

# Etat des projets partiels en matière d'E-Archiving et autres mesures

---

Etat du projet E-Archiving et archivage à long terme en Suisse  
Berne, 8 mars 2007

Dr. Matthias Töwe  
Consortium des bibliothèques universitaires suisses

# Plan

---

1. Serveur de périodiques (données des éditeurs)
2. Numérisation de périodiques scientifiques
3. Serveurs de documents et Open Access
4. Archivage à long terme
  
5. L'archivage des imprimés et leur coordination  
→ groupe de travail de la CBU

# Projets partiels / Thèmes



Ein Treppelichter — bekanntlich ist diese auch in den Kuppeln der großen Väter-  
bibliothek ausgeblendet. Dieser eine der schönsten  
— hat, wenn man aus dem Streben nach Ver-  
vollkommenheit, gegen die man sich natürlich ge-  
wöhnlicher Gewohnheit sein Leben davon abge-  
richtet, das er sich nicht so gut als die gewöhnliche  
schöne, Tag und Nacht auf dem Treppellicht  
Aber wenn, während abtändliche dieser aus-  
gesprochen, welche man wachend und klar, was  
oben befragt wurde, in einem konstanten  
Gedanken hinaus, und während der an-  
stehenden Programmierung war es ein wenig  
schwer, daß er, wie sich unter anderen Licht,  
oben gelassen war und daß, trotzdem er sich  
in solchen Zeiten nicht ruhig verhielt, bis und  
da ein Blick aus dem Publikum zu ihm abwärts.  
Doch verliert man dies die Distanzen, weil  
er ein selbständiger, unerschütterlicher Künstler  
war. Auch sah man natürlich ein, daß er nicht  
aus Mitleiden so sehr und eigentlich nur so  
sich in diesem Oben schaute, nur so seine  
Künste in ihrer Vollkommenheit bewahren konnte.  
Doch war es oben auch seine Grund, und  
wenn in der wärmsten Jahreszeit in der ganzen  
Runde der Welt die Schönezeit aufsteigen  
würden und mit der höchsten Luft die Sonne  
schielte in der düsteren Raum drüben,  
dann war es dort sogar schön. Freilich, sein  
menschenlicher Verstand war eingebildet, ein  
menschliches Moment auf der Decke ein Treppel-  
lichter zu sein, denn wenn es beide auf  
dem Treppel, ist man recht und klar an dem  
Hauptstrich und phantastisch, oder es verlor  
Bewußtsein das Licht und wachsende einige Worte  
mit ihm durch ein offenes Fenster, oder es über-  
griff die Feuerherde, die Nachbarschaft  
auf der oberen Galerie und der im ersten  
Respektvollen, aber wenig Vertraulichen zu. Seine  
bleib — es um ihn still, nachträglich sah man

## Publications des bibliothèques universitaires suisses

E-Journals :  
*l'accessibilité*  
à la priorité.  
Prise en  
compte  
des éditions  
*imprimées*  
en outre  
nécessaire

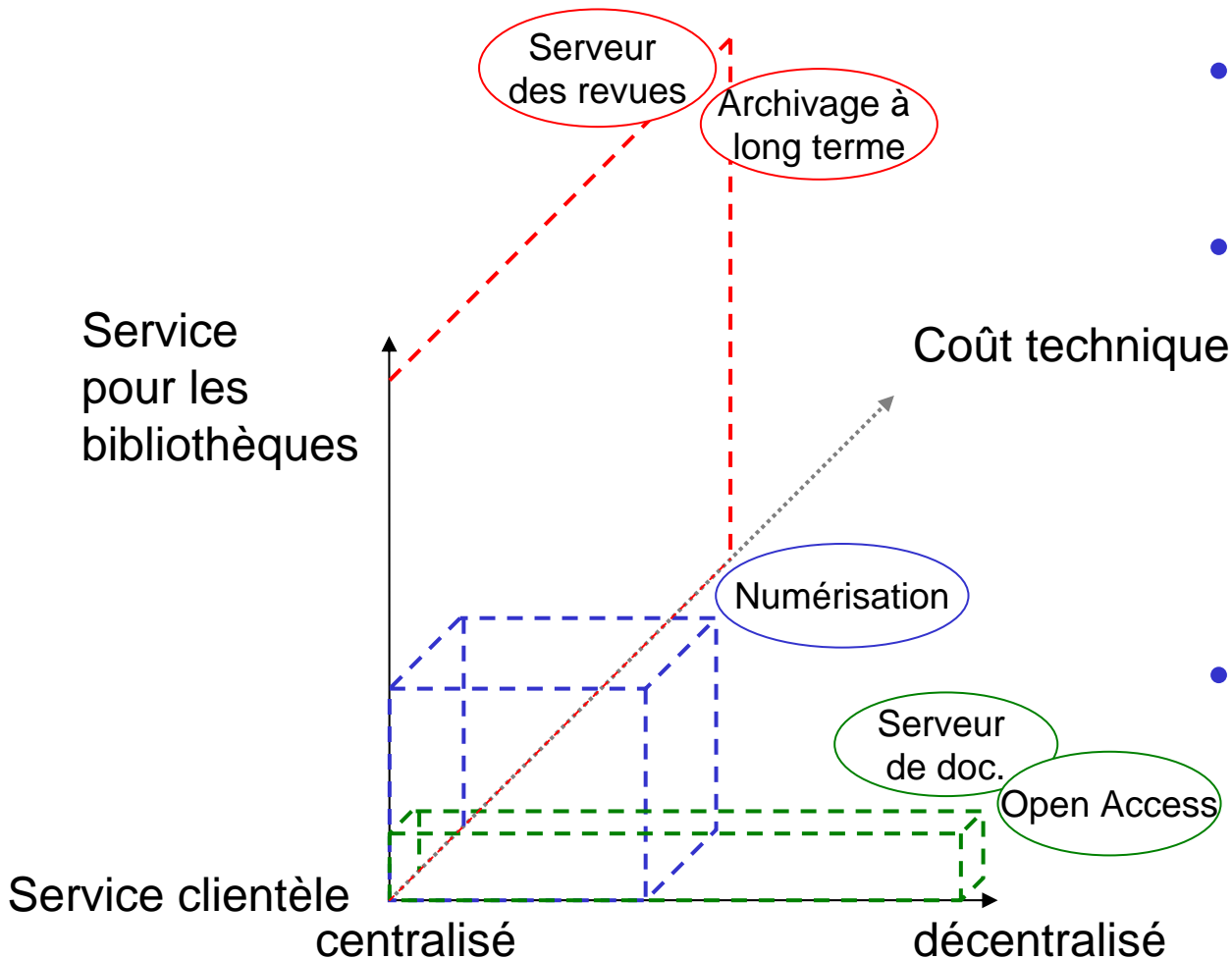
Continuation des  
publications  
universitaires  
traditionnelles sous  
forme électronique:  
*serveur de documents*

Grand intérêt pour  
de nouveaux  
domaines d'activité  
sous le signe de  
*l'Open Access*

Périodiques  
non encore  
numérisés:  
*potentiel*  
et besoins  
*avérés*

*Archivage à long terme*

# Répartition vs. centralisation



- Souhait initial: solutions partagées
- Expérience: répartition difficile à résoudre sur le plan technique et organisationnel; coûts parfois disproportionnés.
- Critères: proximité judicieuse par rapport au client final, coûts (technique, coordination, etc.)

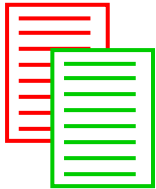
# Serveur de périodiques (données des éditeurs)

---

- Un serveur basé sur DigiTool est en service
  - Raccordement à SWITCHaai (*Authentication and Authorization Infrastructure*) réalisé pour l'essentiel
  - Le filtre pour l'importation d'articles des éditions Elsevier a été mis au point
  - Autorisation d'accès conformément à la licence de 2002 à 2004 en préparation
  - Utilisable comme lieu de stockage pour d'autres contenus
  - Peut être aménagé afin de servir de lieu de stockage (*Repository*) pour certaines institutions ayant leur propre identité institutionnelle (frais supplémentaires)
- Env. 1 million d'articles des éditions Elsevier des années 1995 à 2004 sont prêts à être téléchargés
  - Difficulté: la maison d'édition n'a, elle non plus, aucune expérience et n'a pas de procédure établie pour la transmission de telles données  
→ incertitudes quant aux formats de bande magnétique, de mémoire, etc.

# Serveur de périodiques (données des éditeurs)

---



- Harmonisation des listes de titres du Consortium et de l'éditeur
- les revues ne paraissant plus ou vendues manquent
  - paquets avec plusieurs titres
  - changements de titres...



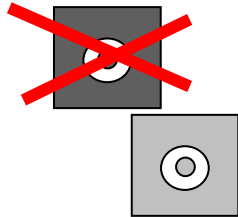
- Livraison de données test soi-disant représentatives sur CD-ROM
- texte intégral en PDF et métadonnées XML ou SGML
  - au moins deux formats différents, en partie non valides
  - néanmoins, filtre pour l'importation de données dans DigiTool réalisé avec succès



- Env. 1 millions d'articles, soit 1,3 téraoctet
- livraison sur bande magnétique
  - production de bandes avec «nos» données: env. 9 mois!
  - offre alternative: livraison de bandes d'archives
- réflexion: avec des données de 2004 → pas trop anciennes → ok

## Serveur de périodiques (données des éditeurs)

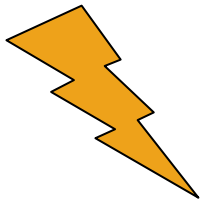
---



Acquisition d'un des derniers modèles de lecteur disponibles sur le marché pour le format de bandes magnétiques de type DLT IV

- les bandes ne sont pas lisibles, parce qu'elles ont été saisies avec un ancien format de sauvegarde

→ lecteur(s) acheté(s) d'occasion



Problèmes de matériel informatique

- alimentation SCSI aujourd'hui avec tension moindre

→ échange de la carte SCSI avec tension plus élevée



- Communication avec l'éditeur nécessitant beaucoup de temps
- Illustre les problèmes pour l'archivage à long terme (veille technologique / « technology watch »)
- Défis ne pouvant être relevés qu'avec l'intervention soutenue des services informatiques de l'ETH-Bibliothek (surtout P. Aliesch)
- Dans cet ordre de grandeur, toute manipulation des données nécessite beaucoup de temps d'un point de vue purement technique: env. trois semaines uniquement pour copier les données de 60 bandes magnétiques sur le serveur!

# Serveur de périodiques (données des éditeurs)

---

- Rôle de la sauvegarde locale des contenus des éditeurs
  - D'une part, il serait judicieux qu'un éditeur prenne lui-même soin de ses données (son capital) le plus longtemps possible – d'autre part, le projet pilote confirme qu'il est déjà difficile, dans des conditions normales, de reprendre des données d'archives classées de manière adéquate.
  - Les frais relatifs à la première reprise des contenus d'une maison d'édition sont élevés.
  - Les frais durant la phase d'exploitation devraient être insignifiants, car les données ne nécessitent pas d'adaptations permanentes → n'est pas encore archivage à long terme!
  - Il n'est pas encore possible de dire pour quels éditeurs une sauvegarde locale s'impose.
  - Des solutions complémentaires, telles qu'une affiliation à Portico ou des expériences avec (C)LOCKSS s'avèrent judicieuses. Reste à définir pour quelles maisons d'éditions et pour quelles bibliothèques tel ou tel projet convient le mieux.

# Serveur de périodiques (données des éditeurs)

---

- Implications pour l'archivage à long terme
  - Les données des éditeurs sont *archivées* par la bibliothèque nationale du pays de l'éditeur, mais ne sont que localement accessibles  
→ Nous nous concentrons sur l'*accessibilité*, toutefois:  
les métadonnées XML et les données en format PDF (avec certaines restrictions) sont appropriées pour l'archivage et, une fois importées, elles sont disponibles localement et indépendantes du support.
  - Compte tenu des problèmes rencontrés actuellement au niveau de la reprise de données dans des conditions «normales», il est peu vraisemblable qu'une maison d'édition puisse encore livrer des données d'archives de manière ordonnée en cas de crise.
  - Il existe une zone grise croissante pour l'accès à des titres qui changent d'éditeur ou ne sont plus édités → le contenu n'est pas livré
  - DigiTool n'est pas encore une archive exhaustive sous sa forme actuelle. Mais, de notre point de vue, il constitue une excellente base pouvant être complétée ultérieurement par des modules, conformément au modèle OAIS.

# Numérisation de périodiques scientifiques

---

- Objectifs

- Périodiques suisses à caractère scientifique  
*de toutes disciplines*
- Améliorer la visibilité sur le plan international et l'accessibilité dans un marché des périodiques toujours plus marqué par la technique numérique
- Pour les périodiques, l'utilisateur profite tout particulièrement de possibilités de recherche performantes (souvent, les articles recherchés sont brièvement comparés au volume complet)
- Plate-forme évolutive

# Numérisation de périodiques scientifiques

---

Déroulement de la numérisation de *tec21/tracés* (direction de projet: Nicole Graf)



## Examen des originaux

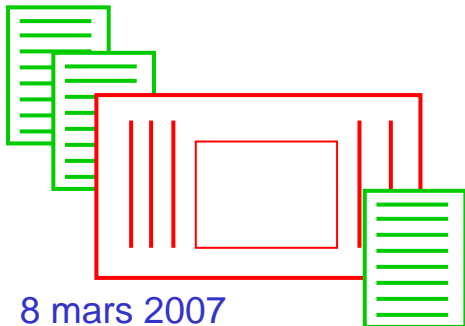
- recherche de volumes/fascicules/feuilles manquants
- appréciation grossière du volume pour l'estimation du temps et des coûts nécessaires
- appréciation détaillée du volume pour le contrôle de qualité ultérieur

## Double existant

- découpage
- séparation des pages de textes (scannées à l'extérieur) et dépliants annexes (scannés en interne)

## Uniquement document original

- scanné à partir du document relié (prêts en interne)



## Assemblage des scans TIFF

- classification suivant le volume
- compression
- gestion des volumes (plusieurs téraoctets)

# Numérisation de périodiques scientifiques



Conversion en série des images

- redénomination des données
- reconnaissance optique de caractères (OCR)
- création de formats de présentation dans divers niveaux de résolution



Saisie manuelle des métadonnées et structuration

→ résultat: structure de document XML

digitalisierte zeitschriften  
part of seals - retro electronic academic library service

Sammlungen > Baugedächtnis Schweiz Online

Baugedächtnis Schweiz Online

Um eine Übersicht über die aktuell verfügbaren Bände einer Zeitschrift zu erhalten, klicken Sie bitte auf das jeweilige Bild der Zeitschrift oder deren Titel.

Titel	Bände	Erscheinungszeitraum	Info
Tracts	106 - 118, 127 - ff	2001 - ff	III
Schweizer Ingenieur und Architekt	97 - 105	1979 - 2001	III
Schweizerische Bauzeitung	1 - 128, 65 - 96	1883 - 1978	III
Die Eisenbahn	1 - 17	1874 - 1882	III
Tracts	127 - ff	2001 - ff	III
Ingenieurs et architectes suisses	105 - 127	1979 - 2001	III
Bulletin technique de la Suisse romande	26 - 104	1900 - 1978	III
Bulletin de la Société vaudoise des ingénieurs et des architectes	1 - 25	1875 - 1899	III

© Eine Dienstleistung des Consortium der Schweizer Hochschulbibliotheken | Impressum - Letzte Änderung 22.01.2007

Intégration dans <http://retro.seals.ch>

- adaptations de fonctions et extensions en raison de la structure complexe du *Bauzeitung* et autres revues ainsi que pour la visualisation adéquate de plans, etc.
- le téléchargement des documents dans le système prend du temps, rien que d'un point de vue technique

# Numérisation de périodiques scientifiques

---

- Etat fin février 2007:
  - dix périodiques (six titres précédents terminés, quatre titres courants) dans deux collections
  - plus de 400 000 pages
  - recherche en texte intégral
  - les collections et les périodiques peuvent être recherchés de manière sélective
- DEMO <http://retro.seals.ch>
  - Titres importants sur le plan scientifique, pas seulement à caractère régional ou local): la plupart financiables dans le cadre du projet.
  - D'autres titres peuvent être numérisés si les coûts sont pris en charge par d'autres instances (p.ex. Mémoire du bâti en Suisse).

# Numérisation de périodiques scientifiques

---

- Titres en préparation (titres précédents inclus)
  - Geographica Helvetica
  - Bulletin suisse de minéralogie et de pétrographie
  - Helvetica Physica Acta
  - situation relativement claire, pas d'objection de la part des éditions Birkhäuser, pour autant que cela les concerne
- Titres en discussion (titres précédents inclus)
  - Revue suisse d'histoire
  - Museum Helveticum
  - vaste intérêt, mais important besoin de clarifier les choses auprès de l'éditeur, de la société et de la Bibliothèque nationale suisse
- Ouverts à d'autres propositions (une liste plus étendue a été soumise à la CBU)

# Numérisation de périodiques scientifiques

---

- Une participation aux coûts d'exploitation a été décidée ou est en discussion avec les gérants ou les éditeurs jusqu'à présent concernés
- Implications pour l'archivage à long terme
  - Les produits numérisés sont des *formats secondaires*, et peuvent donc être remplacés à la longue, lente reconnaissance au tant que form d'archivage  
→ comparaison des coûts
  - Il suffit de conserver les images TIFF et les métadonnées XML. A partir d'elles, on peut en effet reproduire de nouveau le texte intégral et d'autres formats.  
→ TIFF et XML sont des standards ouverts qui sont déjà établis, et qui conviennent à l'archivage
  - A lui seul, le logiciel *n'archive pas*.

# Serveur de documents et Open Access

---

- Pas de *serveur de documents* centralisé, mais un *serveur de données* centralisé
  - Besoins différents sur place – pour que le serveur soit alimenté, cela nécessite un investissement intense des effectifs travaillant sur le site
  - Repository local important pour l'identité de l'institution et pour la confiance des auteurs locaux
  - Indexation du texte intégral partiellement possible, mais pas envisagée
  - En principe, extension au tant que repository local possible, si nécessaire
  - Alternativement: Uniquement archivage en arrière-plan sur serveur des revues (DigiTool)

# Serveur de documents et Open Access

---

- Soutenir le principe d'Open Access en diffusant des informations
- Objectif
  - Meilleure visibilité, qui ne soit pas liée à une institution en particulier, mais qui permette une vision nationale
  - Idéalement accès par matière (vgl. repositories par discipline)  
→ **information doit être saisie localement et mise à disposition via OAI-PMH**
  - Incitation à créer des normes, conformément au modèle OAI-PMH
  - Echanges d'informations

→ DEMO (TEST!), später: <http://meta.seals.ch/>

# Serveur de documents et Open Access

The screenshot shows the homepage of the 'metadaten repository' (part of seals - swiss electronic academic library service). The page features a search bar with 9,257 records, a navigation menu on the left, and two columns of filters: 'Narrow by collection' and 'Focus on'. The 'Narrow by collection' section includes checkboxes for Preprints (111), Postprints / Journalarticles (915), Theses (1,017), Dissertations (145), Conference papers (0), Reports / Studies (19), and Mixed basket (7,050). The 'Focus on' section includes a 'Navigate by insitutions' section with links for Universities (2,162), Universities of applied sciences (45), ETH domain (7,050), and others (0); and a 'Navigate by domain' section with links for 000: Computer science, information & general works (132), 100: Philosophy & psychology (621), 200: Religion (70), 300: Social sciences (472), 400: Language (0), 500: Science (911), 600: Technology (0), 700: Arts & recreation (0), 800: Literature (0), and 900: History & geography (1).

metadaten repository  
part of seals - swiss electronic academic library service

Home

KONS-Testserver lokal 0.92.0

Search 9,257 records for:

Search [input field] any field [dropdown]

Search Browse

Search Tips :: [Advanced Search](#)

Narrow by collection:

- [Preprints](#) (111)
- [Postprints / Journalarticles](#) (915)
- [Theses](#) (1,017)
- [Dissertations](#) (145)
- [Conference papers](#) (0)
- [Reports / Studies](#) (19)
- [Mixed basket](#) (7,050)

Focus on:

**Navigate by insitutions** (9,257)

- [Universities](#) (2,162)
- [Universities of applied sciences](#) (45)
- [ETH domain](#) (7,050)
- [others](#) (0)

**Navigate by domain** (2,207)

- [000: Computer science, information & general works](#) (132)
- [100: Philosophy & psychology](#) (621)
- [200: Religion](#) (70)
- [300: Social sciences](#) (472)
- [400: Language](#) (0)
- [500: Science](#) (911)
- [600: Technology](#) (0)
- [700: Arts & recreation](#) (0)
- [800: Literature](#) (0)
- [900: History & geography](#) (1)

# Serveur de documents et Open Access

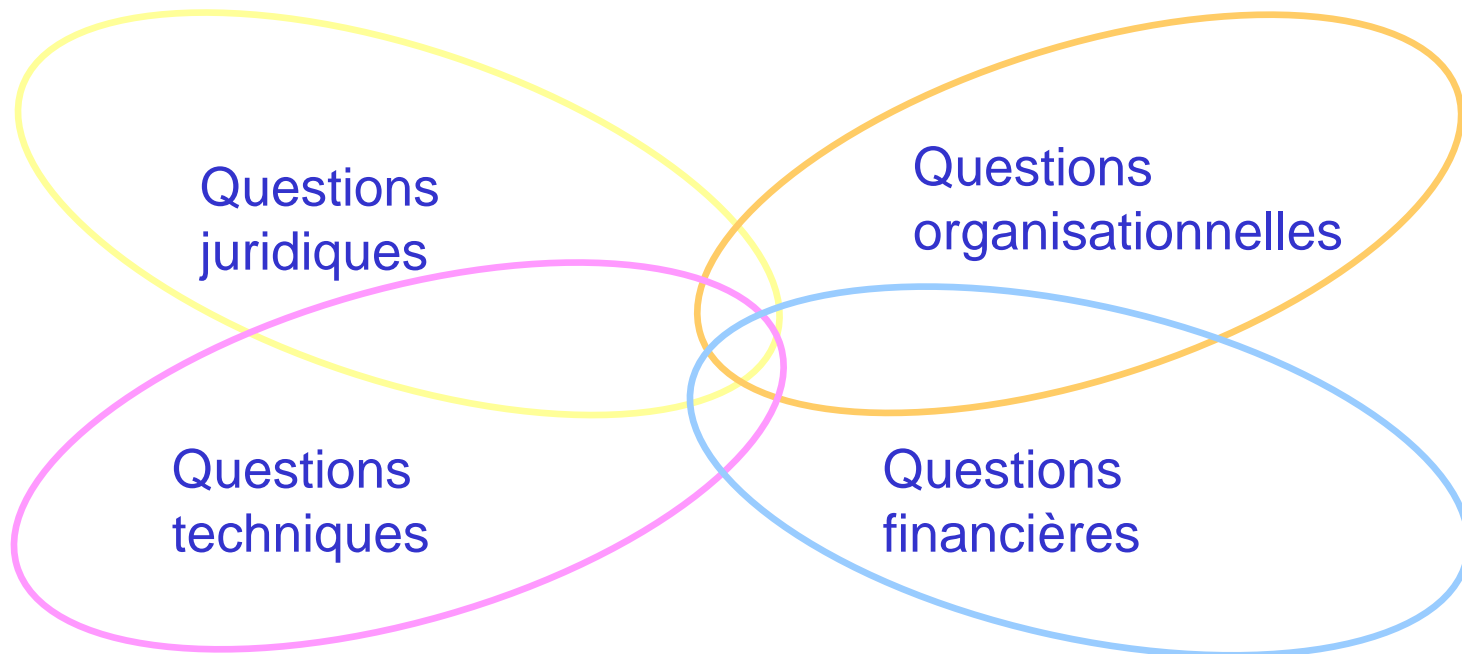
---

- Implications pour l'archivage à long terme
  - Aucun archivage centralisé  
(exception: thèses de doctorat conservées à la BNS)
  - Alternatives pour l'archivage
    - Sauvegarde locale correspondant aux besoins et, si nécessaire, archivage
    - Utilisation de DigiTool comme lieu de stockage central «sombre», à l'arrière-plan (en plus du serveur local)  
→ archivage à long terme pas encore implémenté
    - Utilisation de DigiTool comme serveur de documents productif dans l'organisation de la bibliothèque en question (coûts supplémentaires)  
→ archivage à long terme pas encore implémenté
    - Utilisation d'une option de «consigne» auprès de la BNS

# Accessibilité vs. archivage à long terme

---

Définition: *“The maintenance of digital material over the long-term with a view to ensuring continued accessibility.”*<sup>1</sup>



# Archivage à long terme

---

## Archivage à long terme

- «Conserver la substance»
  - «Préserver le train de bits»
    - «Préserver le flux de données»
    - «Bits in – bits out»
- *surtout migration des médias et des logiciels*

- «*Disponibilité durable*»
- «Utilisation garantie»
- «Maintien du décryptage»
  - *migration de formats, émulation...?*

# Archivage à long terme

---

*«Par “archivage à long terme”, on entend dans ce contexte autre chose que le simple respect des prescriptions légales concernant les délais pendant lesquels les données structurées sous forme de tableaux, importantes sur le plan fiscal, doivent être disponibles.*

*L’expression «long terme» décrit un laps de temps qui n’est pas clairement défini, pendant lequel des transformations essentielles non prévisibles, au niveau des technologies et de la culture sociale, surviendront. Or, celles –ci vont révolutionner complètement la forme ainsi que les conditions d’utilisation des ressources numériques dans des cycles de développement fulgurants. [...] Pour ce qui est de la conservation de la substance, le «long terme» ne signifie pas l’octroi d’une déclaration de garantie durant plus de cinq ou cinquante années, mais le développement responsable de stratégies qui permettront de faire face aux changements permanents induits par l’évolution du marché de l’information.»*

Ute Schwens, Hans Liegmann in: *Grundlagen der praktischen Information und Dokumentation* / comm. de Klaus Laisiepen, Ernst Lutterbeck et Karl-Heinrich Meyer-Uhlenried. - 5., édition entièrement remaniée. - Munich: Saur. vol. 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 2004. pp. 567- 570.

(La numérotation des paragraphes ne correspond pas à celle de l’édition imprimée)

# Archivage à long terme

---

- L'archivage durable n'est garanti par aucun système informatique, mais par des institutions et leurs collaborateurs, qui utilisent les systèmes appropriés.
- Les systèmes d'archivage existant sur le marché sont
  - axés sur le besoin des entreprises qui sont tenues de conserver leurs documents de manière sûre, c.à.d. inaltérable, pour une période donnée.  
(Les fournisseurs n'ont parfois aucune idée des exigences de ceux qui souhaitent conserver des objets qui pourront être décryptés pendant une durée indéterminée au moyen de modifications ciblées. Par conséquent, il n'existe aucun système qui complètement soutienne une telle tâche.)
  - des systèmes pour des serveurs de documents ou des repositories qui sont plus ou moins bien conçus/préparé pour le stockage à long terme.
- Dans une étape intermédiaire, des systèmes prenant déjà en charge certaines fonctions, conformément au modèle de référence OAIS, et susceptibles d'être développés seront utilisés.

# Archivage à long terme




---

## Où en sommes-nous aujourd'hui?

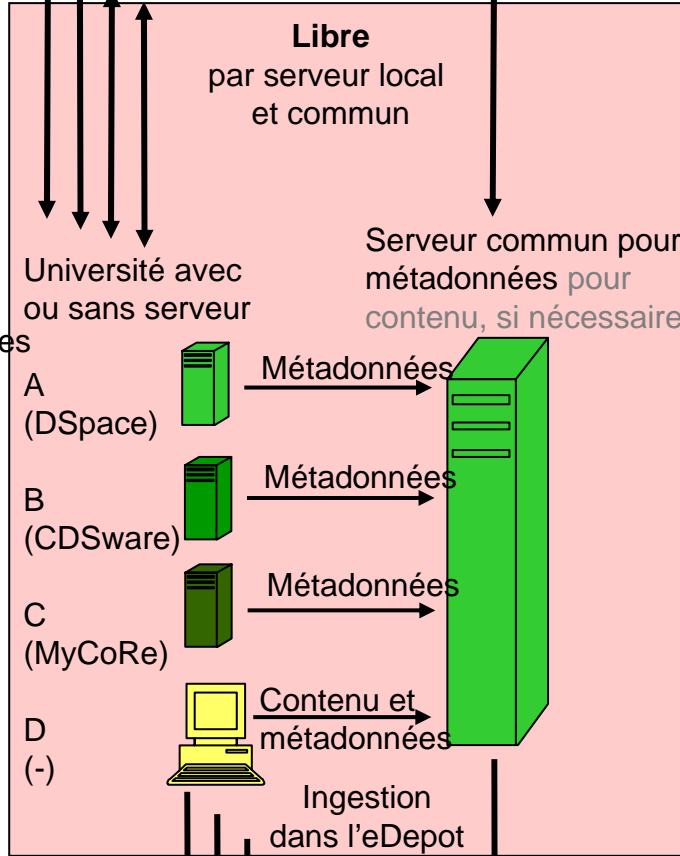
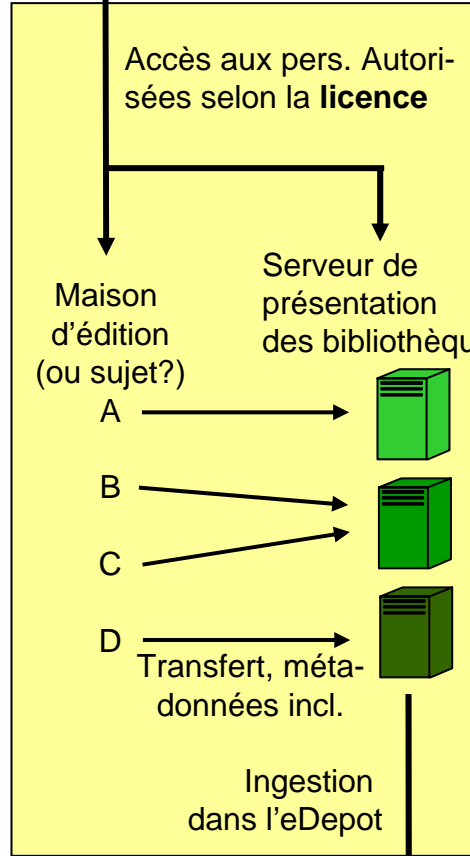
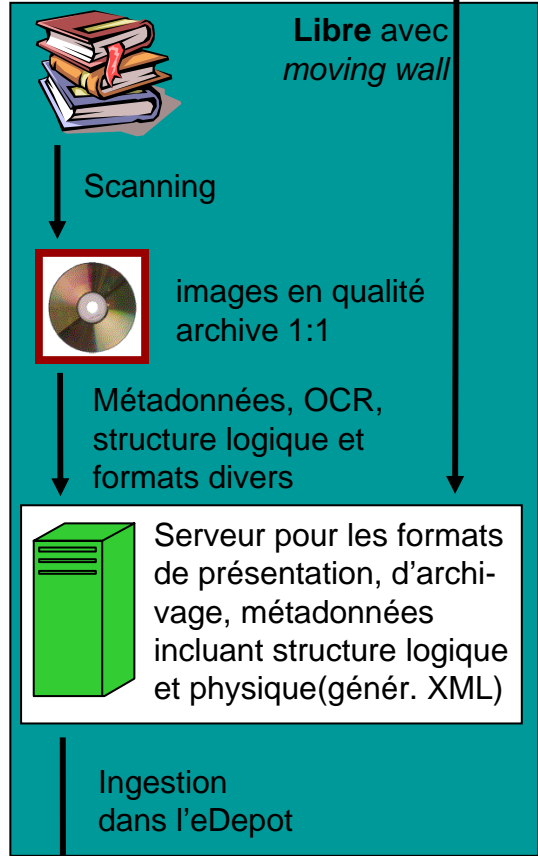
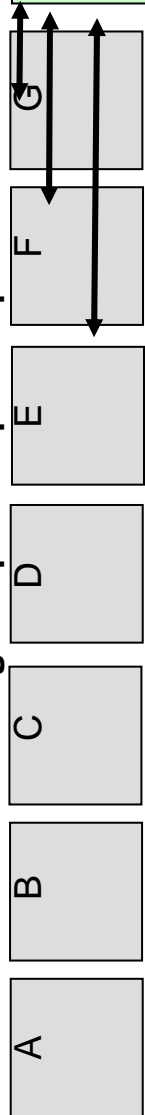
- Dans l'état actuel des choses, on peut et on doit préparer l'archivage à long terme sur le plan local en surveillant soigneusement et continuellement les données existantes. Cela permet de franchir le cap pour des périodes plus courtes.
- Des instructions et des aides seront encore élaborées pendant la durée du projet et mises à disposition des bibliothèques. Pour les appliquer, elles auront besoin de temps (d'heures de travail). Le plus grand nombre d'opérations possibles doit pouvoir être intégré dans des processus de routine. Cela vaut notamment pour l'évaluation et la sélection des objets numériques et de leurs formats. Or, cette tâche ne peut et ne doit pas être centralisée et assumée indépendamment des bibliothèques. La BNS est responsable de l'introduction de procédures pour l'attribution d'URN par les bibliothèques universitaires suisses (en 2007).

# Les projets partiels

<http://www.seals.ch>

**Titres imprimés:** accès OPAC et PSP  **Numérisation:** accès de bases de données bibliographiques externes, d'OPAC (niveau revue), avec recherche séparée en texte intégral, champs, index (<http://retro.seals.ch>)  **Contenus sous licence:** accès depuis OPAC (niveau revue), de listes d'eJournals, de bases de données bibliographiques externes et autres interfaces  **Littérature grise:** accès depuis OPAC et par d'autres interfaces (index par institutions, par disciplines, par types de doc.) avec rech. par champs et par index

Fonds d'archivage de périodiques imprimés



**eDepot / serveur d'archives.** Exemple de concept: projet KOPAL de la DDB, où DIAS (IBM, BR de la Haye) doit être adapté pour une utilisation partagée. Répartition par sujet, par maison d'édition ou autre? Plusieurs des fonctions des serveurs marqués en vert et des tâches d'archivage pourraient éventuellement être assumées par *un seul* système.

# Partenaires

---

- Périodiques de maisons d'édition
  - Editeurs commerciaux
  - Bibliothèques universitaires
- Serveur de documents et Open Access
  - Bibliothèques et universités → groupe-cible: auteurs
- Numérisation de périodiques
  - Organes d'édition et éditeurs
  - Bibliothèques universitaires
- Archivage à long terme
  - Bibliothèque nationale suisse comme bibliothèque d'archive centralisée pour des contenus relatifs à la Suisse (projet e-Helvetica)
  - Bibliothèques universitaires
  - Services informatiques des établissements universitaires

# Perspectives

---

- Périodiques de maisons d'édition
  - Soutien différent suivant les éditeurs
  - Intérêt plus ou moins grand suivant les bibliothèques
  - Quelle priorité donner à la sauvegarde locale en Suisse?
- Serveur de documents et Open Access
  - Soutien formel (Conférence des recteurs, Fonds national suisse (promotion de la recherche), Conférence des Académies scientifiques suisses, diverses universités)
  - Mise en œuvre pratique laborieuse, travail de persuasion nécessaire auprès des auteurs, devant être réalisé sur place
  - Des instruments utilisables sur place sont souhaités
  - Un «Centre de compétences pour OA et les serveurs de documents» aurait-il en outre un rôle à jouer?

# Perspectives

---

- Numérisation des périodiques
  - Modèles pour l'exploitation durable et le financement en discussion
    - au besoin, prestation plus complète / hébergement des titres courants?
- Archivage à long terme
  - Conscience du problème heureusement plus grande. Aucune solution monolithique globale, mais des éléments qui se complètent qualitativement et quantitativement (cf. répartition entre la BNS et le Consortium)
  - Indication d'instruments facilement utilisables sur place pour les opérations de routine décentralisées qui sont nécessaires (validation des formats, etc.)
  - Au besoin, lieu de stockage extensible existant
    - Dans quelle mesure et sous quelle forme des prestations centralisées sont-elles requises et seront-elles financées durablement?

Insatisfaisant: question du soutien financier durable

# Bilan

---

- Le projet E-Archiving est orienté en plus de l'archivage à long terme: numérisation, Open Access, sauvegarde de périodiques commerciaux  
→ Judicieux en principe, mais domaine très vaste et dynamique
- Création d'offres et de prestations pour les bibliothèques et le client final  
→ néanmoins, pratiquement pas de contact avec le client final
- Importance d'avoir des partenaires solides, pour ancrer les offres au-delà de la durée du projet
- Défis supplémentaires posés par l'existence «virtuelle» du Consortium  
→ Institution de services durable, même autonome si nécessaire, comme alternative? Uniquement envisageable en étroite relation avec les bibliothèques.

# Merci de votre attention!

---

Dr. Matthias Töwe  
Consortium des bibliothèques universitaires suisses  
c/o ETH-Bibliothek  
Rämistrasse 101  
CH-8092 Zurich  
0041-(0)44 632 60 32  
[matthias.toewe@library.ethz.ch](mailto:matthias.toewe@library.ethz.ch)  
<http://lib.consortium.ch>