la mise en œuvre systématique du devoir de proposer et de l'évaluation archivistique, tels que mentionnés dans les lois et ordonnances respectives de la Confédération et des cantons. Le contrôle du respect du devoir de proposer ne doit cependant pas intervenir de manière centralisée, mais s'effectuer localement, c'est-à-dire au niveau où s'exerce juridiquement le droit de disposition de chaque institut.
- H2. **National** : le service d'archivage doit, par principe, être à disposition de tous les services concernés par les législations sur l'archivage ainsi que des hautes écoles et instituts de recherche de la Confédération et des cantons.
- H3. **Centralisé** : à l'exception du devoir de proposer et de l'évaluation archivistique (voir H1), le service d'archivage doit être conçu comme centralisé, tout en étant étroitement connecté à un réseau de commissions d'archivage propres à chaque discipline. On attend d'une telle centralisation qu'elle suscite des synergies en matière d'infrastructure et de savoir-faire.
- H4. **Complémentaire** : le service d'archivage est conçu avant tout pour les données primaires qui doivent être conservées au sens des législations d'Etat sur les archives, et ne concerne par conséquent qu'une partie seulement de toutes les données produites par la recherche. Pour celles qui ne sont pas retenues pour un archivage durable, il faut en plus une solution pour la conservation à 10 ans, selon le modèle présenté au chap. 5.1. Le service central n'est pas censé remplacer des solutions d'archivages<sup>42</sup> qui sont déjà opérationnelles dans certains domaines, pour autant que celles-ci respectent les principes re-

---

<sup>42</sup> On mentionnera comme exemple le Schweizerische Informations- und Daten-Archivdienst für die Sozialwissenschaften (SIDOS).

connus d'un archivage à long terme. Et même dans ces cas-là, les contraintes de l'obligation systématique de proposer et d'évaluer doivent être prises en compte.

- H5. Politique de la recherche :** les investissements réalisés dans la production de données primaires doivent être sauvegardés et celles-ci mises à disposition de tiers pour une exploitation secondaire au fil du temps, pour autant qu'aucun motif important d'embargo ou aucune réserve de propriété ne s'y opposent.

### 5.2.2 Justification

Ce modèle définit de manière systématique les conditions cadre nécessaires à la mise en œuvre des lois sur l'archivage des données primaires issues de la recherche. Pour les données qui, au sens de cette loi, ne méritent pas d'être conservées, on recourt au premier modèle (section 5.1), de façon que les chercheurs aient également par ce biais un moyen de satisfaire réellement aux exigences de la conservation aux fins de l'intégrité scientifique.

Ce modèle répond tout particulièrement aux attentes du public, qui veut que les résultats de la recherche restent intelligibles sur le très long terme pour des thèmes sensibles sur les plans éthique et politique, et lorsque les découvertes peuvent avoir des conséquences importantes pour l'homme et son environnement ou influencent de manière notoire les décisions politiques.

Les objectifs stratégiques découlent également du fait que la constitution des dossiers doit être réalisée de manière ordonnée et systématique conformément aux exigences, et que, pour les données qui ne doivent pas être archivées durablement, il existe néanmoins des possibilités de conservation qui permettent aux chercheurs de satisfaire aux critères des règlements sur l'intégrité scientifique.

En plus des exigences qu'elles posent en matière d'intelligibilité et d'intégrité des travaux de recherche, les instances d'encouragement tiennent aussi à favoriser l'exploitation secondaire de données primaires importantes. Il arrive souvent qu'un lot de données primaires ne soit pas totalement exploité au cours d'un projet, ce qui en fait un matériau de départ privilégié pour une autre recherche. On peut parler ici d'une réelle sauvegarde de l'investissement initial consenti.

### 5.2.3 Objectifs opérationnels

Le modèle proposé ici ne couvre que les étapes classiques de l'archivage, à savoir l'évaluation, la préparation, la conservation et la communication de données primaires d'importance.

- Q1. Institution :** c'est à une entité institutionnelle propre qu'est dévolue la prestation d'archivage, entité administrativement rattachée à une institution du monde des hautes écoles suisses ou de la recherche. Une autorité de surveillance assure le contrôle de la gestion et de la stratégie.
- Q2. Autorité de surveillance :** l'autorité de surveillance regroupe des représentants des hautes écoles, des instituts de recherche, des bibliothèques universitaires, des archives officielles ainsi que des organes d'encouragement de la recherche. Elle est instituée par une convention entre les différentes instances, ou plutôt entre leurs tutelles. Elle délègue



ses compétences à un comité de pilotage qui représente son autorité vis-à-vis du service national d'archivage.

- Q3. **Commissions d'archivage** : l'autorité de surveillance institue, d'entente avec les différentes académies et éventuellement avec d'autres instances, les commissions d'archivage propres à chaque branche qui sont appelées à définir et à maintenir, pour leurs domaines respectifs, les directives et normes d'archivage pérenne des données primaires qui les concernent.
- Q4. **Dispense du devoir de proposer** : les services d'archives attirés pour les hautes écoles et instituts de recherche imposent systématiquement le devoir de proposer dans leur domaine de compétence. Ils le font en s'appuyant, pour les modalités d'exécution, sur les prescriptions légales qui prévalent pour eux. D'entente avec leurs organes de contrôle (archives cantonales, Archives fédérales), ils dispensent les données primaires de certaines catégories de travaux du devoir de dépôt, par exemple celles de tous les travaux d'étudiants jusqu'aux niveaux diplôme ou master, à l'exception de ceux qui font partie de grands projets de recherche et apportent de ce fait une contribution décisive.
- Q5. **Evaluation en mode coopératif** : les archives compétentes organisent le recensement des données primaires dignes d'être conservées à leur niveau de compétence tel que défini par la loi. Le modèle présenté ici ne privilégie pas une procédure particulière. Seul le moment auquel on procède à l'évaluation peut être critique : plus on attendra pour juger de la pertinence des données, plus leur préparation et leur documentation en vue de l'archivage seront difficiles et coûteuses. Dans le cadre de ce modèle, nous partons de l'idée que les projets de recherche font toujours l'objet d'une évaluation et d'une approbation, tant au moment de leur soumission qu'à celui de la publication des résultats. On peut dès lors exiger des experts qu'ils prennent pour règle d'émettre, dans leur rapports d'évaluation respectifs, une proposition fondée quant à la nécessité d'un archivage, proposition qui s'inscrira alors dans le processus de décision du centre d'archives. D'ailleurs, dans les rapports finaux généralement exigés pour les projets importants, on attend que soit d'emblée exposée la solution qui sera retenue pour un archivage conforme aux dispositions légales et aux standards de la commission d'archivage. Ces rapports finaux seront à leur tour transmis aux archives compétentes par le biais d'une procédure établie.<sup>43</sup>
- Q6. **Réévaluation** : on peut fixer un délai à partir duquel le centre d'archivage compétent procède à une réévaluation de la nécessité d'archiver en collaboration avec les délégués de la commission d'archivage concernée. Ceci permet de tenir compte du fait que la valeur des données peut changer au cours du temps en raison du renouvellement des connaissances.
- Q7. **Délégation de la compétence d'évaluer** : afin de mieux tirer parti des synergies et des compétences spécialisées, l'archive d'Etat compétente peut déléguer la responsabilité de l'évaluation archivistique (sous réserve de la fonction de contrôle) aux commissions d'archivage instituées par l'autorité du service national d'archivage.
- Q8. **Délai d'embargo et réserves de propriété** : par principe, les données primaires archivées et leur documentation sont librement utilisables. D'éventuels délais d'embargo ou

---

<sup>43</sup> La plupart des services d'archive des universités considèrent déjà que les rapports finaux de projets de recherche méritent d'être archivés et les transmettent aux services compétents.



réserves de propriété seront annoncés par les responsables de projet au moment de la remise des données au centre national d'archivage. Ils devront prouver que, pour d'importantes raisons personnelles ou autres, les données doivent être interdites d'accès à des tiers durant un certain délai, et qu'elles sont soumises à des réserves de propriété ou à des droits d'utilisation (par ex. lorsque le projet a bénéficié d'un financement complémentaire privé).

- Q9. **Préparation** : la préparation des données (et de leur documentation) en vue d'un archivage pérenne s'effectue d'après les règles et standards des commissions d'archivage respectives propres à chaque discipline. Dans la mesure du possible, elle intervient au moment du bouclage du projet pour bénéficier de la participation des personnes impliquées.
- Q10. **Coûts et budget** : les coûts résultant de la préparation des données (et de leur documentation) sont, selon les dispositions légales régissantes, pris en charge soit par l'instance soumise à l'obligation de déposer (l'institut de recherche), soit par l'archive d'Etat compétente. L'autorité de tutelle du service national d'archivage dédommage cette dernière par une contribution annuelle. (Il faut tenir compte ici du fait que la valeur archivistique des données qui seront produites ne peut être clairement évaluée au moment du financement d'un projet de recherche). Lorsqu'il est évident, au moment de l'approbation d'un projet, que les données devront être archivées durablement, notamment lorsque le Fonds national en fait une condition de subventionnement, les coûts doivent alors être imputés au budget de recherche.
- Q11. **Gestion des frais** : le service national établit pour chaque lot de données à archiver un budget concernant la préparation des données et leur archivage à long terme. Ce budget doit être approuvé par l'instance qui doit légalement supporter les coûts (à savoir l'instance déposante, l'archive d'Etat ou l'autorité de tutelle du service national).
- Q12. **Contrôle des coûts** : lorsque les frais de préparation escomptés pour un projet dépassent les possibilités de financement du service national d'archivage, celui-ci lance une demande de crédit extraordinaire auprès de son autorité de tutelle. Si cette demande est refusée, le service national renonce alors à prendre en charge les données.
- Q13. **Coûts d'exploitation** : après la préparation initiale des données, des frais d'exploitation annuels incombent au service national pour la maintenance régulière, ainsi que des frais à plus long terme pour les migrations destinées à garantir la lisibilité des données. Ces coûts augmentent proportionnellement au nombre de lots de données déposés en archive et en fonction également de l'ancienneté des données. L'autorité de tutelle du service national prévoit par conséquent un financement suffisant pour assurer la garde des données à long terme.
- Q14. **Catalogage et référencement** : les données primaires sont cataloguées par le service national conformément aux règles archivistiques en vigueur et dotées d'un *persistent identifier* (DOI ou URN) qui facilite leur référencement. Le catalogage à des niveaux supérieurs se fait de manière cohérente et standardisée<sup>44</sup>. Les données descriptives sont alors mises à disposition des archives compétentes par le biais d'une interface normalisée pour

---

<sup>44</sup> Par „niveaux supérieurs“ on fait référence ici aux dénominations Fonds, Série et Dossier de l'ISAD(G).

l'importation dans leur système d'interrogation. Il peut arriver que les archives d'Etat doivent procéder à un recatalogage selon leurs pratiques dans leur propre système.

- Q15. **Accès aux données** : l'accès aux données primaires du service national s'effectue au travers des systèmes des archives d'Etat compétentes. Pour cela, celles-ci utilisent les indicateurs (DOI et URN) préparés par le service national pour aiguiller vers celui-ci les usagers de leur propre système, via une interface Web. Le service national dispose d'une interface utilisateur homogène qui permet de visualiser les données et leur description. Le contrôle des accès réservés aux personnes habilitées est du seul ressort des archives d'Etat : seules des identifications techniques, et en aucun cas personnelles, sont utilisées entre les systèmes des archives et le service national qui, par conséquent, ne gère aucun compte personnel à son niveau. Ce n'est qu'aux données qui auront été décrétées libres d'accès par les archives d'Etat que le service national pourra octroyer un accès direct. Vu le caractère fastidieux qu'aurait pour le chercheur le fait de lancer sa recherche sur les différents serveurs des archives d'Etat du pays, le service national devrait envisager comme alternative l'édification d'un portail centralisé d'accès.
- Q16. **Responsabilité** : le service national endosse, avec les commissions d'archivage instituées par son autorité de surveillance, la responsabilité pour la préparation des données et leur archivage pérenne. Le service national peut toutefois décliner cette responsabilité s'il manque des moyens, des compétences ou des autorisations nécessaires. C'est l'autorité de tutelle qui décide de la procédure à suivre en pareille circonstance.
- Q17. **Planification de la conservation** : la responsabilité de la planification de la conservation, notamment les migrations rendues nécessaires avec le temps, est assumée en partage par le service national et les commissions d'archivage de disciplines. En cas de mésentente, l'autorité de surveillance tranche après avoir pris l'avis d'un tiers.

### 5.2.3.1 Coûts et financement

Au contraire du modèle décrit au chap. 5.1, les coûts entraînés par la réalisation et l'exploitation d'un service national d'archivage selon ce modèle-ci ne sont pas faciles à estimer, car on ne peut prédire ni la quantité ni la complexité des données primaires à archiver. La préparation soignée des données primaires à archiver et de leur documentation représente un facteur de coût déterminant dans ce modèle. Un autre facteur de coût important relève du fonctionnement des commissions d'archivage. S'y ajoutent les coûts entraînés par l'organisation et l'institutionnalisation du devoir systématique de déposer et ceux de l'évaluation archivistique au sein des instituts de recherche et des archives d'Etat.

Gardons à l'esprit que si les archives compétentes selon la loi devaient se charger elles-mêmes de l'ensemble des tâches d'archivage des données primaires, on aboutirait au total à des coûts sensiblement plus élevés pour un résultat de bien moindre qualité. Un service national d'archivage qui compte les hautes écoles et autres organismes de la recherche dans son autorité de surveillance, et que celle-ci dote de commissions spécialisées *ad hoc*, est globalement bien mieux à même de remplir cette tâche, et plus efficacement.

Les archives d'Etat se sont peu préoccupées jusqu'ici des données issues de la recherche. Le modèle ici proposé comble une lacune importante non seulement dans l'éventail des prestations

offertes par les archives, mais aussi dans la promotion systématique de la recherche. Une exploitation secondaire effective, susceptible de dépasser le court ou moyen terme, n'est possible que si l'on dispose d'un authentique archivage pérenne tel que prévu par ce modèle.

Ce modèle ne couvre toutefois que l'archivage des données primaires qui auront été désignées par les archives d'Etat comme méritant l'archivage sur la base de leurs législations propres. Il n'est pas proposé ici de solution pour une conservation globale des données primaires qui permettrait aux chercheurs de satisfaire aux impératifs du devoir d'intégrité, qu'il s'agisse des données qui n'auront pas été reconnues comme dignes d'intérêt pour l'archivage pérenne, ou de toutes celles qui l'auront été, mais qui doivent être à tel point transformées qu'elles ne répondent plus aux critères d'inaltération exigés.

## 6 Modèle pour les données secondaires

Avec le projet qui nous occupe, la Conférence des bibliothèques universitaires suisses (CBU) vise «l'élaboration d'un modèle pour un archivage pérenne centralisé des données numériques primaires et secondaires issues de la recherche en Suisse» (intitulé du projet<sup>45</sup>).

Le modèle que nous proposons ici pour les publications (*eJournals*, *eBooks* et bases de données scientifiques) se fonde sur les objectifs et prérequis du projet, sur nos analyses des chapitres 3 (besoins et utilisation), 4 (Protagonistes) et 8 (Contexte européen), ainsi que sur la thématique générale énumérée dans la partie introductive 2.3.

Ce modèle a pour contenu la conservation pérenne des données secondaires. Les défis, les stratégies ainsi que des solutions techniques partielles visant à garantir à long terme l'accès aux publications électroniques ont déjà fait l'objet de travaux en 2005 – 2008 dans le cadre du projet «E-Archiving»<sup>46</sup> du Consortium des bibliothèques universitaires, avec notamment l'introduction d'un serveur de documents distribué.

Ce modèle s'appuie en grande partie sur l'«Allianz-Initiative ,Digitale Information'» préconisée par les sociétés scientifiques allemandes (voir chap. 8.5). Il vise à garantir une disponibilité sur la durée des ressources électroniques, revues (*eJournals*) et livres (*eBooks*), après échéance des licences d'utilisation ou coupure des accès sur les serveurs des éditeurs.

### 6.1 Objectifs stratégiques

Les trois objectifs stratégiques suivants ont été fixés pour ce modèle :

- G7. **National** : le service d'archivage pour les données secondaires doit être au service de toutes les bibliothèques représentées au sein du Consortium<sup>47</sup>, ainsi que des universités et HES.

---

<sup>45</sup> On a évoqué au chap. 2.1 la différence de signification entre les termes *archivage* et *conservation*, resp. *archivage* à long terme et *conservation* à long terme, tels qu'ils apparaissent dans les textes législatifs et dans le vocabulaire des archivistes.

<sup>46</sup> <http://lib.consortium.ch>. Le rapport final n'est pas encore accessible publiquement, raison pour laquelle nous nous référons aux résultats d'une étude préliminaire de 2005 [17].

<sup>47</sup> <http://lib.consortium.ch>. Font partie du Consortium fondé en 2000 toutes les universités cantonales, le domaine des ETH, toutes les hautes écoles spécialisées, plusieurs hautes écoles pédagogiques, la Bibliothèque nationale ainsi que plusieurs partenaires secondaires du secteur public suisse. Le Consortium a été financé à 50% par la Confédération jusqu'en 2005, et est supporté depuis lors en totalité par les bibliothèques membres. Sa tâche principale consiste à négocier les licences pour ressources électroniques acquises par les bibliothèques partenaires.

- G8. **Contributif** : le service d'archivage pour données secondaires (publications) entend garantir la disponibilité à long terme des publications pour la recherche, après l'échéance des licences et la suppression des accès sur les serveurs des éditeurs, et contribue ainsi efficacement à la mission d'approvisionnement des bibliothèques universitaires.
- G9. **Transparent** : par souci de transparence, et en raison des coûts en jeu et de la complexité induite sur les plans organisationnel et technique, le service d'archivage doit être assumé par une seule institution qualifiée, reconnue comme digne de confiance par le « Consortium des bibliothèques universitaires suisses » et de ce fait mandatée et soutenue financièrement à cette fin.

## 6.2 Motivation

La problématique de l'archivage pérenne des publications électroniques est abordé dans les chapitres 3.2 et 8.3. Alors qu'une bibliothèque conserve dans ses magasins les exemplaires papier d'une publication à laquelle elle est abonnée, les versions électroniques ne sont stockées que sur les serveurs des éditeurs pour y être consultées via une connexion Internet. L'accès est de ce fait circonscrit aux postes de travail qui se trouvent à l'intérieur du domaine IP d'une université, alors que l'accès aux versions papier est possible entre universités par le biais du prêt interbibliothèques (plus largement ouvert à tous), qui s'effectue généralement par photocopie ou numérisation de l'article demandé.

Plusieurs raisons peuvent faire qu'un éditeur ne soit plus en mesure de garantir l'accès à des ressources électroniques :

- L'éditeur coupe les accès sitôt que la bibliothèque résilie sa licence d'utilisation.
- L'éditeur ferme boutique ou tombe en faillite.
- L'éditeur arrête la publication d'un titre particulier et n'offre plus l'accès aux anciens numéros parus.
- L'éditeur perd ses données suite à un problème technique ou à une catastrophe.

Cette suppression de la disponibilité en ligne est encore plus pénalisante pour les bibliothèques et les chercheurs si la publication en question n'a été publiée que sous forme électronique, ce qui devient de plus en plus le cas.

Les bibliothèques sont ainsi contraintes, si elles entendent garantir une disponibilité<sup>48</sup> sur le long terme, d'archiver à demeure une copie physique des publications électroniques acquises. Mais il faut pour cela que les éditeurs aient prévu un droit d'archiver dans les contrats de licence, et aient livré leurs ressources numériques accompagnées de leurs métadonnées dans un format qui, d'une part, puisse être avalé sans autre par les serveurs d'archivage des bibliothèques, et d'autre part, soit compatible avec les exigences d'un archivage pérenne. Voilà trois conditions essentielles (droit d'archiver, format d'importation, compatibilité avec les critères d'archivage) qui la plupart du temps de sont pas remplies.

---

<sup>48</sup> La copie d'archive ne peut toutefois offrir tout l'éventail des modalités d'interrogation proposées par un éditeur via Internet. Seuls sont conservés les documents et les données bibliographiques, et le service ne peut offrir que des fonctionnalités identiques pour toutes les publications, que ce soit au niveau des descriptions bibliographiques ou des recherches en *fulltext*.

L'effort technique et logistique nécessaire à l'archivage incombe par ailleurs à chaque bibliothèque, même lorsque le même titre est acquis sous licence par plusieurs établissements et que la version électronique, au contraire de l'édition papier, n'a pas à être confinée à un lieu précis. Et ce sont surtout les petites bibliothèques qui vont manquer des ressources nécessaires pour développer leur propre solution d'archivage des *eJournals*, de sorte que, pour autant qu'un droit d'archiver ait été octroyé et utilisé pour certaines copies numériques, celles-ci risquent d'être stockées dans des formats douteux et sur des supports de données éparpillés qui ne pourront jamais garantir une conservation durable.

Certaines bibliothèques ont aussi entrepris de numériser rétrospectivement les éditions imprimées de titres très utilisés („scanning“). Ces copies électroniques ne peuvent être considérées comme des *eJournals*, mais leur sont assimilées ici en ce qui concerne l'archivage. La principale différence avec les *eJournals* est que la bibliothèque a ici une maîtrise totale quant au choix des formats (pour les données et métadonnées).

La problématique est quasi la même en ce qui concerne les livres électroniques scientifiques acquis par les bibliothèques (*eBooks*), puisque ici aussi on ne dispose généralement que d'un droit d'accès<sup>49</sup> aux serveurs de l'éditeur.

Tout ceci donne sens à un service centralisé d'archivage qui, sur mandat des bibliothèques universitaires suisses, archiverait en un seul endroit l'ensemble de leurs *eJournals* et *eBooks*. Il faut pour cela que les bibliothèques obtiennent des éditeurs le droit d'archiver ainsi que le droit de pouvoir continuer à utiliser les copies d'archive après l'échéance des contrats de licence. Cet objectif visé sur le plan juridique est au centre des préoccupations du «Consortium des bibliothèques universitaires suisses»<sup>50</sup>, qui négocie avec les éditeurs les contrats de licence pour les ressources électroniques au nom de tous ses membres. La concrétisation d'un concept de «licences nationales»<sup>51</sup> pour les publications électroniques faciliterait, sur les plans juridique et logistique, la création d'un service d'archivage centralisé.

Il y a cependant des alternatives tout à fait défendables à un service centralisé d'archivage pour les *eJournals* en Suisse. On pourrait penser à des conventions passées avec des bibliothèques étrangères qui archivent déjà les mêmes titres. C'est ainsi que la Bibliothèque nationale des Pays-Bas (Koninklijke Bibliotheek, KB) a déjà signé des conventions avec 13 maisons d'édition internationales et archive 5'000 publications électroniques (voir chap. 8.3 et 8.6.3.1). Mais une telle solution purement contractuelle a aussi ses points faibles :

- Aucune bibliothèque ne peut archiver tous les titres de tous les éditeurs. Il y aura par conséquent toujours quelques titres importants dont l'archivage ne pourra être assuré par des conventions avec l'étranger.

<sup>49</sup> A titre d'exemple la «eBook Collection» proposée par Springer-Verlag comprend déjà 25'000 *eBooks*, avec 3'500 nouvelles parutions annuelles, dont des monographies scientifiques, des manuels d'enseignement, des suites et des ouvrages de référence.

<sup>50</sup> <http://lib.consortium.ch>. Font partie du Consortium fondé en 2000 toutes les universités cantonales, le domaine des ETH, toutes les hautes écoles spécialisées, plusieurs hautes écoles pédagogiques, la Bibliothèque nationale ainsi que plusieurs partenaires secondaires du secteur public suisse. Le Consortium a été financé à 50% par la Confédération jusqu'en 2005, et est supporté depuis lors en totalité par les bibliothèques membres. Sa tâche principale consiste à négocier les licences pour ressources électroniques acquises par les bibliothèques partenaires.

<sup>51</sup> <http://www.nationallizenzen.de>

- Il faudra établir des contrats avec plusieurs bibliothèques étrangères, et pas toujours les mêmes, ce qui représente un investissement assez lourd.
- Pour avoir un statut juridique suffisant, les conventions ne peuvent être passées directement entre bibliothèques, mais doivent l'être par les états (puisque la mission d'archivage de la bibliothèque étrangère est définie par son propre cadre juridique et dépend des financements de sa tutelle).
- Les contrats seront généralement limités dans le temps, si bien que le destin de l'archivage dépendra de décisions de partenaires étrangers sur lesquelles la Suisse n'aura aucune prise.
- Les critères de sécurité et de qualité de l'archivage sont du seul ressort du partenaire étranger, sans qu'il soit possible de l'influencer.
- Il faudrait de toute façon une solution indépendante d'archivage pour les revues rétro-numérisées par des bibliothèques suisses, à moins de les remettre aux bibliothèques étrangères. On aurait alors affaire, dans ce second cas, à une convention de prestation, et non plus d'utilisation partagée.
- Bien que cela soit peu probable, il n'est pas absolument exclu que des conflits politiques ou des guerres ne compromettent l'accès aux données archivées par les bibliothèques étrangères, ou ne conduisent même à la destruction des données d'archive.

La seconde alternative à une solution centralisée consisterait à remettre l'exemplaire électronique d'archive concédé par l'éditeur à un tiers pour qu'il se charge sous mandat de l'archivage pérenne. C'est la solution proposée par la fondation américaine Portico (voir chap. 8.3 et 8.6.3.5), qui archive sur mandat des éditeurs. Portico compte à ce jour près de 8.3 millions d'articles provenant de 5'600 titres de 66 éditeurs – avec des conventions pour 8'085 autres titres – et 476 bibliothèques affiliées. Les coûts de cette prestation sont relativement modérés et ne dépassent pas, par année, 0.1 – 0.5% des dépenses annuelles de fonctionnement des bibliothèques clientes. Mais pour les bibliothèques affiliées, les risques juridiques et relatifs à la sécurité sont encore plus importants avec la solution Portico que dans le cas de la coopération avec des bibliothèques européennes, car Portico est une fondation privée qui a son siège aux Etats-Unis et dont l'existence dépend des contributions annuelles d'autres fondations américaines.

Raison pour laquelle le Consortium et la CBU sont plutôt en faveur d'une infrastructure centralisée pour la Suisse afin de conserver à l'intérieur du pays le contrôle total sur les données d'archives et sur les facteurs critiques pour leur intégrité. Qu'on l'examine sous les aspects organisationnels, logistiques, juridiques ou techniques, une telle motivation nous paraît absolument sensée.

Mais comme pour les données primaires (chap. 5.1.2), nous considérons aussi qu'un *depository* pourrait avoir un sens, à savoir une solution de pur archivage sans fonctionnalités d'interrogation pour l'utilisateur (les bibliothèques affiliées), et qui se limiterait à la conservation pérenne des *eJournals* et des *eBooks* et de leurs métadonnées bibliographiques. Comme pour les objectifs élargis de la conservation des données primaires (voir 5.1.5), la conservation pérenne comprend également la conversion et la migration des données si ces opérations s'avèrent indispensables pour garantir la poursuite de l'archivage.

Ce modèle centralisé reprend en grande partie la stratégie de l'«Allianz-Initiative ,Digitale Information'» [25] des sociétés scientifiques allemandes (voir chap. 8.5).

### 6.3 Objectifs opérationnels pour *eJournals* et *eBooks*

Les objectifs opérationnels suivants nous paraissent découler des objectifs stratégiques fixés en G7 – G9 :

- Z1. **Responsabilité à long terme** : le service d'archivage pour *eJournals* et *eBooks* assume la responsabilité opérationnelle et technique de la conservation à long terme des *eJournals/eBooks* acquis par les bibliothèques affiliées au Consortium. Il garantit la pérennité de l'accès aux archives de leurs publications lorsque l'accès en ligne à la version originale sur le serveur de l'éditeur n'a plus cours.
- Z2. **Prérequis contractuels** : le Consortium règle avec les éditeurs les dispositions juridiques nécessaires pour permettre un archivage via le service central. Cela implique le droit de conserver les données localement, ainsi que le droit de les convertir ultérieurement de son propre chef, pour autant que la sécurité de l'archivage en dépende.
- Z3. **Coûts** : les coûts de la prestation globale Z1 sont entièrement et durablement supportés par le Consortium.
- Z4. **Propriété** : le service d'archivage pour *eJournals* et *eBooks* ne revendique aucune forme de propriété sur les publications qu'il archive.
- Z5. **Réserve de responsabilité** : le service d'archivage pour *eJournals* et *eBooks* remplit sa mission avec la plus haute conscience et dans le respect des standards, mais n'assume aucune responsabilité à l'égard des utilisateurs ou de tiers quant aux dommages qui pourraient découler d'une perte de données. De même, sa responsabilité ne peut être engagée quant au contenu même des données archivées.
- Z6. **Conformité OAIS** : les solutions techniques et organisationnelles du service d'archivage nécessaires à sa mission Z1 respectent les principes et recommandations du standard international ISO 14721:2003 «Open Archival Information System Reference Model» (OAIS) [27], ainsi que des «nestor Kriterienkatalog vertrauenswürdige digitale Langzeitarchiv» [28] ou des «Trustworthy Repositories Audit & Certification: Criteria and Checklist»<sup>52</sup> [29] des «Center for Research Libraries» (CRL). Dans la mesure du possible, le service d'archivage s'efforcera d'obtenir une certification dans ce sens.
- Z7. **Réception des données** : le service d'archivage pour *eJournals* et *eBooks* reçoit des bibliothèques ou directement des éditeurs les copies d'archive des publications électroniques (y c. les métadonnées), en valide la complétude et le format, et les dépose dans l'archive. Des lacunes dans les données déposées, lacunes auxquelles on ne pourrait remédier avec le matériau livré par l'éditeur, seront compensées par le service d'archivage ou par des tiers mandatés pour parvenir à un archivage pérenne correct. Si une telle réparation s'avérait impossible pour des raisons techniques ou financières, le service

<sup>52</sup> <http://www.crl.edu/PDF/trac.pdf>



d'archivage, après consultation du Consortium, pourrait renoncer à charger les données dans l'archive.

- Z8. **Interfaces d'importation** : pour l'importation des données (Z7), le service d'archivage implémente, met en œuvre et maintient les interfaces d'importation propres à chacune des publications des éditeurs. Les interfaces doivent permettre de tester automatiquement les données, de les convertir si nécessaire, et de ne stocker dans l'archive que des objets irréprochables, en isolant ceux qui ne le sont pas. Ces interfaces doivent permettre un contrôle de qualité approprié ainsi que des contrôles par échantillonnage.
- Z9. **Formats** : le service d'archivage pour *eJournals* et *eBooks* utilise dans la mesure du possible des formats normalisés et compatibles avec l'archivage à long terme.
- Z10. **Signalement** : le service d'archivage pour *eJournals* et *eBooks* gère, sous forme structurée (base de données), un jeu minimal de métadonnées bibliographiques qui permettent l'identification univoque d'une publication quelconque. Il génère par ailleurs ses propres métadonnées pour établir un catalogue par preneur de licence, pour protocoler les étapes du processus d'archivage ou retracer l'historique d'un dépôt.
- Z11. **Sécurité** : le service d'archivage pour *eJournals* et *eBooks* veille à ce que les données soient stockées de manière sécurisée et professionnelle sur le long terme, et les protège aussi bien contre les accès non autorisés que contre une éventuelle perte due à une catastrophe.
- Z12. **Preservation planning et migration** : le service d'archivage pour *eJournals* et *eBooks* garantit que les documents confiés et leurs métadonnées conservent leur forme authentique sur le long terme et restent lisibles par machine. Il convient à cet effet avec les éditeurs que, chaque fois que c'est possible, les données seront livrées dans un format standard ou, à défaut, il s'efforce de les convertir lui-même dans ces formats (voir Z7). Durant la période d'archivage, notamment lorsque des formats peu appropriés auront été utilisés, il planifie suffisamment tôt la conversion des données dans de nouveaux formats, bien avant que les anciens ne soient devenus obsolètes ou ne puissent plus être reconnus.
- Z13. **Preuve d'intégrité** : le service d'archivage pour *eJournals* et *eBooks* fait en sorte qu'on puisse prouver l'intégrité (inaltération) des documents depuis le moment de leur prise en charge en les munissant automatiquement, au moment de l'importation (voir Z8), d'un cachet temporel conforme à la loi suisse sur les signatures électroniques ZertES<sup>53</sup> [26]. Il veille en outre à renouveler ce cachet à temps lorsque les procédés cryptographiques utilisés en Suisse ne paraissent plus assez sûrs, et fait remonter la preuve d'inaltération à l'instant du premier cachet. Pour des documents qui auront été convertis (Z12), une nouvelle chaîne de certification commence au moment de l'opération.
- Z14. **Identification de provenance** : le service d'archivage pour *eJournals* et *eBooks* dote chaque document au moment de l'importation (voir Z8) d'une estampille filigrane qui permet ultérieurement d'indiquer que le document provient des collections „du service d'archivage“. Il veille en outre à renouveler cette estampille à temps lorsque les procédés cryptographiques utilisés en Suisse ne paraissent plus assez sûrs.

---

<sup>53</sup> [http://www.admin.ch/ch/d/sr/c943\\_03.html](http://www.admin.ch/ch/d/sr/c943_03.html)

- Z15. **URN/UID/DOI** : au moment de l'importation (voir Z8), le service d'archivage pour *eJournals* et *eBooks* dote chaque publication d'un Uniform Resource Name (URN) ou d'un identificateur unique (UID), respectivement d'un Digital Object Identifier (DOI).
- Z16. **Accès** : le service d'archivage pour *eJournals* et *eBooks* n'autorise aucun accès depuis l'extérieur aux données archivées, à l'exception de Z18.
- Z17. **Interfaces d'exportation** : s'il arrive qu'un *eBook* ou un *eJournal* ne soit plus accessible en ligne sur le site d'un éditeur, le service d'archivage exporte alors sa version archivée de la publication de manière à ce que les métadonnées bibliographiques puissent être importées dans les catalogues de bibliothèques (par ex. le catalogue collectif NEBIS, etc.) et les documents eux-mêmes dans un serveur de documents centralisé. Les interfaces d'exportation nécessaires seront produits de cas en cas, car ils doivent tenir compte du format des données archivées par le service et du format des systèmes de catalogage et des serveurs de données. Z18 présente une alternative.
- Z18. **Interface intégrée/Hosting<sup>54</sup>** : le service d'archivage pour *eJournals* et *eBooks* offre une interface Web qui permet à un simple navigateur, à l'aide de l'URN/UID de la publication en question (voir Z15), d'en retrouver le fichier archivé (sans les métadonnées). Pour que cette intégration soit possible, il faut que les bibliothèques concernées aient auparavant intégré les URN/UID dans les notices correspondantes de leurs catalogues. Il faut aussi que le format du document d'archive soit supporté par le navigateur (par ex. PDF). Cette interface vérifie l'autorisation d'accès en fonction du domaine IP à partir duquel l'interrogation est lancée (par ex. celui d'une université). Les domaines autorisés sont gérés<sup>55</sup> dans le cadre de la fonction Z10, car ils sont rattachés au preneur de licence de la publication en question.

**Remarque** : Le modèle proposé aux points Z1 – Z18 pour les *eJournals* et *eBooks* ne concerne que les produits pris sous licence par les membres du Consortium des bibliothèques universitaires (dont les universités cantonales et les HES, ainsi que le domaine des EPF). Il est clair qu'il peut être facilement étendu aux publications en «Open Access»<sup>56</sup>. Comme celles-ci ne donnent pas lieu à une licence, mais sont offertes gratuitement sur Internet, tant la licence comme critère de sélection que la protection de l'investissement comme motivation de l'archivage perdent de leur pertinence.

Il pourrait s'avérer difficile pour le service d'archivage d'opérer un choix raisonnable parmi ces publications en «Open Access» et, partant, de maîtriser les coûts. Car même pour ces publications en accès libre, les éditeurs réclament au minimum un dédommagement qui couvre les frais de livraison de l'exemplaire d'archive et des métadonnées. Raison pour laquelle nous estimons qu'il est préférable de renoncer à un archivage systématique des publications en «Open Access» dans le cadre du service d'archivage national, et d'en faire un autre projet.

<sup>54</sup> Cette fonctionnalité n'est pas essentielle à notre sens. C'est un élément de la stratégie «Allianz-Initiative ‚Digitale Information‘» [22] des sociétés de recherche en Allemagne.

<sup>55</sup> La possibilité d'effectuer le contrôle d'accès via la «Authentifizierungs- und Autorisierungs-Infrastruktur (AAI)» (<http://www.switch.ch/aai/>) de SWITCH dépendra des modalités d'implémentation de cette interface intégrée. On peut partir du principe que l'accès de l'utilisateur au catalogue à distance s'effectuera déjà via AAI.

<sup>56</sup> Il existe déjà de par le monde près de 4000 publications scientifiques en «Open Access» (<http://open-access.net>, <http://www.doaj.org>), qui pour la plupart sont des publications normales d'éditeurs mises gratuitement à disposition, et pour lesquelles il n'y a par conséquent pas besoin de licence.

### 6.3.1 Coûts et financement

L'ensemble des fonctionnalités esquissées sous Z1 – Z18 devrait pouvoir être confié à un Système de gestion de documents (DMS) du marché de classe de prix intermédiaire. On retiendra comme critère essentiel de sélection la flexibilité qu'offrira le produit pour l'implémentation des procédures d'import-export, pour la prise en charge des cachets temporels certifiés (surtout des cachets réactualisés) et des estampilles filigranes, ainsi que pour une migration sans perte de tout le contenu archivé vers une nouvelle plate-forme. Le choix des formats de données destinées à l'archivage, bien qu'indépendant du produit, est tout aussi important.

Les coûts d'acquisition d'une solution DMS dépendent en grande partie du nombre de licences utilisateur. Le service d'archivage (si on excepte Z18) n'a besoin pour son usage que de 2 à 3 licences, pour les administrateurs qui importent et gèrent les données d'archive.

Comme on peut prouver leur intégrité à l'aide des cachets temporels, les objets archivés peuvent être déposés sur n'importe quel système d'archivage prévu pour un archivage pérenne. Mais il serait bon d'utiliser pour cela la même solution que pour les données primaires (voir chap. 5.1.5.1).

Le financement d'un modèle conforme à Z1 – Z18 comporte une partie d'investissements (étude de détail, achat et configuration du logiciel, réalisation et mise en exploitation) et des coûts récurrents d'exploitation (notamment pour le développement des interfaces d'import-export).

Nous estimons les frais d'investissement à 0.6 millions de francs en gros, abstraction faite des interfaces d'import-export et de la mise en place de l'infrastructure de stockage et de télécommunication. Cette dernière devrait être mise à disposition sans frais par un partenaire du monde académique. On ne tient pas compte par ailleurs, dans ces calculs, de synergies possibles avec la solution du service d'archivage pour les données primaires (chap. 5.1.2).

Nous estimons que ce montant devrait faire l'objet d'un subventionnement direct et unique de la Confédération et des cantons, des organes d'encouragement de la recherche ainsi que de fondations et, si possible, d'entreprises privées de la branche IT (sponsoring).

S'y ajoutent les coûts des interfaces d'import-export (Z8 et Z17), qu'on peut imputer soit aux coûts d'investissement, soit aux coûts de fonctionnement, puisque il faudra réaliser une interface ad hoc pour quasi chaque nouvel éditeur et pour chaque système externe mentionné en Z17. Nous estimons ces frais à environ 20 – 40'000 CHF par interface, ce qui inclut l'analyse, la spécification détaillée, le développement et les tests. Tout cela dépendra cependant de la facilité avec laquelle le produit à archiver se prêtera ou non à l'implémentation de l'interface, des exigences posées en matière d'assurance de qualité et d'éventuelles conversions de formats, et du fait que l'éditeur livrera ses données dans un format acceptable ou pas.

En revanche, les coûts de fonctionnement devraient rester modiques, puisque l'importation pourrait être largement automatisée. Des frais plus conséquents pourraient intervenir de cas en cas, par exemple pour le développement de nouvelles interfaces (voir ci-dessus) ou, tous les 5 à 10 ans, pour la migration de l'ensemble de l'archive vers un nouveau système, ou pour la conversion de certaines collections.

Conformément à Z3, les frais de fonctionnement doivent être pris en charge par les membres du Consortium des bibliothèques universitaires suisses.

Nous pensons qu'il pourrait y avoir une alternative à la création d'une solution ex nihilo : la Bibliothèque nationale (BN) dispose déjà d'un système d'archivage pour les publications électroniques produites dans le pays. Il vaudrait la peine d'examiner si un tel système pourrait satisfaire aux exigences d'une solution généralisée à l'ensemble de la Suisse pour tous les *eJournals* de toutes les bibliothèques universitaires (ou pourrait être élargi dans ce sens), et si les bases légales de la BN lui permettraient d'assumer un service d'archivage tel que présenté en Z1 – Z18 sur mandat du Consortium (et contre dédommagement).

## 6.4 Objectifs opérationnels pour d'autres types de données

### 6.4.1 Bases de données scientifiques

Beaucoup de partenaires du Consortium ont contracté par son entremise des licences d'accès à des bases de données<sup>57</sup>. Il s'agit de corpus de références bibliographiques, d'ouvrages de références thématiques, de corpus d'abstracts ou de citations, de dictionnaires et de bases de données en *fulltext* (par ex. articles de périodiques), etc. Ces bases de données proposent également des publications fortement utilisées par les chercheurs de certaines disciplines. Il peut arriver toutefois que ces bases de données, pour les mêmes raisons que pour les revues électroniques, ne soient tout à coup plus disponibles lorsque l'éditeur met un terme à son activité.

Or l'archivage de bases de données (c.à.d. de données structurées) est sensiblement plus compliqué que l'archivage de données de type documents comme les *eJournals* et les *eBooks*.

- **Structure complexe** : alors que les *eJournals* sont des fichiers bien identifiables, qui chez la majorité des éditeurs se présentent dans un format PDF (Portable Document Format) facilement accessible, chaque base de données est particulière, et beaucoup d'entre elles présentent une structure interne complexe.
- **Dépendance du logiciel** : une base de données est toujours dépendante du logiciel système sous-jacent. De plus, bien des bases de données ne peuvent être exploitées pleinement qu'avec un logiciel applicatif additionnel qui offre à l'utilisateur des fonctions avancées de recherche et de dépouillement.
- **Dynamique** : alors que la version publiée d'une revue électronique présente un caractère achevé et ne change plus, les bases de données ont un caractère dynamique du fait que leur contenu change constamment chaque fois que l'éditeur en ajoute ou en retranche. Il s'ensuit que le contenu d'une base de données ne peut être archivé que „en l'état du moment X“, donc seulement comme un „instantané“.

Comme il n'est généralement pas possible d'archiver avec les données tout le logiciel d'application qui va avec, même avec des outils d'émulation, seule une extraction pure et simple des

---

<sup>57</sup> [http://lib.consortium.ch/products\\_categories\\_lizenzen.php](http://lib.consortium.ch/products_categories_lizenzen.php)

données à un moment donné peut être envisagée. Il faut ensuite que ces données soient transposées dans un format et une structure qui les rendent utilisables et lisibles pour un futur utilisateur. Cette opération de transposition dans un format d'archivage doit être préparée et mise en œuvre pour chaque base de données spécifiquement, ce qui représente un investissement non négligeable.

On a donc besoin, pour un archivage à long terme efficace, d'un logiciel d'archivage adéquat qui extrait les données de la base de l'éditeur et les contrôle, les transpose dans un format standard adapté à l'archivage pérenne et qui, finalement, administre collectivement toutes les bases de données archivées de manière à les diffuser via une interface utilisateur unique, différente de l'environnement fonctionnel d'origine qu'on trouvait sur le site de l'éditeur.

Nous ne connaissons que deux solutions<sup>58</sup> susceptibles de remplir cette fonction de manière relativement acceptable. Elles exigent toutefois une connexion directe à la base de données (alors que l'utilisateur normal n'y accède que via l'interface utilisateur). Mais si l'éditeur se borne à n'exporter que des fichiers qu'il aura lui-même préparés, la situation est généralement désespérée pour ce qui est de l'archivage à long terme.

Il n'est à notre sens pas réaliste que les bibliothèques entreprennent un archivage systématique de nombreuses bases de données - pour autant encore que les éditeurs l'autorisent - bien que cela soit faisable dans des cas bien précis. Seule une analyse plus détaillée des besoins et des possibilités, qui dépasserait toutefois le cadre de cette étude de concept, permettrait d'émettre des recommandations plus précises.

- D1. **Bases de données** : le service d'archivage pour les bases de données scientifiques peut initier des projets pour un archivage pérenne de certaines bases de données importantes, contractées sous licence par les bibliothèques, et dont l'accessibilité n'est plus assurée. Les coûts doivent être couverts par les partenaires. Mais il n'est pas question d'un archivage systématique de toutes les bases de données auxquelles les bibliothèques se sont abonnées.

#### 6.4.2 Objets de *eLearning*

La problématique de l'archivage durable de supports électroniques pour l'enseignement et l'apprentissage a été discutée dans le chapitre 3.3. On compte parmi ceux-ci des enregistrements vidéo de cours, des versions en ligne de supports de cours, des présentations Powerpoint, des outils d'autoévaluation, mais aussi des présentations multimédia complexes, des blogs relatifs à un cours, des „chat rooms“ d'assistants, etc.

L'ETH Zurich et la Fondation SWITCH ont en cours quelques projets destinés à produire, administrer et diffuser des contenus pédagogiques numériques. Chez SWITCH en tous cas – qui met à disposition des universités une infrastructure de production de documents *eLearning* – la question de la conservation à long terme de ce type de documentation n'est pas encore au centre des préoccupations [18]. Il faut dire que les objets de *eLearning* se présentent la plupart du

<sup>58</sup> CHRONOS ([www.datenbankarchivierung.de](http://www.datenbankarchivierung.de)) de la firme allemande CSP, qui a été développé notamment grâce à des subventions du Land de Bavière et de la haute école d'informatique de Landshut, ainsi que SIARD des Archives fédérales suisses (<http://arxiv.org/abs/cs.DL/0408054>).

temps dans des formats très propriétaires, donc impropres à l'archivage. Il nous semble par ailleurs que les réalisations des établissements d'enseignement en matière de *eLearning* en sont encore à un stade relativement expérimental. Et le catalogage de ces objets (description et classement systématique) représente un travail très lourd en raison de leur multiplicité et de leur hétérogénéité.

Pour toutes ces raisons, il ne nous a pas été possible d'émettre des recommandations plus précises au sujet de l'archivage de ces objets dans un service centralisé sans procéder à une analyse plus détaillée qui dépasserait le cadre de cette étude.

### 6.4.3 Copies *master* de rétronnérisations

La problématique de l'archivage pérenne des *master*, c.à.d. des produits bruts du processus de numérisation de documents, a été discutée au chapitre 3.4. On ne fait pas allusion ici aux journaux rétronnérisés par les bibliothèques et dont l'archivage est assuré par le service d'archivage pour les *eJournals* et les *eBooks* (voir chap. 6).

On pense plutôt ici aux *master* en haute définition (résultats bruts du *scan*) d'autres documents numérisés, comme par exemple des photos, des films, des enregistrements sonores ou des ouvrages précieux. C'est à partir des *master* qu'on produit des copies de moindre qualité, mais plus faciles à manipuler lors d'une utilisation courante. Celles-ci ne sont pas archivées : si on les perd ou qu'elles ne conviennent plus à une utilisation plus exigeante, on en produit de nouvelles à partir des *master*. Ce sont donc ceux-ci qui doivent être archivés, ce qui n'est toutefois pas toujours le cas aujourd'hui. En effet, l'évolution des technologies de numérisation fait de tels progrès qu'après quelques années on préfère procéder à une nouvelle numérisation des originaux.

L'archivage pérenne des copies *master* reste néanmoins impératif lorsque les originaux sont menacés d'une destruction physique ou chimique qui interdirait de les soumettre à une nouvelle numérisation.

On trouve aujourd'hui de telles copies *master* un peu partout, notamment dans les services d'archives des cantons et de la Confédération qui produisent des copies de sauvegarde pour leurs collections, mais aussi dans des bibliothèques universitaires comme par exemple dans le cadre du projet «e-rara»<sup>59</sup> qui rétronnérisé des imprimés anciens. L'expérience montre que, pour ces projets limités dans le temps, les coûts récurrents pour un archivage sécurisé à long terme de grandes quantités de copies *master* ne sont pas pris en compte, ce qui fait que ni l'archivage pérenne ni la sauvegarde de l'investissement ne sont garantis.

Interrogés à propos d'un service national d'archivage tel que nous le proposons ici, le «Centre de coordination pour l'archivage à long terme de documents électroniques» (CECO/KOST) des archives cantonales, le projet «e-rara» ainsi que les autres projets de rétronnérisation en cours ont déclaré être surtout intéressés par une simple „bitstream preservation”, soit par un espace de stockage à bon marché. Une offre allant dans ce sens pour des projets de numérisation financés par la manne publique pourrait facilement être intégrée au service d'archivage pressenti. Cela pourrait se faire via la solution pour les données primaires (voir chap. 5.1.2), avec les quelques différences suivantes :

---

<sup>59</sup> <http://www.e-rara.ch>

- Une telle offre n'aurait de sens que pour des quantités de données importantes (par client).
- L'ingestion des données ne se fait pas en ligne, mais par envoi de bandes magnétiques ou de disques durs.
- Il n'est pas effectué de catalogage bibliographique ou archivistique.

Nous proposons par conséquent les objectifs opérationnels suivants :

- S1. **Copies numériques** : le service d'archivage des copies numériques peut, sur mandat d'institutions publiques, se charger de la conservation de copies numériques selon le principe de la „consigne“.
- S2. **Quantités et coûts** : l'offre ne vaut qu'à partir d'une quantité de 200 GB de données par institution/projet. L'institution dépositante doit couvrir les frais coûtants annuels du dépôt auprès du service d'archivage. Si elle n'entend plus le faire, elle doit reprendre ses données.
- S3. **«Bitstream preservation»** : sur mandat de l'institution dépositante, le service d'archivage des copies numériques est responsable de la conservation sécurisée et de l'accessibilité aux "chaines de bits" qui constituent les données déposées. Cela inclut la protection contre les accès non autorisés et les catastrophes, ainsi que le stockage sécurisé.
- S4. **Pas de transfert de droits** : le service d'archivage n'endosse aucun droit, de propriété ou d'utilisation, sur les données déposées. Tous les droits de propriété, d'utilisation, d'auteur, de copyright, etc., pour autant qu'il y en ait, restent acquis à l'institution dépositante.
- S5. **Réserve de responsabilité** : le service d'archivage remplit sa mission au plus près de sa conscience en se conformant à des standards exigeants, mais, en cas de perte des données déposées, il ne peut être tenu responsable des dommages encourus par l'institution dépositante ou d'autres tiers. De même il ne peut être tenu responsable du contenu des données elles-mêmes.
- S6. **Réception des données, sans contrôle** : le service d'archivage ne reçoit pas les données en ligne, mais seulement via des bandes magnétiques ou des disques durs envoyés par courrier. Il contrôle le nombre des fichiers et leurs *hashes*, mais ne se préoccupe pas de leur format de données.
- S7. **Intégrité** : le service d'archivage attribue à chaque fichier qui fait partie d'une livraison de l'institution un cachet d'horodatage certifié conforme à la loi fédérale sur la signature électronique ZertES<sup>60</sup> qui prouve l'inaltération des données depuis le moment du dépôt. Il veille en outre au renouvellement de ce cachet en temps voulu lorsque les procédés cryptographiques utilisés en Suisse ne paraissent plus assez sûrs, et reporte alors la preuve d'inaltération au moment du premier cachet.

---

<sup>60</sup> [http://www.admin.ch/ch/d/sr/c943\\_03.html](http://www.admin.ch/ch/d/sr/c943_03.html)



- S8. **UUID** : le service d'archivage identifie chaque fichier de manière univoque à l'aide d'un „Universally Unique Identifier“ (UUID)<sup>61</sup> et transmet cet UUID avec les noms de fichiers et leurs *hashes* à l'institution dépositante.
- S9. **Accès** : le service d'archivage permet à l'institution dépositante d'accéder en tout temps à ses données, moyennant préavis. Ceci peut se faire en ligne ou par courrier : dans ce dernier cas, l'institution envoie au service d'archivage les UUID des différents fichiers désirés, et celui-ci les transmet par courrier sur un support approprié. Si c'est en ligne, l'institution bénéficie d'un accès à ses données via SSH (Secure Shell) à partir d'un compte utilisateur ouvert temporairement.

---

<sup>61</sup> Les Universally Unique Identifiers (UUID) sont des nombres de 128 bits, générés par des méthodes standardisées sur le plan international (ITU-T Rec. X.667, ISO/IEC 9834-8:2005, RFC 4122). Ils sont exprimés sous la forme d'une suite de 36 chiffres hexadécimaux, comme par ex. 550e8400-e29b-41d4-a716-446655440000. Une fois créé, un UUID ne peut mathématiquement pas être généré une seconde fois, et revêt de ce fait une forme unique, universellement valable.



## 7 Modèles alternatifs

En complément aux deux modèles pour les données primaires décrits au chapitre 5, nous esquissons ici quatre alternatives qui s'écartent notablement d'un modèle de service national centralisé pour l'archivage à long terme des données primaires issues de la recherche. Nous n'avons pas creusé ces modèles alternatifs dans les détails, parce qu'ils ne répondent pas de manière adéquate à l'urgence du problème qui motive le présent projet de la CBU.

Les deux modèles décrits dans le chapitre 5 reposent, en ce qui concerne l'infrastructure et le fonctionnement opérationnel du service, sur une approche centralisée. Les alternatives qui suivent prennent donc le contre-pied en proposant une approche fédéraliste qui laisserait à l'initiative des universités le soin de répondre au besoin d'avoir une solution réellement opérationnelle pour la conservation et l'archivage des données primaires dans un avenir pas trop lointain. Dans ce cas de figure, le service national „se bornerait“ à motiver, conseiller et coordonner de telles initiatives.

Il est évident que certains éléments de ces alternatives pourraient être intégrés dans les deux modèles du chapitre 5. Mais nous risquerions alors de „surcharger“ les modèles et, partant, de réduire leurs chances d'être acceptés

En ce qui concerne l'archivage pérenne des publications des éditeurs acquises par les bibliothèques, nous ne voyons pas d'alternative raisonnable à une solution centralisée commune à toute la Suisse. En revanche, pour d'autres types de données secondaires comme par exemple la littérature grise, d'autres solutions sont envisageables.

### 7.1 Comité stratégique / Task Force nationale

Ce modèle consiste à mettre sur pied une espèce de comité national, ou une *task force* pour „l'archivage à long terme des données primaires et secondaires issues de la recherche en Suisse“ (ALT), composé de représentants des organes de la recherche, des universités et des hautes écoles, des bibliothèques et des archives d'Etat. De tels organismes ont existé ou existent dans de nombreux autres pays (voir par ex. chap. 8.6.1), mais généralement avec des résultats concrets de moindre portée.

Cet organe définit les options stratégiques, les conditions cadre et les directives qui doivent prévaloir dans tout le pays en ce qui concerne l'ALT des données de la recherche.

La mise en œuvre de la stratégie nationale est de ce fait confiée aux institutions individuelles, à savoir à leurs représentants dans cet organe, qui soutiennent l'initiative dans chacune de leurs institutions et zones de compétence. Cet organe a notamment pour fonction :

- de fixer les objectifs stratégiques de l'ALT pour les données de la recherche dans le cadre des infrastructures de la recherche en Suisse
- d'établir le statut juridique des données issues de la recherche dans les cadres légaux des services d'archives des cantons et de la Confédération
- d'établir des directives, critères, conditions cadre et procédures de contrôle pour la mise en œuvre des directives des académies suisses, du Fonds national et des universités sur la conservation des données primaires en matière d'„intégrité de la science“
- de faire des recommandations auprès des universités et hautes écoles concernant le financement ou l'encouragement de solutions et de projets ALT pour les données primaires.

Cet organe, ou cette *task force*, dispose d'un bureau qui coordonne les travaux et prépare les bases de discussion à son intention.

## 7.2 Centre national de compétences

Ce modèle propose la création d'une institution dotée de personnel scientifique propre, appelée à apporter conseil aux chercheurs et instituts de recherche de tout le pays, à maintenir un portail Internet professionnel, à soutenir ou à lancer elle-même des projets en relation avec l'archivage à long terme des données scientifiques.

Ce centre peut aussi édicter des standards et développer des outils pour l'archivage à long terme des données primaires. Mais il n'est pas de son ressort d'offrir de quelconques prestations en matière d'archivage.

On peut citer comme exemple de ce modèle le «Digital Curation Centre» (DCC) en Angleterre (voir chap. 8.6.1.2.4).

## 7.3 Réseau national de compétences

Ce modèle envisage la création d'une association qui, grâce à un subventionnement direct de la Confédération, constitue un réseau informel de personnes et d'institutions intéressées par l'archivage à long terme des données scientifiques ou impliquées dans des projets y relatifs.

Le réseau institue des groupes de travail et organise des réunions régulières de discussion, d'échange d'expériences et de préparation de recommandations concernant l'archivage à long terme des données primaires.

On peut prendre comme illustration de ce modèle les réseaux de compétences respectifs des pays voisins que sont l'Allemagne et la France: NESTOR (Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland, voir chap. 8.6.1.2.2) et PIN (Groupe Pérennisation des Informations Numériques, voir chap. 8.6.1.2.5).

## 7.4 Répertoire, inventaire et portail

Ce modèle repose sur un service rédactionnel qui repère les *repository* et *depository* de données primaires existants ou projetés en Suisse et les répertorie dans une base de données avec description sommaire et identification des personnes de contact.

Toutes ces informations sont publiées sous forme d'un répertoire sur un portail Web accessible à tous, qui peut également proposer un service de *gateway* pour les cas où les serveurs autorisent un accès en ligne public. Le portail permet aussi aux chercheurs et aux institutions d'inscrire eux-mêmes en ligne leurs propres serveurs d'archivage.

Ce genre de service serait surtout utile pour l'utilisation secondaire des données et l'établissement de contacts avec des responsables de projets.

## 8 Environnement européen

Le mandat qui fonde la présente étude demandait aussi qu'on rende compte d'activités semblables ailleurs en Europe. Nous dressons tout d'abord dans ce chapitre une brève typologie en distinguant diverses catégories de projets et en nous appuyant, à titre d'illustration, sur les projets financés par la Commission européenne portant sur l'archivage à long terme en général, sur les serveurs institutionnels et sur les bibliothèques nationales de dépôt.

Nous nous concentrons ensuite sur une importante initiative européenne et sur une autre initiative nationale de poids, toutes deux consacrées aux données primaires et qui forment une toile de fond idéale pour les deux modèles présentés au chapitre 5.

Enfin nous présentons, sous forme de catalogue, les résultats de nos recherches au sujet des stratégies nationales ou de projets individuels d'archivage à long terme de données primaires et secondaires qui pourraient être significatifs pour cette étude ou pour l'évolution de ce sous-projet de «E-lib.ch».

Rien qu'en Europe, on a peine à cerner la quantité et la variété des projets et des réalisations financés par des fonds publics qui visent à diffuser largement, ou à conserver durablement, des ressources informationnelles numériques volatiles qui résultent de l'activité scientifique, culturelle ou administrative. On peut répartir grossièrement ces activités en cinq catégories :

1. Les projets financés par la Commission européenne à travers ses programmes FP5 (1998-2002), FP6 (2002-2006) et FP7 (2007 - 2013), qui par définition supposent une participation et une orientation internationales (européennes) très larges.
2. Des stratégies, des projets ou des réalisations spécifiques, mais conduits à un échelon supranational par des associations professionnelles, des institutions ou des organismes liés à la recherche.
3. Des projets, initiatives et réalisations pluridisciplinaires, mais qui s'adressent à un cercle d'utilisateurs d'un seul pays, et qui sont la plupart du temps supportés par la bibliothèque nationale ou par des bibliothèques universitaires (parfois aussi par les archives nationales) ainsi que par les organismes du pays liés à la recherche.
4. Des projets et réalisations propres à un domaine conduits par des sociétés scientifiques ou des instituts de recherche académiques à l'intention d'un cercle d'utilisateurs de tout un pays.
5. Des projets et réalisations spécifiques limités à une seule institution ou au cercle d'utilisateurs d'un même ensemble de données.

## 8.1 Les projets EU dans le cadre des FP6 et FP7

La grande majorité des projets et réalisations tombent dans les catégories 4 et 5, et ont de ce fait soit une portée très locale, soit une ambition très limitée techniquement parlant. Les projets financés par l'Union européenne ont pris beaucoup d'ampleur depuis 2004. Rien que dans le domaine de recherche «IST-2005-2.5.10: Access to and preservation of cultural and scientific resources» des FP6 (2002-2006), 18 projets ont été financés pour un montant global de plus de 70 Mio Euro, dont les deux plus gros que sont PLANETS<sup>62</sup> et CASPAR<sup>63</sup> dotés chacun de 15 Mio Euro. Les recherches sur la conservation et l'accès pérennes aux ressources numériques de la culture et de la science se poursuivent dans le cadre du FP7 (2007 – 2013) dans le créneau «ICT-2007.4.1: Digital libraries and technology-enhanced learning» doté d'un financement global d'environ 200 Mio Euro<sup>64</sup>, dont le projet SHAMAN<sup>65</sup> pour environ 13 Mio Euro. S'ajoute à cela dans le FP7 le domaine «e-Infrastructures»<sup>66</sup> avec un budget d'environ 600 Mio Euro pour financer des projets destinés à encourager l'accès aux données scientifiques par delà les frontières.

L'engagement relativement élevé de la Commission européenne dans le domaine „Preservation and Access“, à savoir la conservation et la diffusion des ressources numériques culturelles et scientifiques, ne doit pas faire oublier que d'importants projets comme PLANETS et CASPAR (comme tous les projets FP d'ailleurs) sont d'abord des projets de recherche et non de création d'infrastructures, et que tant leurs objectifs que les consortia de partenaires qui les portent sont extrêmement diversifiés. Ce qui fait qu'ils n'ont que peu d'effets directs sur la réalisation de solutions nationales. Habituellement une quinzaine de partenaires des secteurs public et privé se répartissent l'argent alloué au projet, et les objectifs vont de l'échange d'informations d'ordre général („awareness raising“) au développement de logiciels d'application, en passant par l'élaboration de directives générales et de scénarios conceptuels („frameworks“) ou la préparation d'environnements de test („testbeds“).

L'idée est que les „frameworks“ élaborés dans le cadre de ces projets soient ensuite concrétisés au niveau national au travers d'infrastructures opérationnelles. Mais il manque en général, à ce niveau national, aussi bien le financement nécessaire à la mise en œuvre que l'ancrage effectif du projet dans la politique d'encouragement de la recherche, car ce sont les universités à titre individuel qui sont les partenaires des projets européens et non les organes officiels de la politique de la science.

<sup>62</sup> PLANETS (Preservation and long-term access to our cultural and scientific heritage), 14 Mio. € (contribution EU: 8.6 Mio. €), 2006 – 2010. Les partenaires de droit public sont les Bibliothèques nationales d'Autriche, du Danemark et des Pays-Bas, les Archives nationales des Pays-Bas et de Suisse, ainsi que quatre universités d'Angleterre, d'Autriche et d'Allemagne. IBM et Microsoft sont les partenaires du monde de l'industrie.

<sup>63</sup> CASPAR (Cultural, artistic and scientific knowledge preservation, for access and retrieval), 15.1 Mio. € (contribution EU: 8.8 Mio. €), 2006 – 2009. Les partenaires de droit public sont l'Agence spatiale européenne (ESA), l'Institut National de l'Audiovisuel pour la France, ainsi que dix universités et centres de recherche de France, Angleterre, Italie, Grèce et de la République tchèque. Avec notamment IBM pour les partenaires du monde de l'industrie.

<sup>64</sup> <http://cordis.europa.eu/fp7/ict/programme/>

<sup>65</sup> SHAMAN (Sustaining heritage access through multivalent archiving), 12.72 Mio. € (contribution EU: 8.4 Mio. €), 2008 – 2011. Les partenaires de droit public sont la Bibliothèque nationale allemande, la bibliothèque universitaire de Göttingen, le parlement flamand de Belgique ainsi que neuf universités du Portugal, d'Angleterre, d'Allemagne, de Suède et des USA. Parmi les partenaires industriels figurent entre autres PHILIPS Innovation Lab et le Xerox Research Centre Europe.

<sup>66</sup> <http://cordis.europa.eu/fp7/ict/e-infrastructure/>

## 8.2 Serveurs institutionnels

La typologie des projets présentée plus haut se fonde sur la force d'impact potentielle de chaque projet. Mais il faut ajouter pour les besoins de cette étude deux remarques importantes en ce qui concerne les contenus :

- Dans la plupart des projets, ce n'est pas tant de la conservation et de la diffusion des données primaires qu'il s'agit, mais de données secondaires au sens large, soit, à côté des publications (y c les *preprints* et la littérature grise) et des objets pédagogiques numériques, surtout des données qui résultent du dépouillement et de l'agrégation de données primaires, ainsi que de la consolidation de résultats d'expérience provenant de différentes sources<sup>67</sup>.
- De même, l'accent ne porte pas sur la conservation à long terme de la documentation, mais sur le fait qu'elle doit être accessible de manière ouverte sur Internet, par ex. dans le cadre de l'initiative Open Access<sup>68</sup>.

Ces deux orientations se défendent, car c'est d'abord un accès aisé et ouvert à des ressources documentaires validées et publiées qui fertilise la recherche, conférant par là une grande publicité à ces projets. Mais un accès durable aux informations („access“) exige des solutions correspondantes pour leur préservation et leur conservation à long terme („preservation“). C'est pourquoi tous les projets condensent ces deux aspects en une seule formulation „access and preservation“. Il n'en reste pas moins que, dans la plupart des réalisations importantes (de même que dans des projets plus limités dans le temps), la conservation à long terme des ressources numériques en jeu n'est nullement garantie, ni financièrement, ni techniquement.

On compte en Europe des centaines de serveurs institutionnels, souvent dénommés de manière un peu dépréciative „serveurs de documents“, dédiés à la collecte et à la mise en ligne de données secondaires essentiellement de type documentaire (textuelles) telles que des revues, des *preprints*, des thèses et de la littérature grise. Les projets DRIVER<sup>69</sup> et DRIVER-II<sup>70</sup> (Digital Repository Infrastructure Vision for European Research), dotés dans le cadre du FP6 (IST) et du FP7 («e-infrastructures») d'environ 5.2 Mio Euros par l'Union européenne, ont pour objectif de réaliser un inventaire et une interconnexion des serveurs institutionnels scientifiques, «*to establish a cohesive, pan-European infrastructure of Digital Repositories, offering sophisticated functionality services to both researchers and the general public*». <sup>71</sup>

<sup>67</sup> Un exemple parmi d'autres est la base de données utilisée dans le monde entier de l'Institut suisse de la bioinformatique UniProtKB (<http://www.uniprot.org>).

<sup>68</sup> Voir par ex. <http://open-access.net>

<sup>69</sup> DRIVER (Digital Repository Infrastructure Vision for European Research), 2.5 Mio. € (contribution EU: 1.84 Mio. €), 2006 – 2007. Les partenaires de droit public sont le Centre National de la Recherche Scientifique (CNRS) pour la France, la fondation hollandaise SURF, la Bibliothèque universitaire de Göttingen, le Conseil national italien de la recherche ainsi que cinq universités d'Allemagne, de Belgique, d'Angleterre et de Pologne.

<sup>70</sup> DRIVER-II (Digital Repository Infrastructure Vision for European Research), contribution EU: 2.7 Mio. €, 2008 – 2010. Les partenaires de droit public sont le Centre National de la Recherche Scientifique (CNRS) en France, la fondation hollandaise SURF, la Bibliothèque universitaire de Göttingen, la Bibliothèque universitaire de Gand, le Conseil national italien de la recherche ainsi que neuf universités d'Allemagne, de Grèce, d'Angleterre, de Pologne, du Danemark, du Portugal et de Slovénie.

<sup>71</sup> <http://www.driver-repository.eu> und <http://www.driver-support.eu>, voir aussi [20].

### 8.3 Les bibliothèques nationales de dépôt

Un grand nombre de serveurs institutionnels d'universités ou d'instituts de recherche („serveurs de documents“, voir section 8.2 ci-dessus) offrent en ligne sur Internet des thèses, des *preprints*, de la littérature grise, des revues en *open access*, des répertoires et autres ressources de tous types. Plusieurs bibliothèques nationales et de grandes bibliothèques universitaires de par le monde ont par ailleurs mis en place des solutions d'archivage pour les publications électroniques qui, bien qu'elles soient toujours souscrites en abonnement sous leur forme papier par les universités, sont également fournies sous forme électronique par les éditeurs. Ceci devient la règle, et à moyen terme bien des publications ne seront plus éditées que sous forme électronique. Ce qui soulève une problématique d'archivage particulière, décrite au chapitre 3.2.

Car des problèmes organisationnels viennent s'ajouter aux problèmes techniques : ce sont les bibliothèques universitaires, dans le cadre de leur mission d'approvisionnement en documentation scientifique, qui sont responsables de la conservation des exemplaires papier des revues auxquelles elles sont abonnées. Ce principe devrait valoir aussi pour les exemplaires électroniques, mais les universités de moindre taille ne peuvent faire front ni techniquement ni financièrement lorsqu'il s'agit d'assurer une conservation à long terme adéquate de ces ressources.

Pour les bibliothèques nationales, la situation juridique varie de pays à pays. Elles sont en tous cas obligées d'archiver toutes les publications éditées dans le pays, dans le cadre d'un mandat de dépôt qui, selon les pays, fait explicitement abstraction de la nature du support d'information. La Bibliothèque nationale suisse par exemple archive, à côté des publications des hautes écoles suisses, les revues éditées en Suisse sous forme papier ou électronique en tant que partie des eHelvetica. Mais ceci ne représente en fait qu'une infime partie résiduelle des revues auxquelles les universités et hautes écoles suisses sont abonnées.

D'autres bibliothèques nationales ont un mandat plus étendu en ce qui concerne les publications électroniques (ou interprètent leur mandat dans un sens plus large), et archivent également des revues étrangères pour lesquelles leurs universités ont acquis des licences. A titre d'exemple, le «e-Depot»<sup>72</sup> de la Bibliothèque nationale des Pays-Bas (Koninklijke Bibliotheek, KB) comprend plus de 10 millions d'objets numériques, dont la plupart sont des publications d'éditeurs internationaux. L'objectif de la KB est de signer des conventions d'archivage avec les 20 plus grands éditeurs internationaux ; il en existe 13 pour le moment.

Indépendamment des missions attribuées par la loi à la bibliothèque nationale, la question se pose dans chaque pays de savoir quelles sont les institutions qui seraient capables de garantir un accès à long terme aux publications électroniques acquises sous licence par les universités, et ceci même une fois que la licence est échue ou que l'éditeur a disparu du marché (voir chap. 3.2). Il vaut mieux qu'une seule institution assume cette tâche de manière centralisée pour toutes les universités. Elle doit alors se contenter de n'offrir que des fonctionnalités d'archivage, sans accès en ligne, aussi longtemps que les membres de la communauté académique accèdent aux ressources directement sur les serveurs des éditeurs.

---

<sup>72</sup> <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

Il va sans dire que la concrétisation d'un concept de licences nationales<sup>73</sup> rendrait plus aisée la mise en œuvre d'une solution centralisée d'archivage pour les publications électroniques, puisque ce serait une instance nationale qui traiterait avec les éditeurs, et que les restrictions usuelles qui frappent les accès en ligne réciproques entre universités seraient levées pour des exemplaires archivés à un seul endroit dans le pays.

Il faut mentionner aussi ici la fondation américaine Portico<sup>74</sup> (voir 8.6.3.5), fondée en 2005, à laquelle est affiliée entre autres la Bibliothèque du Congrès. Elle offre aux éditeurs et bibliothèques du monde entier, en échange d'une participation aux frais, la possibilité d'archiver leurs publications numériques dans l'infrastructure Portico et d'y accéder en ligne en tous temps. Pour ce faire, Portico convertit les formats originaux en „formats d'archivage standardisé” et assume la responsabilité de leur migration au fil du temps dans de nouveaux formats, si nécessaire. Portico renferme à ce jour environ 8.3 millions d'articles et a conclu des conventions avec 66 éditeurs de revues pour environ 14'000 titres.

## 8.4 L'Alliance for Permanent Access to the Records of Science

On peut dire en résumé que, si la grande majorité des projets mentionnés dans les chapitres 8.1 et 8.2 sur le thème „Access to and preservation of cultural and scientific resources” (pour reprendre l'intitulé du domaine prioritaire IST du FP6) ont dans l'ensemble beaucoup à voir avec la problématique qui nous occupe, ils n'ont que très peu d'implications concrètes pour la thématique centrale du présent projet de la CBU, à savoir l'établissement d'un service national multidisciplinaire pour la conservation à long terme des données scientifiques primaires.

Il existe par ailleurs quelques autres projets qui pourraient avoir une implication directe, mais qui bénéficient d'une dotation financière beaucoup plus modique que les gros projets européens à haute visibilité que sont PLANETS, CASPAR ou SHAMAN.

L'«Alliance for Permanent Access to the Records of Science in Europe» (Alliance PARSE, en abrégé: „L'Alliance”), créée officiellement en 2006, joue également un rôle important dans la mesure où elle se préoccupe elle aussi de la conservation de documentation scientifique, et expressément des données primaires. Mais le concept de „Records of Science” est compris dans un sens très large [24] :

- *'Science' covers the natural sciences and engineering, mathematics and computer sciences, the life sciences and the medical sciences, and the social sciences and the humanities;*
- *'Science' is understood to include both science and technology;*
- *The 'records of science' include the 'document' record, in all its modern variety, and the 'data' record, again in all the variety in terms of data formats, media and origin, including observational data gathered primarily for operational purposes;*
- *The focus will be on public, non-military data, and where privacy and ethics pose considerable additional conditions on the preservation of medical data, the latter will not be the primary focus.*

<sup>73</sup> <http://www.nationallizenzen.de>

<sup>74</sup> <http://www.portico.org>



L'Alliance fait porter ses efforts sur la création et l'interconnexion d'infrastructures nationales et l'établissement de standards transdisciplinaires, ceci en partenariat non seulement avec les institutions patrimoniales, mais surtout avec les producteurs de données eux-mêmes. Ce qui a l'avantage d'impliquer les groupes d'intérêt directement au niveau stratégique et non pas via des projets de coopération. Car l'Alliance ne cherche pas à lancer des projets elle-même, mais bien à gagner de l'influence sur les organes qui soutiennent les infrastructures nationales afin de les coordonner. L'Alliance se démarque par là clairement des autres initiatives de l'Union européenne mentionnées plus haut [24] :

- *Its focus is squarely on long-term preservation and access. Inevitably this will imply some involvement with ongoing efforts to build digital collections. But the Alliance will restrict itself to activities that follow from its perspective to establish a Europe-wide infrastructure that takes care of providing long-term access for the various user communities.*
- *Its focus is on infrastructure, and not simply on a multitude of projects. Infrastructure should be taken here to mean a virtual 'infrastructure' of a number of physical archives (or organisations that function as archives) together with a set of standards and conditions that enable them to function interoperably. [...] A vital task is therefore the establishment of a consensus in various communities as to which organisations will be the mainstays of a cost-effective infrastructure for a particular community.*
- *Its focus is on science and technology in a broad sense. It not only covers all fields of science and technology including the social sciences and the humanities, but also those areas where large data and information collections are needed for operational as well as for scientific use. This provides a direct link to the specific activities of libraries (especially research libraries) as well as, for example, archives (the collections of which serve for example historical research).*
- *It is crucial to involve not only traditional custodians of collections, such as libraries, but major stakeholders in the production ('first-hand' such as major laboratories, or 'secondhand' such as scientific publishers) or users of the records of science as well.*
- *It first of all involves these stakeholders at a strategic level, and not at the level of project cooperations. Such project-level cooperations are important; after all the real work of building digital repositories starts and ends at the project level. However what Europe has missed up until now [...] is a commitment of the major stakeholders to jointly create a digital European information infrastructure, as a direct consequence of their own strategies, i.e. as a 'board-level' issue.*

L'Alliance déclare explicitement soutenir les recommandations de l'OCDE (voir 3.1.3.1 et 3.1.3.4) sur l'accès libre aux données primaires issues de recherches financées par des fonds publics [30]. Elle insiste enfin pour qu'on ne se focalise pas uniquement sur les besoins des centres de recherche internationaux, mais qu'on prenne aussi en compte la situation particulière des chercheurs individuels et des diverses universités spécialisées, et que la question de la responsabilité de la conservation soit clairement thématifiée [24] :

- *Universities, or even individual departments, seldom have explicit deposit policies. It is usually left (and mostly implicitly) to researchers or research teams, or the communities these researchers belong to, to decide whether and how to store and maintain data. Scientific journals impose certain obligations, but the incentives for researchers to spend time and money on maintaining data collections are often absent. Yet, there is a growing feeling among universities (albeit seldom at management levels) that this will require greater attention. Some are considering university repositories, and discussions are currently being held with some research councils as to who should be responsible for providing access to (in the first place) publicly-funded research. Yet it would seem unlikely – in contrast to the world of paper publications where university libraries*

*have often served and still serve as major collections of past research output in many areas – that this will continue in the same way in the digital world.*

L'Alliance dispose aujourd'hui d'un large soutien. Ses membres comptent plus de vingt institutions réputées, comme par exemple la European Science Foundation, la European Space Agency, le CERN, la Max Planck Gesellschaft, le Science and Technology Facilities Council, la British Library, la Deutsche Nationalbibliothek, la Koninklijke Bibliotheek des Pays-Bas, le Joint Information Systems Committee britannique, la Digital Preservation Coalition, l'International Association of Scientific, Technical and Medical Publishers et le Kompetenznetzwerk Langzeitarchivierung Nestor en Allemagne. La Suisse n'y est pas représentée.

#### 8.4.1 PARSE.insight

Quelques membres de l'«Alliance for Permanent Access to the Records of Science in Europe» (PARSE) ont lancé PARSE.insight (2008 – 2010)<sup>75</sup>, financé à hauteur de 2 Mio Euro par l'Union européenne dans le cadre du volet «e-infrastructures» du FP7. Les buts du projet sont<sup>76</sup> :

- Mettre l'accent sur la conservation et l'intelligibilité à long terme de la documentation scientifique
- En conséquence, cerner plus précisément les besoins des différents domaines de recherche
- Dresser une *roadmap* unifiée qui intègre les points de vue et les stratégies nationales, européennes et globales
- Dresser un inventaire détaillé des infrastructures existantes ou planifiées
- Analyser les lacunes par comparaison entre la *roadmap* et l'inventaire
- Elaborer des directives et des outils pour faciliter l'analyse d'impact (*impact analysis*) en ce qui concerne la disponibilité à long terme des données scientifiques
- Etablir un catalogue des *best practices* ainsi que des *benchmarks* pour les serveurs d'archivage de données scientifiques

### 8.5 L'Allianz-Initiative Digitale Information en Allemagne

Lors d'une table ronde organisée début 2008 par la Deutsche Forschungsgemeinschaft (DFG), les participants ont constaté que [31] :

- c'est un fait qu'il existe aujourd'hui des pertes incontrôlées de données primaires,
- qu'il est par conséquent nécessaire d'agir dans ce domaine, car on ne peut compter par tout sur l'auto-organisation de la communauté scientifique,

<sup>75</sup> Science and Technology Facilities Council (UK), Koninklijke Bibliotheek (NL), Deutsche Nationalbibliothek (DE), Max Planck Gesellschaft (DE), International Association of Scientific, Technical and Medical Publishers (NL), European Space Agency ESRIN (FR), Fernuniversität in Hagen (DE), European Organization for Nuclear Research CERN (CH), Niedersächsische Staats- und Universitätsbibliothek Göttingen (DE).

<sup>76</sup> <http://www.parse-insight.eu>

- qu'il existe entre les différentes disciplines une grande hétérogénéité des données, des métadonnées, des environnements de recherche et des standards utilisés,
- que les scientifiques jouent un rôle central pour assurer la qualité des données et améliorer l'indispensable documentation contextuelle, et qu'en conséquence des incitations judicieuses doivent être prévues à leur intention.

Neuf organismes de recherche allemands<sup>77</sup> concluent alors en juin 2008 l'«Allianz-Initiative ,Digitale Information'» [25] pour la période 2008 – 2012. L'un des thèmes principaux de cette initiative concerne les données scientifiques primaires [25] :

- *Il est incontestable que beaucoup de ces données, après une période relativement courte d'exploitation par des chercheurs individuels ou de petits groupes, sont vouées à l'oubli, voire à la perte. Tous les organismes scientifiques s'accordent pour dire qu'il y a là un urgent besoin de conserver systématiquement ces données, de les archiver et de les diffuser en vue d'une réutilisation par des tiers.*
- *L'objectif, qui touche toutes les disciplines pour lesquelles il n'existe pas encore d'infrastructure pour les données primaires, est de mettre en place des structures qui permettent de les collecter, de les archiver et de les tenir à disposition pour une réutilisation – que cela soit pour vérifier les résultats de la recherche dans le sens de l'intégrité de la science, ou pour une réutilisation dans le cours d'une autre recherche. Une étroite coopération entre les chercheurs et les services de traitement de l'information représente le défi majeur de l'entreprise, la condition sine qua non de sa réussite.*

Cette Allianz-Initiative est en quelque sorte le pendant allemand de l'alliance internationale PARSE (voir chap. 8.4), si ce n'est qu'ici ce ne sont pas seulement quelques instituts de recherche, mais les sociétés scientifiques elles-mêmes qui veulent prendre les choses en main.

L' Allianz-Initiative poursuit trois buts :

- Définir une politique commune concernant les données primaires destinée à mieux faire prendre conscience aux chercheurs de la nécessité de pouvoir disposer d'une infrastructure pour le traitement et l'exploitation de ces données.
- Promouvoir et accompagner la collaboration entre spécialistes de disciplines et spécialistes des services d'information, et dégager des financements incitatifs pour des projets visant à définir des standards et des procédures spécifiques qui doivent permettre de construire des archives fiables et accessibles pour les données primaires, compatibles avec les interfaces d'accès internationales ou interdisciplinaires.
- Construire des serveurs d'archivage de données primaires adaptés aux besoins de chaque discipline et interconnectés sur le plan international.

D'après l'Alliance, le troisième objectif ne pourrait et de devrait «cependant être abordé que lorsque l'on disposera de suffisamment d'enseignements à partir des projets conceptuels financés pour être certain que les nouvelles structures seront bien en adéquation avec les besoins de chaque discipline, et ne passeront pas à côté».

<sup>77</sup> Alexander von Humboldt-Stiftung, Deutscher Akademischer Austauschdienst, Deutsche Forschungsgemeinschaft, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft Deutscher Forschungszentren, Hochschulrektorenkonferenz, Leibniz-Gemeinschaft, Max-Planck-Gesellschaft, Wissenschaftsrat

Les partenaires de l'Alliance se sont entendus pour coordonner leurs initiatives dans le domaine des données scientifiques primaires et évaluer le moment venu la création d'infrastructures communes pour ces données.

## 8.6 Autres activités du même ordre

Nous présentons brièvement, dans les trois sections qui suivent, d'autres activités en cours en Suisse et à l'étranger qui nous paraissent intéressantes pour le projet de la CBU en raison de leur portée nationale et interdisciplinaire. La question reste ouverte de savoir si ces activités présentent un réel intérêt pour le projet de la CBU. Seule une analyse détaillée de chaque projet permettrait d'y répondre, mais qui dépasserait le cadre de cette étude conceptuelle.

L'énumération (malgré l'intitulé de ce chapitre 8) contient également quelques projets qui se déroulent aux Etats-Unis, au Canada et en Australie. Elle se limite cependant aux projets :

- dont l'accent porte sur la conservation à long terme (au minimum 10 ans) ou l'archivage pérenne et la diffusion de données primaires structurées et de données secondaires publiées (publications) résultant de financements publics ;
- qui ont une portée nationale et interinstitutionnelle, donc qui ne se limitent pas aux données d'une seule institution ;
- qui ont un caractère multidisciplinaire en tous cas au sein d'un même domaine, donc qui ne se limitent pas à un certain type de données propres à une branche bien spécifique.

De ce fait, pour les données secondaires documentées (textuelles), on considérera essentiellement les «serveurs de documents institutionnels» dont le but est de collecter et de rendre accessibles en ligne des collections d'articles, de *preprints*, de thèses et autre littérature grise. Pour les données structurées, on retiendra quelques projets qui ou bien ne concernent pas des données primaires, mais plutôt des agrégations de données ou des dépouillements de résultats, ou bien concernent des données primaires propres à une seule institution<sup>78</sup>.

La première partie énumère des projets nationaux et internationaux, des plans d'action et des centres de compétence. La seconde recouvre des projets et prestations propres à une discipline de sciences humaines ou de sciences naturelles, mais qui ont une portée nationale. La troisième est consacrée aux projets et solutions de portée nationale dans le domaine de l'archivage des publications électroniques.

Nous nous sommes limités dans la seconde partie aux sciences humaines et aux sciences naturelles. D'une part pour fixer à cette étude des limites raisonnables. Et d'autre part parce que nous estimons que ces deux domaines ont déjà atteint un degré relativement élevé de concrétisation en ce qui concerne la conservation des données primaires, en regard des sciences économiques ou des sciences de l'ingénieur. La recherche médicale a été laissée de côté, car la problématique y est beaucoup plus complexe vu qu'il n'est en général pas possible de faire une distinction nette entre les recherches financées par des fonds publics et celles qui le sont par des fonds privés,

---

<sup>78</sup> C'est pour cette raison que n'apparaissent pas dans cette liste certains grands centres de recherche réputés d'institutions scientifiques nationales ou internationales, comme par exemple le CERN, l'ESA ou la NASA.

et que la gestion des données issues de la recherche y est dans une grande mesure bien cadrée par des lois et autres réglementations.

## 8.6.1 Stratégies, plans d'action et infrastructures nationales

### 8.6.1.1 Sur le plan international

#### 8.6.1.1.1 PARSE, PARSE.insight

Buts	Voir chap. 8.4. Sur le plan européen, l'Alliance for Permanent Access to the Records of Science in Europe (PARSE) s'est fixé pour but une action stratégique et transdisciplinaire visant à mettre en place et à coordonner les infrastructures européennes pour la conservation et l'exploitation en ligne des données scientifiques primaires.
Organisation	Créée en 2006 sous forme de fondation, l'Alliance est supportée par plus de 20 organismes réputés des domaines de la recherche européenne et des bibliothèques, parmi lesquelles la European Science Foundation, la European Space Agency, le CERN, la British Library, la Deutsche Nationalbibliothek et la Koninklijke Bibliotheek des Pays-Bas.
Financement	L'Alliance est financée par des contributions de ses membres. Le projet PARSE.insight reçoit pour sa part de la Commission européenne un financement d'environ 2 Mio Euro dans le cadre du volet «e-infrastructures» du FP7.
Site Web	<a href="http://www.parse-insight.eu">http://www.parse-insight.eu</a>

#### 8.6.1.1.2 Committee on Data for Science and Technology (CODATA)

Buts	CODATA a été fondé en 1966 comme groupe de travail de l'«International Council for Science (ICSU)», avec pour tâche principale d'améliorer la qualité, la gestion et la mise à disposition des données scientifiques en poussant à la coopération internationale
Organisation	CODATA est piloté par un comité exécutif de six membres, et dispose de plusieurs cellules et «Task Groups». Les membres en sont des organismes nationaux en charge du soutien de la recherche, des associations scientifiques internationales et quelques institutions individuelles.
Financement	Le financement de CODATA est essentiellement assuré par les contributions de ses membres.
Site Web	<a href="http://www.codata.org">http://www.codata.org</a>

### 8.6.1.2 Sur le plan national

#### 8.6.1.2.1 Allemagne : l'Allianz-Initiative 'Digitale Information' 2008 – 2010

Buts	Voir chap. 8.5. Cette alliance de neuf sociétés de recherche allemandes estime qu'il y a urgence à se préoccuper de la conservation à long terme et de l'accessibilité des données scientifiques primaires. Elle entend par conséquent renforcer la prise de conscience des chercheurs, encourager des projets destinés à définir des standards communs et mettre en place ultérieurement des infrastructures communes d'archivage des données primaires.
Organisation	L'alliance est supportée par sept sociétés scientifiques nationales, dont la Deutsche Forschungsgemeinschaft, ainsi que par le Conseil de la science et la Conférence des recteurs des universités
Financement	Promotion de projets conceptuels par l'alliance.

#### 8.6.1.2.2 Allemagne : NESTOR

Buts	NESTOR (Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland) est une plate-forme d'information et de communication à l'intention d'institutions, d'organisations et d'entreprises qui sont actives ou conduisent des projets dans le domaine de l'archivage pérenne de ressources numériques.
Organisation	NESTOR est supporté par la Deutsche Nationalbibliothek et des partenaires au projet, conseillé par un comité scientifique et dispose de cinq groupes de travail qui travaillent chacun sur un sous-thème spécifique de l'archivage à long terme.
Financement	NESTOR est financé par le Bundesministerium für Bildung und Forschung (BMBF) et par les membres partenaires (dont la Deutsche Bibliothek et la Deutsche Bundesarchiv). Pour la période 2006 – 2009, NESTOR reçoit du BMBF un subside de 1.5 Mio Euro
Site Web	<a href="http://www.langzeitarchivierung.de/">http://www.langzeitarchivierung.de/</a>

#### 8.6.1.2.3 Allemagne : KoLaWiss

Buts	Le projet KoLaWiss («Kooperative Langzeitarchivierung für Wissenschaftsstandorte») a pour ambition, en prenant comme exemple le Parc scientifique de Göttingen, de développer un modèle d'organisation et de fonctionnement pour l'archivage pérenne des données scientifiques primaires, qui doit évaluer l'interfonctionnement des infrastructures informatiques locales et leurs interconnexions avec des institutions suprarégionales et internationales sous l'angle des exigences propres à l'archivage pérenne. Ainsi les corpus de données numériques provenant des différentes institutions sont également saisis à Göttingen, analysés des points de vue des solutions techniques, du droit, des coûts, de l'organisation et des mesures de soutien, et des concepts sont élaborés dans la perspective d'une répartition des tâches, des compétences et des responsabilités. On aimerait obtenir à terme des modèles économiques à propos des coûts et de la facturation
Organisation	Limité à une année (März 2008 – März 2009) KoLaWiss est conduit par la Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG). Sont parties prenantes la Staats- und Universitätsbibliothek Göttingen (SUB) ainsi que le Geschäftsbereich Informationstechnologie et le département Medizinische Informatik de l'Université Georg-August de Göttingen. Les premiers résultats seront probablement rendus publics en décembre 2008.
Financement	KoLaWiss est soutenu par la Deutsche Forschungsgemeinschaft (DFG).
Site Web	<a href="http://kolawiss.uni-goettingen.de">http://kolawiss.uni-goettingen.de</a>

#### 8.6.1.2.4 Angleterre : Digital Curation Centre (DCC)

Buts	Le DCC se veut être le „Center of Excellence“ dans le domaine de l'archivage numérique à long terme, avec un accent national. Il se présente en outre comme capable de fournir «a wide range of resources, software, tools and support services».
Organisation	Le DCC-Konsortium est supporté par trois départements de l'Université d'Edimbourg, par le «National e-Science Centre», par les instituts HATII à l'Université de Glasgow, par UKOLN à l'Université de Bath, ainsi que par le «Science and Technology Facilities Council (STFC)».
Financement	Le budget du DCC, qui s'élève à environ 3.4 Mio Euro pour la période 2007 – 2010, est financé à près de 80% par le «Joint Information Systems Committee (JISC)». Les 20% restants proviennent de programmes e-Science du «Research Council UK (RCUK)», des organes britanniques de soutien à la recherche ainsi que des universités partenaires de Glasgow, Edimbourg et Bath

---

Site Web <http://www.dcc.ac.uk>

---

#### 8.6.1.2.5 France : Groupe Pérennisation des Informations Numériques (PIN)

---

**Buts** Le groupe PIN est un groupe de réflexion composé de représentants d'une vingtaine d'instituts de recherche, d'organes gouvernementaux et d'entreprises privées, qui se réunissent tous les trois mois en session plénière pour échanger leurs informations et discuter des problèmes liés à l'accès durable aux informations numériques.

---

**Organisation** Le groupe PIN a son siège au Centre National d'Etudes Spatiales (CNES) à Toulouse. En font partie entre autres les Archives de France, la Bibliothèque nationale de France, l'Institut National de l'Audiovisuel, France Telecom, ainsi que les ministères de la Justice et de l'Agriculture.

---

**Financement** Pas d'information

---

Site Web <http://vds.cnes.fr/pin/>

---

#### 8.6.1.2.6 Grande-Bretagne : National Digital Archive of Datasets (NDAD)

---

**Buts** Archivage de corpus de données numériques du gouvernement britannique et des autorités administratives.

---

**Organisation** NDAD est subordonné aux Archives nationales britanniques ; ses données sont gérées et stockées par le «University of London Computer Centre (ULCC)»

---

**Financement** NDAD est financé en totalité par les budgets du gouvernement britannique.

---

Site Web <http://www.ndad.nationalarchives.gov.uk>

---

#### 8.6.1.2.7 Grande-Bretagne : National Data Repository (NDR)

---

**Buts** Le NDR archive des „données de tous types” selon les désirs de ses clients issus des milieux académiques et culturels, ainsi que d'organisations à but non lucratif et de services de l'administration publique. Les prestations vont du simple dépôt avec *bitstream-preservation* à la conception, la réalisation et l'hébergement de solutions d'archivage pérenne complètes adaptées au client, en passant par le conseil, la formation, la numérisation, la saisie des métadonnées et l'archivage Web.

---

**Organisation** Le NDR est une prestation de l'«University of London Computer Centre (ULCC)».

---

**Financement** Les prestations du NDR sont calculées au prix coûtant et facturées aux clients en fonction des frais induits.

---

Site Web <http://www.ulcc.ac.uk/digital-preservation/current-activities/ndr.html>

---

#### 8.6.1.2.8 USA : National Digital Information Infrastructure and Preservation Program

---

**Buts** Le NDIIPP développe depuis 2000 une stratégie nationale pour la collecte, l'archivage, et la conservation essentiellement d'informations numériques originales qui n'existent pas sous forme analogique, dont des données scientifiques. Avec le soutien stratégique et financier du NDIIPP, les institutions partenaires collectent et archivent leurs contenus numériques, édifient et mettent en réseau leurs infrastructures d'archivage notamment dans le domaine de la e-science.

---

**Organisation** Sur mandat du Congrès, la Library of Congress a mis sur pied avec NDIIPP un réseau national de spécialistes de l'archivage pérenne venant de plus de 130 institutions partenaires publiques ou privées. En 2005, le NDIP a signé un partenariat avec la National Science Foundation.

---



---

Financement En 2000 le Congrès a libéré 100 Mio US\$ dont l'engagement est géré par la Bibliothèque du Congrès.

---

Site Web <http://digitalpreservation.gov>

---

#### 8.6.1.2.9 USA : DataNet

---

Buts Le projet «Sustainable Digital Data Preservation and Access Network Partners (DataNet)» est un programme encore au stade de la planification de l'organisme d'encouragement américain «National Science Foundation (NSF)», qui a pour but la mise en place et la conduite d'un réseau décentralisé pour l'archivage des données scientifiques à partir de 2010.

---

Organisation L'organisation de DataNet n'est pas encore bien établie. Les intéressés sont priés de s'annoncer d'ici mi-2009. Les partenaires (jusqu'à 5) doivent établir entre eux un réseau solide afin de garantir la disponibilité à long terme des données qui leur sont confiées

---

Financement DataNet est financé en totalité par la NSF. Une somme d'environ 100 Mio US\$ est à disposition pour ce programme, prévu pour cinq ans et pour cinq institutions au maximum.

---

Site Web <http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm>

---

#### 8.6.1.2.10 Canada : National Data Archive, Data Canada

---

Buts Un groupe de travail du «Social Sciences and Humanities Research Council of Canada (SSHRC)» et des Archives nationales canadiennes a mené en 2002 une étude en vue d'un «National Data Archive». En 2005, une autre étude de la «Task Force for National Consultation on Access to Scientific Research Data» aboutissait aux mêmes conclusions. Mais il semble que plus rien ne se soit passé depuis la publication de ces deux rapports.

---

Organisation Le premier rapport propose soit la création d'un réseau d'universités et d'autres institutions avec un service central de coordination, soit la création d'une agence nationale sous l'autorité des Archives nationales canadiennes. Le second rapport propose la création d'un organe central dénommé «Data Canada», qui devrait être supporté par toutes les institutions scientifiques du pays.

---

Financement Les deux rapports partent du principe que le financement proviendrait essentiellement des budgets d'encouragement de l'Etat.

---

Site Web [http://www.sshrc.ca/web/about/publications/da\\_finalreport\\_e.pdf](http://www.sshrc.ca/web/about/publications/da_finalreport_e.pdf)  
[http://ncasrd-cnadrs.scitech.gc.ca/finalrpt\\_05\\_e.shtml](http://ncasrd-cnadrs.scitech.gc.ca/finalrpt_05_e.shtml)

---

#### 8.6.1.2.11 Australie : Australia National Data Service, Australian Data Commons

---

Buts ANDS planifie la création d'un centre national pour la conservation et la disponibilité à long terme des données scientifiques. ANDS opère ainsi comme service central de coordination, les données continuant à être stockées localement. En parallèle, elles doivent être centralisées dans un Australian Data Commons afin d'y être disponibles, utilisables et consultables à long terme.

---



Organisation	La structure organisationnelle d'ANDS n'est pas encore définitivement fixée. Ce sont aujourd'hui la Monash University à Melbourne, l'Australian National University (ANU) à Canberra et la Commonwealth Scientific and Industrial Research Organisation (CSIRO) qui ont pris l'initiative. Il est prévu d'instituer par la suite un comité de pilotage au sein duquel tous les partenaires auront un siège. ANDS est dirigé par un directeur ainsi que deux sous-directeurs à Monash University et à l'Australian National University. Le siège se trouve à Monash. Le nombre de collaborateurs devrait avoisiner 20 à 22 personnes
Axes d'action	Le domaine d'activité d'ANDS comprend quatre secteurs. Developing Frameworks : élaboration d'un cadre juridique et institutionnel commun pour la conservation, la gestion et l'utilisation des données scientifiques. Providing Utilities : développement d'outils pour la recherche, la consultation et l'exploitation des données scientifiques stockées (portail commun, identification unique, etc.). Seeding the Commons : soutien aux institutions associées pour l'archivage et la mise à disposition à long terme de leurs données. Building Capabilities : constitution d'un corpus de connaissances en matière de production, de stockage, de gestion et d'utilisation à long terme de données scientifiques.
Financement	Il n'existe à ce jour qu'une esquisse de budget, qui prévoit d'ici 2011 des dépenses pour quelque 24.5 Mio AU\$. La majeure partie des coûts sera supportée par la «National Collaborative Research Infrastructure Strategy (NCRIS)» du «Department of Innovation, Industry, Science and Research (DIISR)» australien. Il est prévu que les institutions partenaires participent également aux frais d'ANDS, que ce soit par la mise à disposition de personnel, de services ou de produits. On estime que ces subventions en nature atteindraient à peu près le même niveau que la subvention globale du NCRIS.
Site Web	<a href="http://ands.org.au">http://ands.org.au</a>

## 8.6.2 Conservation des données primaires propres à certaines disciplines

### 8.6.2.1 Sciences sociales

#### 8.6.2.1.1 Sur le plan international : Council of European Social Science Data Archives (CESSDA)

**Buts** L'objectif principal est l'édification d'un CESSDA-catalogue qui permette de consulter en ligne les serveurs d'archivage des partenaires nationaux et d'en exploiter les données. L'association organise par ailleurs des „Expert Seminars“ annuels consacrés aux problématiques d'archivage dans les pays concernés

**Organisation** CESSDA est l'organe faîtière de 20 centres de service nationaux dans le domaine des sciences sociales

**Financement** CESSDA se finance essentiellement par les contributions des partenaires.

**Site Web** <http://www.cessda.org>

#### 8.6.2.1.2 Suisse : Informations- und Datenarchivdienst für die Sozialwissenschaften (SIDOS)

**Buts** SIDOS archive les données issues de la recherche en sciences sociales, et établit un inventaire des projets de recherche.

**Organisation** Fondé en 1992, SIDOS est depuis le 1<sup>er</sup> janvier 2008 une section de la «Fondation suisse pour la recherche en sciences sociales (FORS)» et a son siège à l'Université de Lausanne.

---

Financement SIDOS est financé essentiellement par les subventions de l'Académie suisse des Sciences humaines et sociales (SAGW).

---

Site Web <http://www.sidos.ch>

---

#### 8.6.2.1.3 Allemagne : Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS)

Buts GESIS-ZA archive du matériau primaire (données, questionnaires, plans de code) et les résultats d'enquêtes empiriques. Le matériau est préparé pour une recherche analytique secondaire et mis à disposition publiquement.

---

Organisation L'ancien «Zentralarchiv für Empirische Sozialforschung» est maintenant une section de l'Institut GESIS récemment créé. Une collaboration étroite persiste avec l'Université de Cologne.

---

Financement Le GESIS est financé par les subventions d'encouragement étatiques que l'Etat fédéral et les Länder accordent selon disponibilité sur la base de conventions passées entre l'Etat et les Länder. Aucune contribution n'est demandée aux partenaires

---

Site Web <http://www.gesis.org>

---

#### 8.6.2.1.4 Grande Bretagne : UK Data Archive (UKDA)

Buts L'UKDA, créée en 1967, est un centre de compétences attaché à la production, à la conservation et à l'exploitation de données relatives aux sciences sociales, qui, avec plus de 5000 collections de données, dispose de la plus importante archive dans ce domaine en Grande-Bretagne. L'UKDA est reconnue par les Archives nationales britanniques comme lieu privilégié de conservation («Place of Deposit») pour les données numériques.

---

Organisation L'UKDA a son siège à l'Université d'Essex (Colchester) et est l'une des quatre organisations partenaires de l'«Economic and Social Data Service (ESDS)» créé en 2003. Les treize départements de l'UKDA, qui occupent environ 50 collaborateurs, sont dirigés par un directeur.

---

Financement L'UKDA est financée par l'«Economic and Social Research Council (ESRC)», le «Joint Information Systems Committee (JISC)» et l'Université d'Essex.

---

Site Web <http://www.data-archive.ac.uk>

---

#### 8.6.2.1.5 USA : Data Preservation Alliance for the Social Sciences (Data-PASS)

Buts L'objectif de ce projet est d'identifier, avec les partenaires, des corpus de données en sciences sociales qui méritent l'archivage et sont menacés de disparition, de les sauver et de les archiver durablement sur une plate-forme commune. Il s'agit surtout de sondages d'opinion, de résultats de scrutins et d'enquêtes à grande échelle sur l'évolution ou le revenu des familles.

---

Organisation Le projet Data-PASS est actuellement supporté par sept institutions partenaires aux Etats-Unis, dont la «Library of Congress», les Archives nationales américaines («National Archives and Record Administration NARA») ainsi que l'ICPSR (voir 8.6.2.1.6).

---

Financement Data-PASS est financé par une contribution du «National Digital Information Infrastructure and Preservation Program (NDIIPP)» de la «Library of Congress» (voir 8.6.1.2.8).

---

Site Web <http://www.icpsr.umich.edu/DATAPASS>

---

#### 8.6.2.1.6 USA : Inter-university Consortium for Political and Social Research (ICPSR)

Buts	L'ICPSR rassemble et archive les données issues de la recherche en sciences sociales pour les mettre ensuite à disposition de la collectivité. Un point fort d'ICPSR est le soutien actif fourni aux utilisateurs lors de l'exploitation effective des données, par la mise à disposition d'une documentation exhaustive et par des séances d'introduction aux collections et aux méthodes analytiques de dépouillement.
Organisation	Fondé en 1962, l'ICPSR compte aujourd'hui près de 660 collaborateurs (pour la plupart universitaires) de par le monde. Il est dirigé par un conseil de douze membres élus pour quatre ans. Son siège, qui se trouve à l'Université de Michigan à Ann Arbor, occupe environ 27 collaborateurs à plein-temps.
Financement	22% des recettes globales de 14.25 Mio US\$ en 2007 provenaient des contributions des partenaires, et 46% de crédits de subventionnement liés à des projets ainsi que de dons.
Site Web	<a href="http://www.icpsr.umich.edu">http://www.icpsr.umich.edu</a>

## 8.6.2.2 Sciences naturelles

### 8.6.2.2.1 Sur le plan international : ICSU World Data Centers (WDC)

Buts	Les World Data Centers (WDC) forment un réseau d'entités décentralisées qui archivent les données scientifiques d'une discipline particulière pour les mettre à disposition de la science.
Organisation	Les 52 WDC répartis dans 12 pays sont rassemblés sous le toit de l'International Council for Science (ICSU), mais de manière relativement peu contraignante. Les informations qui figurent sur le site de l'ICSU datent déjà de plusieurs années et n'ont plus été actualisées depuis au moins deux ans.
Financement	Il n'y a pas de financement central ; chaque WDC dispose de son propre budget, qui la plupart du temps provient d'une université, d'un organisme ou d'un projet de recherche.
Site Web	<a href="http://www.ngdc.noaa.gov/wdc/">http://www.ngdc.noaa.gov/wdc/</a>

### 8.6.2.2.2 USA : SDSC Data Central, San Diego Supercomputing Center (SDSC)

Buts	Offrir du <i>know-how</i> , des prestations et des capacités de stockage pour l'archivage de données issues de projets de recherche à financement public ou privé, ou de données de bibliothèques ou de centres d'archives américains. Le SDSC archive à ce jour près de 120 collections de données scientifiques qui occupent 5 petabytes, alors que sa capacité totale est de 27 petabytes.
Organisation	Le SDSC est une entité de recherche de l'Université de Californie, à San Diego, qui compte environ 400 collaborateurs. Le Data Central est une section du SDSC.
Financement	Le financement du SDSC est essentiellement assuré par des moyens provenant de la «National Science Foundation (NSF)». L'archivage des données n'est pas facturé, mais les intéressés doivent toutefois introduire une demande d'adhésion et répondre aux critères de la National Science Foundation.
Site Web	<a href="http://datacentral.sdsc.edu">http://datacentral.sdsc.edu</a>

### 8.6.2.2.3 Allemagne : World Data Center for Climate (WDCC)

Buts	Le WDCC n'archive pas de données primaires, mais prend en charge des lots de données agrégées après dépouillement de relevés sur les modèles climatiques et de séries de mesures pour les mettre à disposition des scientifiques. Le Deutsche Klimarechenzentrum (DKRZ) ainsi que quelques centres de calcul internationaux actifs dans le domaine du climat en sont les principaux fournisseurs.
Organisation	Le WDCC se trouve au Max-Planck-Institut für Meteorologie à Hamburg, qui lui-même fait partie du Zentrum für Meeres- und Klimaforschung de l'Université de Hamburg. Le WDCC travaille par ailleurs en étroite collaboration avec le Deutsche Klimarechenzentrum (DKRZ) à Hamburg, dont les parrains sont également la Max-Planck-Gesellschaft et l'Université de Hamburg.
Financement	Le financement du WDCC est assuré par les quatre sociétaires du DKRZ et par le Bundesministerium für Bildung und Forschung (BMBF).
Périmètre	Le corpus de données du WDCC comprenait à fin 2007 environ 340 terabytes stockés sous forme de plus de 6 milliards de BLOBs (Binary Large Objects) dans une banque de données. A côté des données produites par le centre sont également stockées des données comparables provenant de sources extérieures avec leurs métadonnées. Ces données externes représentent environ 5% de l'ensemble.
Prestations	En plus de l'archivage de données climatiques, le WDCC offre encore d'autres prestations. Avec son Integrierende Modell- und Daten-Infrastruktur (IMDI), le centre offre un environnement pour compiler et exécuter des programmes de modélisation du système terrestre et pour l'analyse et l'archivage des résultats. Les collections de données dont l'exploitation est terminée, et qui ont une portée significative pour l'avancement de la recherche, sont dotées d'un identificateur permanent (Digital Object Identifier (DOI) et mises à disposition de tous via le catalogue central de la Technische Informationsbibliothek (TIB) à Hannovre.
Site Web	<a href="http://wdc-climate.de">http://wdc-climate.de</a>

### 8.6.3 Archivage de publications électroniques

La plupart des bibliothèques nationales disposent de solutions pour l'archivage de publications électroniques, de journaux rétronumérisés et de collections iconographiques, de même que, parfois, pour des sites Web relevant du domaine Internet national (voir chap. 8.3.) Nous n'avons retenu que trois exemples européens ainsi qu'un prestataire international. Mais la liste est loin d'être close.

#### 8.6.3.1 Pays-Bas : e-Depot Koninklijke Bibliotheek

Buts	Le e-Depot de la Bibliothèque nationale des Pays-Bas (Koninklijke Bibliotheek, KB) constitue une solution d'archivage à long terme de ressources numériques supportée par l'environnement informatique (DIAS) de la firme IBM. Il renferme aujourd'hui la collection numérique nationale de la KB, et accueillera à l'avenir l'archive des sites Web hollandais ainsi que les masters de rétronumérisations réalisées par la KB.
Quantité	Le e-Depot de la KB comprend à ce jour 5 000 titres de publications électroniques (eJournals) représentant 10 millions d'articles. Cette masse de données occupe 11 TB ; et la capacité actuelle de 25 TB peut être étendue à 1,5 PB. La KB a conclu des „archiving agreements” avec 13 éditeurs de revues qui lui livrent périodiquement les versions numériques pour importation dans le e-Depot.

---

**Financement** On ne connaît pas le montant des investissements pour le e-Depot. Les coûts internes qui incombent à la KB pour la „Digital Archiving and R&D Digital Preservation“ s’élèvent à 1.1 Mio Euro pour le personnel et la maintenance du système et à 1.2 Mio Euro pour le secteur recherche et développement.

---

**Site Web** <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

---

### 8.6.3.2 Allemagne : Kopal

---

**Buts** Kopal gère un système d’archivage destiné à garantir la disponibilité à long terme de publications numériques. Il s’agit d’un service coopératif offert aux bibliothèques, aux archives et aux musées, ainsi qu’aux universités et aux organismes de recherche. L’infrastructure technique recourt au logiciel DIAS d’IBM, en usage aussi à la Bibliothèque nationale des Pays-Bas (voir 8.6.3.1).

---

**Organisation** Kopal est supporté par la Deutsche Nationalbibliothek (DNB), la Niedersächsischen Staats- und Universitätsbibliothek (SUB), la Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) et la firme IBM. Le centre opérationnel de l’archive pérenne se trouve au GWDG. La DNB et la SUB utilisent Kopal également pour archiver leurs propres publications.

---

**Financement** Le projet de création du système d’archivage actuel (2004 – 2007) a été soutenu par une subvention à hauteur de 4.2 Mio Euro du Bundesministerium für Bildung und Forschung (BMBF).

---

**Prestations** Kopal propose essentiellement trois manières différentes de procéder. En tant que participant : une institution fait archiver ses données par l’intermédiaire d’un mandataire. En tant que mandant : une institution gère de manière autonome son propre secteur du système d’archivage Kopal. Avec une installation propre : une institution gère son propre système d’archivage à l’aide du software Kopal en bénéficiant des enseignements du projet Kopal.

---

**Site Web** <http://kopal.langzeitarchivierung.de>

---

### 8.6.3.3 Suisse : e-Helvetica

---

**Buts** La Bibliothèque nationale suisse (BN) archive, avec son projet e-Helvetica, les publications électroniques qui tombent dans la catégorie des eHelvetica. Celle-ci comprend, à côté des publications officielles, les publications académiques, les thèses ainsi que les sites Web du domaine .ch, de même que des journaux électroniques publiés en Suisse (qui ne représentent toutefois qu’une part de plus en plus modeste des revues utilisées dans les universités du pays).

---

**Financement** Le financement du projet e-Helvetica provient des moyens propres de la BN ainsi que de crédits que la Confédération a libérés pour le volet eGovernment.

---

**Prestations** La BN offre à des tiers une solution de conservation des ressources numériques selon le principe de la mise en consigne.

---

**Site Web** [http://www.nb.admin.ch/slb/slb\\_professionnel/01693/index.html?lang=de](http://www.nb.admin.ch/slb/slb_professionnel/01693/index.html?lang=de)

---

### 8.6.3.4 France : Hyper Article en Ligne (HAL)

---

**Buts** L’objectif général est l’édification d’une archive nationale ouverte pour les publications scientifiques.

---

---

Organisation	HAL est un projet du «Centre pour la Communication Scientifique Directe (CCSD)» qui fait partie du «Centre national de la recherche scientifique (CNRS)», l'organe officiel français pour l'encouragement de la recherche. HAL est supporté par de nombreuses organisations scientifiques du pays, qui en composent le comité stratégique (avec quelques représentants du gouvernement). Le projet tourne opérationnellement au CCSD, qui occupe six personnes à cet effet.
Financement	Le financement de HAL est assuré en majeure partie par le CCSD.
Périmètre	HAL renferme aujourd'hui plus de 100'000 publications scientifiques.
Site Web	<a href="http://hal.archives-ouvertes.fr">http://hal.archives-ouvertes.fr</a>

---

### 8.6.3.5 Sur le plan international : Portico

---

Buts	Portico propose aux bibliothèques et éditeurs du monde entier d'archiver leurs publications numériques dans son infrastructure ad hoc en leur garantissant un accès en ligne en tout temps. Pour ce faire, Portico convertit les formats originaux en „formats d'archivage standardisés“ et assume la responsabilité de la migration ultérieure vers de nouveaux formats, au cas où cela serait nécessaire. A cet effet, les clients doivent signer une „licence d'archivage“ qui autorise Portico à effectuer la conversion originale et les migrations ultérieures de leurs contenus numériques.
Organisation	Portico est une fondation sans but lucratif fondée en 2005 aux USA par plusieurs partenaires, dont les fondations JSTOR et Ithaka, ainsi que la Library of Congress et la Andrew W. Mellon Foundation.
Financement	Portico facture ses prestations. Les éditeurs paient une contribution annuelle qui se monte à 0.03 – 0.1% du chiffre d'affaires global du secteur revues. Les contributions des bibliothèques représentent 0.1 – 0.5% de leurs dépenses d'équipement. Portico reçoit d'autres subventions conséquentes de ses fondateurs (voir ci-dessus)
Périmètre	Portico renferme aujourd'hui près de 8.3 millions d'articles tirés de 5'600 titres de 66 maisons d'édition et de 476 bibliothèques participantes. Il existe des accords avec des éditeurs pour 8'085 autres titres.
Site Web	<a href="http://www.portico.org">http://www.portico.org</a>

---

## 9 Bibliographie

1. Wolfram Neubauer, Antrag für Einzelvorhaben im Rahmen des nationalen Projektes Elektronische Bibliothek Schweiz „E-lib.ch“: Konzeptentwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten für die Schweiz, 20.08.2008.
2. Bundesgesetz über die Archivierung (BGA, SR 152.1). 26.6.1998.
3. Pascalia Boutsouci, Daniel Metje, Digitale Langzeitarchivierung. Erhebung an der ETH Zürich und der Georg-August-Universität Göttingen, in Deutscher Bibliothekartag 2008, Forum Langzeitarchivierung, Mannheim, 5 juin 2008. 2008.
4. Thomas Severiens, Eberhard R. Hilf, Institute for Science Networking Oldenburg GmbH and an der Carl von Ossietzky Universität Oldenburg, Studie zum Stand vorhandener Forschungsdaten und Rohdaten aus wissenschaftlichen Tätigkeiten: Erfordernisse und Eignung zur Archivierung bzw. Zurverfügungstellung in Deutschland (Primärdaten). in nestor - Materialien 6 (<http://nbn-resolving.de/urn:nbn:de:0008-20051114018>). 2006.
5. American Economic Review. Data Availability Policy ([http://www.aeaweb.org/aer/data\\_availability\\_policy.html](http://www.aeaweb.org/aer/data_availability_policy.html)). 2008.
6. European Science Foundation (ESF), Good scientific practice in research and scholarship. European Science Foundation Policy Briefing, 2000(10).
7. Organisation for Economic Co-Operation and Development (OECD), OECD Principles and Guidelines for Access to Research Data from Public Funding (<http://www.oecd.org/dataoecd/9/61/38500813.pdf>). 2007.
8. Deutsche Forschungsgemeinschaft (DFG), Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ (<http://www.dfg.de/antragstellung/gwp/>). 1998, Weinheim: Wiley-VCH.
9. Bundesgesetz über die Forschung (SR 420.1). 7. Oktober 1983 (Edition du 25 février 2008).
10. Akademien der Wissenschaften Schweiz, Wissenschaftliche Integrität - Grundsätze und Verfahrensregeln (<http://www.akademien-schweiz.ch/Publikationen/Richtlinien/>). 2008, Bern.
11. Schweizerischer Nationalfonds, Beitragsreglement. Reglement des Schweizerischen Nationalfonds über die Gewährung von Beiträgen ([http://www.snf.ch/SiteCollectionDocuments/allg\\_reglement\\_d.pdf](http://www.snf.ch/SiteCollectionDocuments/allg_reglement_d.pdf)). 14.12.2007 (approuvé par le Conseil fédéral le 13 février 2008).
12. Schweizerischer Nationalfonds, Reglement über die Information, die Valorisierung und die Rechte an Forschungsergebnissen, N. Forschungsrat, Editor. 17.6.2008.
13. Senat der Universität Bern, Reglement über die wissenschaftliche Integrität vom 1.5.2007 ([http://www.forschung.unibe.ch/content/richtlinien/integritaet/index\\_ger.html](http://www.forschung.unibe.ch/content/richtlinien/integritaet/index_ger.html)).
14. Schulleitung ETH Zürich, Richtlinien für Integrität in der Forschung und gute wissenschaftliche Praxis an der ETH Zürich vom 14. November 2007 ([http://www.rechtssammlung.ethz.ch/pdf/414\\_Integrit%C3%A4t-Forschung.pdf](http://www.rechtssammlung.ethz.ch/pdf/414_Integrit%C3%A4t-Forschung.pdf)).
15. Tony Mayer, Nick Steneck, Final Report to ESF and ORI: First World Conference on Research Integrity: Fostering Responsible Research (Lisbon, Portugal, 16-19 September 2007):

- [http://www.icsu.org/5\\_abouticsu/PDF/WC\\_final\\_report.pdf](http://www.icsu.org/5_abouticsu/PDF/WC_final_report.pdf). 2007, European Science Foundation, and U.S. Office of Research Integrity.
16. Peter Doorn, Heiko Tjalsma, Introduction: archiving research data. *Archival Science*, 2007. 7(1)
  17. Matthias Töwe unter Mitarbeit von Arlette Piguet, Konzeptstudie E-Archiving, Version 1.2 ([http://lib.consortium.ch/external\\_files/Konzeptstudie\\_D\\_V1\\_2.pdf](http://lib.consortium.ch/external_files/Konzeptstudie_D_V1_2.pdf)). 2005, Konsortium der Schweizer Hochschulbibliotheken: Zürich.
  18. Peter Keller-Marxer, Niklaus Bütikofer, «E-lib.ch»-Projekt „Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten für die Schweiz“: Workshop II mit erweiterem Personenkreis, 26.08.2008, ETH Zürich. Bericht zum Workshop. 2008.
  19. Verordnung zum Bundesgesetz über die Archivierung (VBGA, SR 152.11). 8.9.1999.
  20. Botschaft des Bundesrates über das Bundesgesetz über die Archivierung. Bbl, 26.2.1997. 1997(II).
  21. Archiv der ETH Zürich, Richtlinien für die Bewertung von Unterlagen am Archiv der ETH Zürich ETH-Bibliothek, Editor. 2002.
  22. Staatskanzlei Kanton Bern, Gesetz über die Archivierung. Fassung vom 12. Juni 2008 für die Vernehmlassung. 2008.
  23. C. Wallace et al., IETF RFC 4810: Long-Term Archive Service Requirements (<http://www.ietf.org/rfc/rfc4810.txt>). 2007, Internet Engineering Task Force (IETF).
  24. European Task Force Permanent Access, PERMANENT ACCESS TO THE RECORDS OF SCIENCE: Strategic Action Programme 2006 - 2010 (<http://www.alliancepermanentaccess.eu/Strategic%20Action%20Programme.pdf>). 2005, Amsterdam: Koninklijke Bibliotheek.
  25. Deutsche Forschungsgemeinschaft, Hochschulrektorenkonferenz, Wissenschaftsrat, Max-Planck-Gesellschaft, Leibniz-Gemeinschaft, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft Deutscher Forschungszentren, Alexander von Humboldt-Stiftung u.a., Schwerpunktinitiative „Digitale Information“ der Allianz-Partnerorganisationen: Allianz-Initiative Digitale Information – Das Leitbild ([http://www.dfg.de/aktuelles\\_presse/das\\_neueste/download/pm\\_allianz\\_digitale\\_information\\_details\\_080612.pdf](http://www.dfg.de/aktuelles_presse/das_neueste/download/pm_allianz_digitale_information_details_080612.pdf)).
  26. SR943.03: Bundesgesetz vom 19. Dezember 2003 über Zertifizierungsdienste im Bereich der elektronischen Signatur (Loi fédérale sur la signature électronique, ZertES). 2003.
  27. Consultative Committee for Space Data Systems (CCSDS), ISO 14721:2003: Space data and information transfer systems - Open archival information system - Reference model. 2003, Geneva: International Organization for Standardization (ISO).
  28. nector-Arbeitsgruppe Vertrauenswürdige Archive - Zertifizierung, Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 1: <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>. nector Materialien. Vol. 8. 2006, Frankfurt am Main: nector - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland.
  29. Inc. (OCLC) The Center for Research Libraries (CRL) and Online Computer Library Center, Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), Version 1.0, February 2007: <http://www.crl.edu/PDF/trac.pdf>. 2007, Chicago: The Center for Research Libraries (CRL) and Online Computer Library Center, Inc. (OCLC).
  30. Luigi Fusco, Co-ordinating national efforts: Alliance for Permanent Access (<http://www.wepreserve.eu/events/fp6-2007/presentations/Co-ordinating-national-efforts-luigi.ppt>), in DPE, PLANETS, and CASPAR Second Annual Conference: Progress towards Addressing Digital Preservation Challenges, 5-6 September 2007. 2007.
  31. Thekla Kluttig, Bericht über das DFG-Rundgespräch „Forschungsprimärdaten“ am 17.01.2008 ([http://www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/lis/download/forschungsprimaerdaten\\_0108.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/forschungsprimaerdaten_0108.pdf)). 2008: Deutsche Forschungsgemeinschaft (DFG).