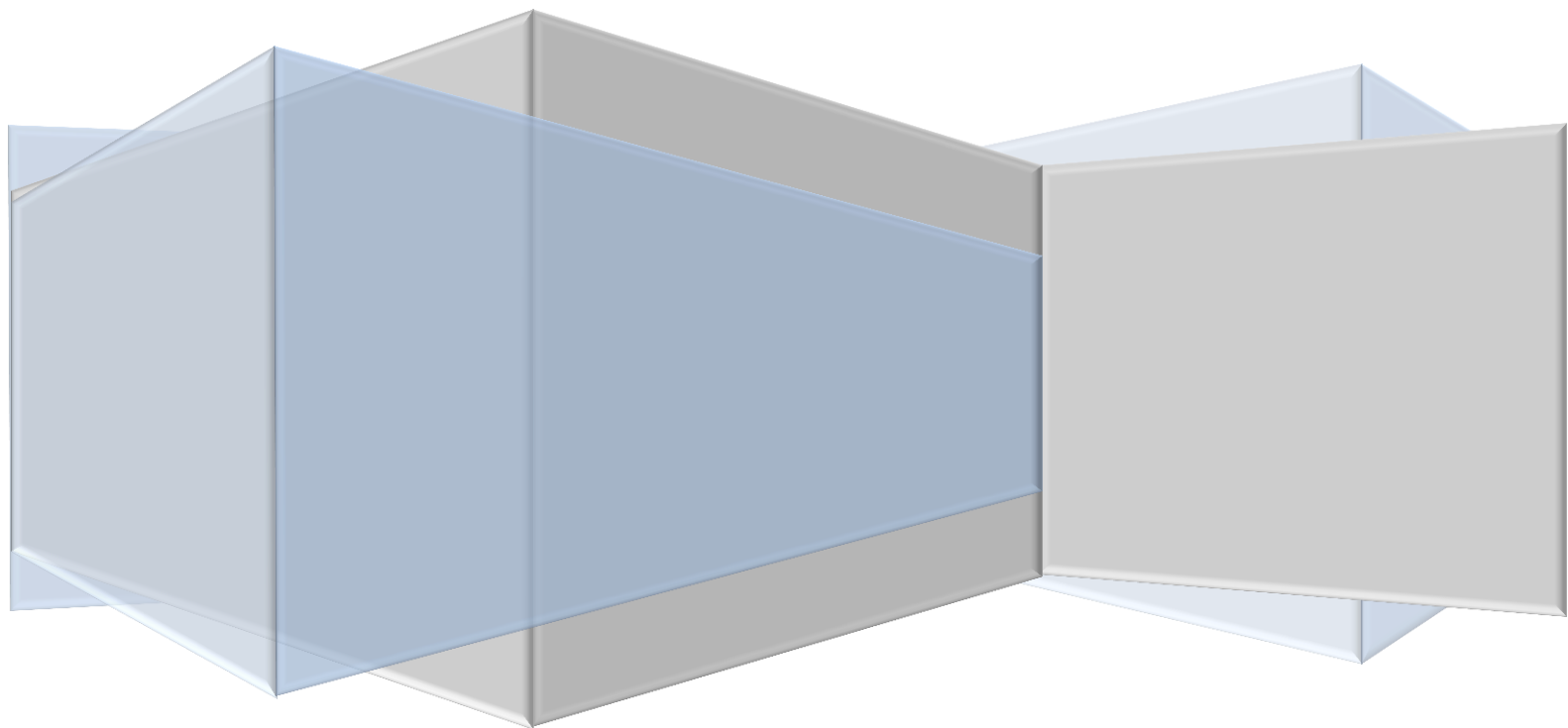


Archivage papier et/ou électronique et accès pérenne aux ressources documentaires électroniques en texte intégral en Fédération Wallonie-Bruxelles

Aude Alexandre



L'objet du présent rapport est d'éclairer le choix d'une stratégie de préservation et d'accès à long terme aux e-journaux électroniques pour les bibliothèques universitaires de la Communauté française de Belgique, afin qu'elles puissent aborder ces questions de manière globale et concertée, en analysant les solutions existantes et en identifiant les meilleures pistes – et leur limites, en termes de coûts, de garanties et de service.

Table des matières

1	Introduction.....	5
1.1	La conservation pérenne des e-journaux : une question urgente	5
1.2	Les problématiques de la conservation des e-journaux.....	6
1.3	Les concepts de l'accès pérenne aux e-journaux.....	8
1.4	Collaboration et mutualisation des ressources.....	11
2	Cadre légal et clauses contractuelles	11
2.1	Cadre légal.....	12
2.2	Les clauses contractuelles des abonnements	12
2.2.1	Droits d'accès post-abonnement et d'accès continu	14
2.2.2	Garanties d'archivage pérenne	20
2.2.3	Les bouquets d'archives	21
2.3	Conclusions et recommandations	21
2.3.1	Négociations avec les éditeurs	21
2.3.2	Gestion de l'information	23
3	Les solutions d'archivage à long terme des e-journaux	25
3.1	Critères minimaux et typologie	26
3.2	Les solutions sans mandat spécifiquement dédié.....	27
3.2.1	Les dépôts institutionnels.....	28
3.2.2	Les logiciels de préservation numérique (commerciaux et Open Source).....	28
3.2.3	Les plateformes d'archives.....	29
3.3	Les solutions avec mandat spécifiquement dédié	30
3.3.1	LOCKSS.....	30
3.3.2	Portico	33
3.3.3	CLOCKSS.....	35
3.3.4	e-Depot.....	36
4	LOCKSS, CLOCKSS et Portico : évaluation	36

4.1	Critères d'évaluation	37
4.1.1	Les garanties d'archivage pérenne.....	37
4.1.2	L'accès	39
4.1.3	La couverture.....	39
4.1.4	Le degré de contrôle.....	40
4.1.5	Les coûts.....	40
4.2	Evaluation et comparaison de LOCKSS, CLOCKSS et Portico	42
4.2.1	Garanties d'archivage pérenne	42
4.2.2	Accès.....	49
4.2.3	Couverture.....	53
4.2.4	Contrôle.....	72
4.2.5	Coûts.....	74
4.2.6	Conclusion et recommandations.....	89
5	Les initiatives nationales ou régionales.....	91
5.1	Les licences nationales	91
5.2	Les initiatives des bibliothèques nationales liées au dépôt légal et les services de tiers archivage	92
5.3	Les réseaux de préservation nationaux et les initiatives nationales liées à LOCKSS.....	94
5.4	Les stratégies nationales ou consortiales à l'étude.....	95
5.4.1	Suisse	95
5.4.2	Allemagne.....	96
5.5	Conclusion	96
6	L'archivage papier partagé.....	97
6.1	Aspects économiques.....	97
6.1.1	Éléments de coût de la préservation papier	97
6.1.2	Éléments de comparaison des coûts de la préservation papier avec les coûts de la préservation électronique via LOCKSS	99
6.2	Autres aspects	101

6.3	Conclusion et recommandations.....	101
7	Scénarios envisageables pour la BICfB	102
7.1	Une échelle de solutions d'accès pérenne électronique	102
7.1.1	Scénario prioritaire.....	102
7.1.2	Scénarios complémentaires	103
7.1.3	Synthèse	104
7.2	Recommandations.....	105
8	Conclusion	109
9	Bibliographie.....	112

1 Introduction

1.1 La conservation pérenne des e-journaux : une question urgente

Le passage de la documentation papier à la documentation électronique initié dans les années 90 et exponentiel depuis lors, s'il a constitué une révolution en terme d'accès pour les usagers (accès immédiat et à distance, augmentation importante du volume des collections, fonctionnalités de recherche avancée), et en matière de gestion de l'espace (gain de place et réaménagement des espaces vers d'autres fonctions), a posé et pose encore avec une acuité grandissante différentes questions immédiates en termes d'acquisition (surcoût TVA et modèles commerciaux des éditeurs impliquant des augmentations difficiles à supporter pour les bibliothèques) et de gestion (flux et outils de gestion des abonnements, accès, données d'utilisation...), mais aussi, plus fondamentalement et de manière moins immédiatement visible, en termes d'accès et de conservation. La préservation des journaux électroniques ne va en effet pas de soi et pose des questions juridiques et techniques qui ne sont pas résolues actuellement.

Les bibliothèques universitaires qui ne définissent pas rapidement une stratégie pour aborder cette question risquent de se retrouver dans une impasse : le principe de précaution limite en effet le passage aux abonnements e-only tant que leur préservation à long terme n'est pas garantie, alors que presque tous les autres facteurs y mènent, de facto¹ : l'offre des éditeurs, la demande des usagers, la structure de l'espace². Le problème n'est pas mineur vu le volume de documents et l'investissement financier concernés³. Pour la BICfB, le montant des acquisitions électroniques a dépassé depuis 2011 le montant des acquisitions papier, avec le passage à l'e-only des bouquets d'e-journaux Wiley et Elsevier, et l'évolution en ce sens continue.

¹ Un rapport du JISC publié en novembre 2010 (E-only scholarly journals: overcoming the barriers | Research Information Network, novembre 2010. <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/e-only-scholarly-journals-overcoming-barriers>), cite p. 8 un rapport de l'Association of Learned and Professional Society Publishers (ALPSP) qui établissait en 2008 déjà que plus de 90% de la littérature scientifique était disponible en ligne. Un autre facteur important de limitation du passage à l'électronique seul reste la question de la TVA, plus élevée sur l'électronique (21%) que sur le papier (6%).

² La pénurie de l'espace de stockage pour la préservation des collections de périodiques imprimés, qui continuent de croître en parallèle aux collections électroniques, augmente la pression vers le passage à l'électronique. L'initiative de préservation partagée en cours au sein des institutions de la BICfB (via l'initiative GT-CPP du CIUF) comme un peu partout en Europe et dans le monde, permet des économies d'espace [réelles ou potentielles], mais est une démarche de longue haleine, car coûteuse en ressources humaines.

³ Le coût d'abonnement aux ressources électroniques négociées en consortium pour les institutions de la BICfB dépassait en 2013 les 5 millions d'euros.

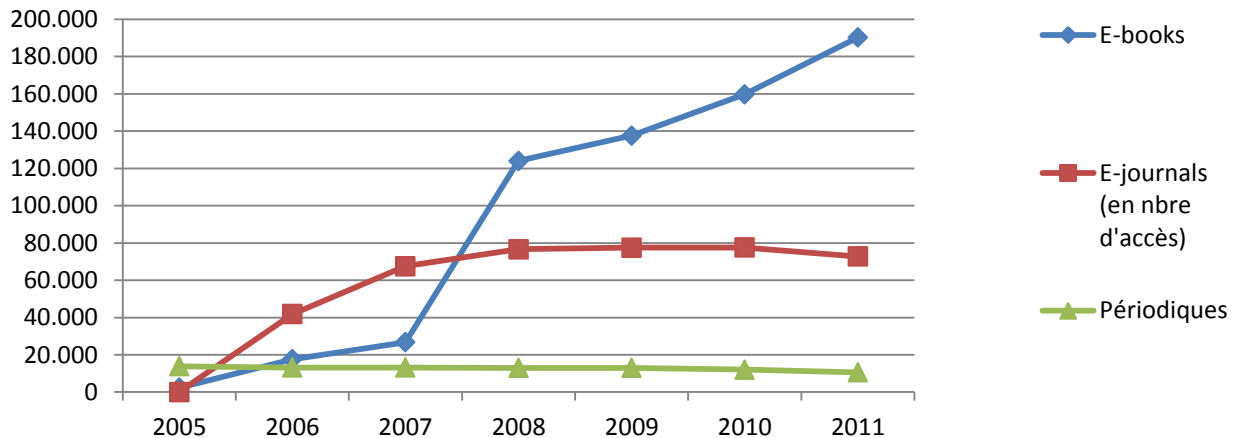


Figure 1 - Statistiques CIUF⁴ – Evolution du nombre de souscriptions aux ressources électroniques (E-books, E-journals) et imprimées (Périodiques), pour l'ensemble des institutions de la BICfB.

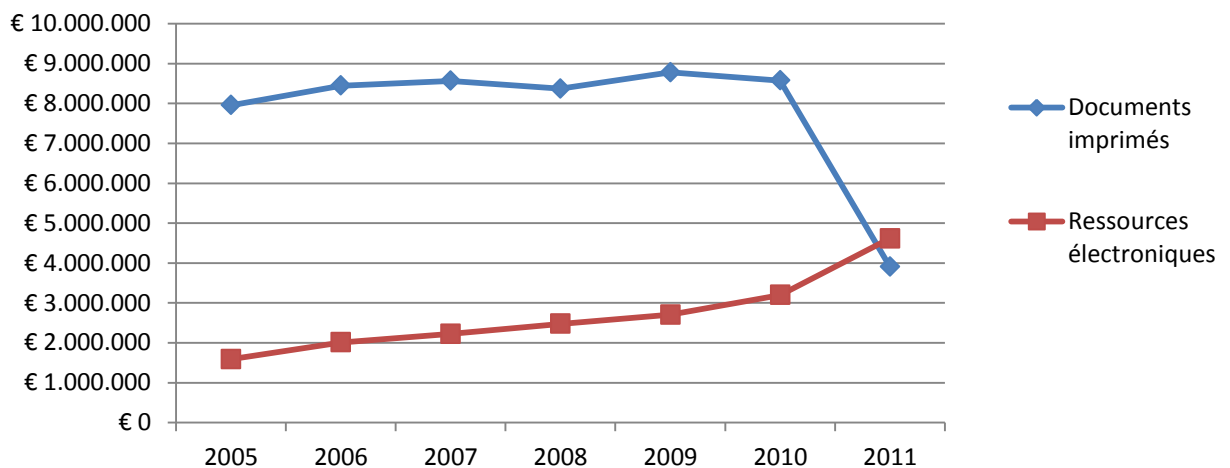


Figure 2 - Statistiques CIUF – Evolution des dépenses en ressources électroniques et imprimées, pour l'ensemble des institutions de la BICfB.

C'est la raison pour laquelle la BICfB a commandité la présente étude de manière à identifier les meilleurs choix en termes de stratégie de préservation et d'accès à long terme aux périodiques électroniques pour les bibliothèques universitaires de la Communauté française de Belgique, à en identifier au mieux les coûts, les risques et les garanties.

1.2 Les problématiques de la conservation des e-journaux

Si la gestion des collections de périodiques électroniques présente des similitudes avec la gestion des collections de périodiques imprimés, leur préservation et accès à long terme pose des problématiques concrètes spécifiques qui à la fois se distinguent de celles de la conservation des périodiques imprimés et dépassent celles de la conservation des données numériques.

Dans le monde des publications imprimées, la stratégie de garantie d'accès à court et moyen terme et de préservation à long terme des collections imprimées relève d'une part des choix des bibliothèques en termes de politique de gestion des collections et d'autre part de la qualité de stockage des documents. L'achat des documents auprès des éditeurs ou des agents implique le transfert physique des volumes au sein des collections des bibliothèques et, de manière tacite, le

⁴ Conseil interuniversitaire de la Communauté française (<http://www.ciuf.be>).

droit à en disposer comme bon leur semble, dans le respect du droit d'auteur. L'accès à long terme est donc garanti si les conditions de stockage des documents sont adéquates (sauf action a posteriori d'élimination⁵) et, dans une perspective plus large, si le nombre de copies et leur répartition géographique sont suffisants. En pratique, on sait que ces critères ne sont pas une évidence, même s'ils sont bien connus et gérés à différents niveaux (bibliothèques universitaires, bibliothèques nationales et dépôt légal). La gestion régulière des collections – passant aussi par la conservation partagée – est devenue cruciale avec la limitation de l'espace disponible auquel sont confrontées la majorité des bibliothèques universitaires et nationales. Toutefois, même en cas d'absence de gestion, les chances de conservation des volumes ne sont pas nulles.

Le passage à l'électronique de la documentation scientifique la soumet aux problématiques de la conservation de l'information numérique. L'information numérique est médiante, concentrée, labile et évolutive, ce qui pose des problèmes accrus pour sa préservation. Ces problèmes et leurs solutions possibles sont détaillés plus bas (cf. point 4.1.1), mais on peut déjà souligner que dans le cas de l'électronique, les points de défaillance sont beaucoup plus sensibles que pour la conservation du papier et les délais de gestion et d'action beaucoup plus réduits. Si la conservation des documents papier peut être envisagée sur des centaines d'années, celle des données numériques, dans le cadre des solutions actuelles, n'est prévisible, au mieux, que sur quelques dizaines d'années.

En outre, la question de la conservation pérenne de la documentation électronique ne se limite pas à ces problématiques. Le passage des abonnements aux périodiques du papier à l'électronique s'est accompagné, en effet, de modifications majeures en termes de modèle commercial, de garanties et responsabilités, qui sont autant de facteurs délicats à prendre en compte dans la stratégie de conservation pérenne :

1. **un déplacement et une dilution des responsabilités (perte de contrôle)** : les documents, s'ils sont accessibles via l'Internet au moyen d'une identification sécurisée liée à la bibliothèque, restent désormais (que la bibliothèque achète ou s'abonne aux contenus) stockés sur les serveurs des producteurs ou fournisseurs de contenus, où ils bénéficient par ailleurs – et de plus en plus – de nombreuses et utiles fonctionnalités dynamiques (hyperliens, statistiques...);
2. **une orientation vers des modèles d'accès au détriment de l'acquisition (perte de droits)** : les éditeurs, pour tirer profit des possibilités de l'électronique, ont modifié – et multiplié – leurs modèles commerciaux, en les centrant sur la notion d'accès et de service, désormais distincte de la propriété des données. Les modèles *big deals*, offrant un accès à un nombre de ressources beaucoup plus important pour un surcoût réduit par rapport aux abonnements proprement dits, en sont le meilleur exemple. Ceci a entraîné un glissement progressif de la notion d'achat (ou fourniture), vers la notion d'accès (ou abonnement) aux contenus. Les bibliothèques paient désormais pour un accès aux contenus limité dans le temps, qui peut, secondairement, s'accompagner de garanties d'accès à plus long terme (cf. infra chapitre 2.2). Le problème est que cette modification des modèles ne s'est pas toujours accompagnée d'une réduction de coûts correspondante ;
3. **une augmentation exponentielle des volumes (et des coûts) (perte de visibilité)** : en particulier via les *big deals*, les collections des institutions ont évolué de manière exponentielle – mais les coûts également. Ce phénomène exacerbe la distinction entre accès et garanties de conservation, et l'urgence du problème de l'accès à long terme (les sommes

⁵ Cet effet est à souligner en particulier dans le contexte de la Communauté française, où l'élimination des collections ne se fait qu'à corps défendant, a posteriori. La situation est semble-t-il différente en France par exemple, où la rotation des documents est beaucoup plus rapide.

investies dans les publications électroniques représentant des montants de plus en plus considérables), mais met aussi en évidence la problématique de la gestion de l'information sur les collections électronique des institutions.

Les risques liés à ce nouveau modèle en matière d'accès à la documentation et de préservation du patrimoine sont donc majeurs puisque :

1. **l'accès, à court, moyen ou long terme, relève désormais, en premier lieu, des droits et garanties accordés par l'éditeur sur ces données aux bibliothèques.** Ceci implique de la part des institutions un travail de négociation et dans certains cas d'adaptation en profondeur des modèles commerciaux actuels, et nécessite une clarification des rôles et responsabilités dans la question de la conservation pérenne. Ce point est déterminant dans la mise en place de solutions d'archivage pérenne électroniques, et les distingue des solutions partagées mises en place pour les périodiques imprimés, où le transfert de droits, la propriété étant acquise à l'origine auprès de l'éditeur, concerne uniquement les bibliothèques ;
2. **le processus d'archivage pour conservation à moyen et long terme ne peut plus être une décision et un investissement a posteriori⁶,** mais nécessite des investissements et décisions de préservation a priori, et en continu⁷, tant pour obtenir les contenus (ou les droits d'accès aux contenus) que pour préserver ces contenus. Il s'agit évidemment d'une difficulté, non seulement parce que nous ne disposons pas encore d'une perspective historique suffisante sur la préservation des données numériques, mais aussi sur le plan de la gestion des collections. Les statistiques d'usage et citations n'arrivent qu'a posteriori, et toutes les informations nécessaires pour la sélection des documents ne sont généralement pas (encore) centralisées. Les investissements (matériels et humains) destinés à la préservation numérique doivent également être consentis a priori, lors la souscription ou de l'achat de la documentation⁸. La croissance exponentielle des ressources et fonctionnalités de la documentation électronique, l'évolution rapide des formats complexifient le problème.

1.3 Les concepts de l'accès pérenne aux e-journaux

Même si les deux notions sont imbriquées et complémentaires, il est fondamental pour identifier les enjeux et évaluer les solutions de préservation à long terme de distinguer, en fonction de l'échelle du temps et de la réalité juridico-technique qu'elles recouvrent, deux notions distinctes dans le concept d'accès pérenne⁹.

⁶ "Paper as the medium for the world's memory has one great advantage; it survives benign neglect well."
Rosenthal, D. S., Rosenthal, D. C., Miller, E. L., Adams, I. F., Storer, M. W., & Zadok, E. (2012). *The economics of long-term digital storage*. In Memory of the World in the Digital Age Conference, Vancouver, BC. Consulté à l'adresse http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Rosenthal_et_al_27_B_1330.pdf, p. 1.

⁷ Voir par exemple la déclaration commune, en 2005, de bibliothécaires académiques à la Mellon Foundation *Urgent Action Needed to Preserve Scholarly electronic journals* (<http://old.diglib.org/pubs/waters051015.htm>), et, plus récemment, la notion de COI (Cost of Inaction) discutée dans le post de Paul Weatherley sur le blog du projet W4C (<http://4cproject.eu/news-and-comment/4c-blog/16-guest-blog-digital-lifecycles-and-the-costs-of-curation-by-paul-wheatley>). Les exemples de données électroniques perdues ne manquent pas ; à titre illustratif, voir <http://www.atlasofdigitaldamages.info>

⁸ Dans le cas de Portico, l'investissement est consenti a priori mais les institutions individuelles peuvent y participer a posteriori. Voir point 3.3.1.2.

⁹ Ces distinctions ont été établies déjà par le JISC en 2007 (Jones, M. (s. d.). *e-Journals: Archiving and Preservation Briefing paper*. Consulté 29 juillet 2013, à l'adresse http://sitecore.jisc.ac.uk/publications/briefingpapers/2007/pub_ejournalspreservationbp.aspx), de manière un

- **Accès post-abonnement et accès continu (*post-cancellation access (PCA) et perpetual access*) :** l'accès post-abonnement est la garantie d'accès (par les bibliothèques) aux contenus souscrits après la fin de l'abonnement à ces derniers. Cet accès est avant tout fonction des garanties contractuelles négociées (ou non) par l'institution avec l'éditeur. L'accès perpétuel ou continu est souvent défini de manière plus large que l'accès post-abonnement, en incluant d'autres cas de figure liés à la pérennité des droits d'accès aux contenus souscrits : transfert des droits d'un éditeur à un autre ou cessation de publication sans transfert de droits.
- **Préservation ou archivage pérenne (*long-term preservation*) :** la question de l'accès à long terme envisage, au-delà des questions contractuelles liées à la durée de l'accès aux contenus, les garanties de conservation des contenus sur le (très) long terme telles que définies par l'OAIS (voir point 3.1.)¹⁰, incluant la couverture des risques liés à la technologie (pérennité des formats et des supports, sécurité des serveurs), mais aussi les risques sociaux et économiques (dans le cas des e-journaux, faillite de l'éditeur sans rachat par une autre société par exemple).

On désigne généralement dans la littérature les événements particuliers rendant inaccessible le contenu sur la plateforme originale de publication par la notion d' « **événement déclencheur** » ou *trigger event*.

Il est important de souligner que les questions de l'accès à long terme se posent indépendamment de l'abonnement aux contenus et des questions de droits d'accès : **les risques existent même pendant la durée de l'abonnement**. L'inclusion des clauses concernant les droits d'accès (en particulier une clause de post-abonnement) dans les licences est une condition nécessaire pour l'accès à long terme et la rentabilisation des coûts consacrés à l'achat des ressources électroniques, mais n'est pas une condition suffisante pour garantir l'accès pérenne aux ressources. L'accès à long terme suppose à la fois des garanties juridiques et des garanties techniques, économiques et sociales, qui ne peuvent, de facto, être assumées unilatéralement par les éditeurs.

En outre, dans le contexte de la garantie de préservation des e-journaux pour les usages de la communauté universitaire, la préservation à long terme des ressources électroniques n'est pertinente que dans la mesure où elle offre également une **disponibilité et des droits d'usage suffisants** :

- en cas d'événement déclencheur, et plus ou moins sans doute selon le type d'événement, les contenus doivent rester accessibles ou l'être à nouveau le plus rapidement possible ;
- en termes de **fonctionnalités et droits d'usage** : consultation immédiate, impressions et téléchargement, réserves de cours... et fonctionnalités associées (hyperliens, contenus dynamiques...). Si ces fonctionnalités ne sont peut-être pas toutes compatibles avec la conservation pérenne (certaines solutions privilégient la préservation du contenu « source » – *source file*, distingué du contenu mis en forme et dynamique – *presentation file*), elles constituent la référence des attentes des usagers.

peu différente : The terms 'perpetual access', 'archiving', and 'long-term preservation' are sometimes used interchangeably. **Perpetual access** is most commonly associated with e-journal licence clauses designed to provide assurance of continued access to subscribed material in certain circumstances, including post-cancellation. **Archiving** describes the process and procedures whereby e-journal content may be managed for the short or long term. **Long-term preservation** refers to the processes and procedures required to ensure content remains accessible well into the future, regardless of any technical or organisational changes."

¹⁰ Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in an OAIS. This period extends into the indefinite future.

La concrétisation de l'accès pérenne aux ressources électroniques souscrites (ou achetées) pour les bibliothèques doit donc passer à la fois par 1) des garanties d'accès continu, en particulier post-abonnement – accordées par les éditeurs (via leur plateforme, via une plateforme tierce ou via la fourniture de fichiers aux bibliothèques) – ceci concerne les titres souscrits auprès d'éditeurs ou d'agrégateurs de contenus, et 2) par des garanties d'archivage pérenne des contenus (nécessairement via un système tiers ou par délégation de cette responsabilité aux bibliothèques) – ceci concerne à la fois les titres souscrits et les titres publiés en Open Access (Gold).

L'ensemble de la problématique de l'accès pérenne peut être schématisé comme suit :

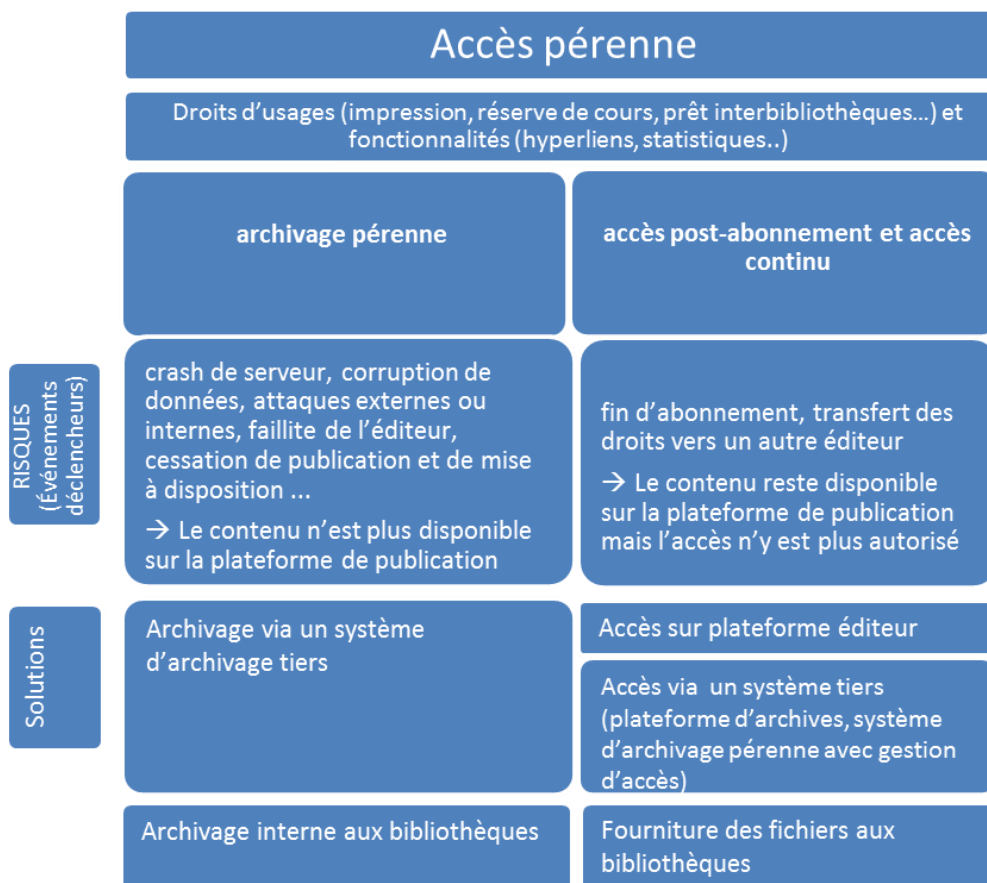


Figure 3 - schématisation des différentes composantes de l'accès pérenne des contenus souscrits ou achetés

Dans le cas où la fonction d'archivage pérenne est assumée par un tiers, on parle de **dark archive** ou de **light archive** selon que l'archive est destinée uniquement à l'archivage pérenne et soumise à des règles d'accès strictes ou qu'elle combine les fonctions de préservation à long terme et de communication et propose un accès plus ouvert. L'archivage via certains systèmes tiers peut donc couvrir à la fois les nécessités de l'archivage pérenne et de l'accès continu.

Le point 3.3 du présent rapport décrit et compare les solutions tierces d'archivage pérenne : CLOCKSS et e-Depot sont des *dark archives*, distribuée (réseau de pairs) dans le cas de CLOCKSS, centralisée (sites miroirs) dans le cas d'e-Depot, tandis que LOCKSS et Portico sont plutôt des *light archives*, distribuée dans le cas de LOCKSS, centralisée dans le cas de Portico, car elles autorisent l'accès post-abonnement.

1.4 Collaboration et mutualisation des ressources

Pour assurer la conservation pérenne des ressources électroniques, la nécessité de collaboration, entre les éditeurs et les bibliothèques, mais aussi entre les bibliothèques elles-mêmes et avec d'autres organisations, a été souvent soulignée¹¹, pour deux raisons principales :

- les coûts de mise en place d'une solution complète d'archivage pérenne dépassent à la fois les moyens des éditeurs et ceux des bibliothèques (à l'exception d'acteurs majeurs du paysage documentaire, telles les bibliothèques nationales et les très grands éditeurs) – le risque majeur de l'archivage pérenne est un risque économique¹² ;
- les nécessités de l'archivage pérenne supposent en elles-mêmes une distribution des risques (techniques, économiques et sociétaux) pour éviter les *single points of failure*.

L'archivage des e-journaux présente en outre deux particularités qui les distinguent d'autres contenus numériques à préserver (les contenus propriétaires notamment) et encouragent également la mutualisation des ressources :

- les contenus à préserver sont les mêmes pour de nombreuses bibliothèques dans le monde (comme dans le cas des périodiques imprimés) – la mutualisation du stockage prend donc tout son sens ;
- toutes les bibliothèques doivent négocier les droits d'accès et d'archivage avec les mêmes éditeurs – les négociations peuvent donc être effectuées au bénéfice de tous.

Ces facteurs expliquent que les systèmes d'accès pérenne dédiés aux e-journaux sont soit portés par des bibliothèques nationales (e-Depot, voir point 3.3.1.4) ou des projets nationaux (PECAN, licencesnationales.fr, voir point 5.1), soit des solutions reposant sur une collaboration entre un grand nombre de bibliothèques et d'éditeurs (LOCKSS, CLOCKSS et Portico, voir point 3.3).

2 Cadre légal et clauses contractuelles

Nous avons évoqué l'importance pour les bibliothèques d'obtenir des droits suffisants pour l'accès pérenne aux contenus acquis dans les modèles d'abonnement ou d'achat des ressources électroniques qui prévalent actuellement. Les données physiques restent la propriété des éditeurs, et accessibles sur leurs serveurs, l'accès à celles-ci, avant et après la fin de l'abonnement, n'est en effet possible que si l'éditeur l'autorise. Nous examinons ci-dessous quels droits sont accordés par le droit d'auteur et les clauses contractuelles signées dans le cadre de l'acquisition des ressources électroniques par la BICfB, et identifions des points d'actions pour le consortium afin de combler la distance entre les contrats existants et les droits nécessaires pour l'accès pérenne aux contenus.

¹¹ « Alle Bibliotheken sind sich darüber einig, dass unter dem wachsenden Druck betriebswirtschaftlichen Denkens keine Institution allein alle digitalen Ressourcen dauerhaft archivieren kann, sondern dass geeignete nationale Kooperations- und Austauschmodelle greifen müssen. » nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. (2010). Consulté 24 mai 2012, à l'adresse http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf, p. 31.

¹² C'est aussi l'une des conclusions de la 1^{re} phase du projet LIFE : « Collaborative tool development will significantly reduce the cost of ingest and metadata creation ». McLeod, R.; Wheatley, P.; Ayris, P. (2006). LIFE 1 : Final Report. Consulté à l'adresse <http://discovery.ucl.ac.uk/1854/1/LifeProjMaster.pdf>, p. 5. Les coûts d'ingestion et de création des métadonnées pouvant atteindre selon ce rapport 60% du coût du cycle de vie du document. Les conférences et publications d'Educopia soulignent également ce constat.

2.1 Cadre légal¹³

Le cadre légal, belge, européen et international, ne constitue pas à l'heure actuelle une réponse définissant de manière suffisante les droits et responsabilités des différentes parties impliquées dans l'archivage pérenne des ressources électroniques.

Les exceptions au droit d'auteur existant en droit belge¹⁴ comme dans d'autres pays européens (notamment à la suite de la directive InfoSoc¹⁵) qui autorisent la reproduction numérique à des fins d'archivage visent spécifiquement la reproduction (numérique) de documents imprimés menacés d'obsolescence. Elles imposent des limites en nombre de copies, restreignent la mise à disposition à des terminaux locaux et ne s'appliquent que dans les limites de l'absence de préjudice à l'exploitation normale de l'œuvre. En outre, le droit des bases de données limite de facto la possibilité d'effectuer des copies de contenus souscrits et accessibles via des plateformes électroniques.

Par ailleurs, en droit belge, les conventions prévalent dans le cas où les œuvres sont mises à disposition en ligne; les exceptions peuvent donc effectivement être neutralisées par les termes du contrat de licence signé avec l'éditeur lorsque les œuvres sont accessibles par services interactifs à la demande, comme c'est le cas, notamment, pour les bases de données et les revues électroniques. Le système mis en place laisse donc une très large place à la négociation contractuelle et n'établit pas un régime d'exceptions qui serait pleinement opposable.

Finalement, la majorité des ressources électroniques étant souscrites auprès d'éditeurs ou fournisseurs étrangers ou d'entreprises multinationales, l'utilisation d'exceptions au droit d'auteur supposerait une harmonisation du droit au niveau international, la définition de la juridiction étant également un problème complexe.

2.2 Les clauses contractuelles des abonnements

Il est important de souligner en premier lieu que dans la majorité des cas, à l'exclusion des contrats importants négociés directement avec l'éditeur, aucune licence spécifique n'est négociée, voire signée, avec l'éditeur ou l'agrégateur du contenu souscrit et que ce sont dès lors les conditions de l'éditeur par défaut qui s'appliquent. Par ailleurs, l'agence d'abonnement n'intervient normalement pas dans la négociation des termes des licences avec les éditeurs¹⁶.

35 contrats d'abonnements aux ressources électroniques parmi les plus importants (en nombre de titres et en coût) des institutions de la BICfB ont été examinés :

- contrats négociés par la BICfB :
 - 11 (bouquets de) périodiques électroniques, pour lesquels les institutions disposent dans la majorité des cas d'un contrat signé, souvent négocié¹⁷ : *Science* (American Association for the Advancement of Science), *American Chemical Society*, *American Physical Society*, *PsycArticles* (American Psychological Association), *Elsevier Cell Press*,

¹³ L'auteur remercie Laurence Thys, juriste auprès de la BICfB, de lui avoir fourni les informations nécessaires et l'assistance utile pour la rédaction de ces paragraphes.

¹⁴ Loi sur le droit d'auteur et les droits voisins du 30 juin 1994, article 22, §1er, 8° et 9°.

¹⁵ Directive 2001/29/DC du 22 mai 2001.

¹⁶ Wang, J., & Schroeder Jr., A. T. (2005). *The Subscription Agent as E-journal Intermediary*. *Serials Review*, 31(1), 20-27. doi:10.1016/j.serrev.2004.11.007, recommandait aux agents de s'investir dans cette problématique.

¹⁷ A défaut, par exemple pour *Science*, ce sont les clauses contractuelles publiées sur le site de l'éditeur qui ont été examinées.

Elsevier Science Direct, Nature, PNAS, Sage Premier ou HSS, Springer Online, Wiley Online Library ;

- 1 contrat d'archives de périodiques négocié par la BICfB : *Springer Online Journal Archives ;*
- 3 contrats d'agrégateurs de contenus, pour lesquels la BICfB ne disposait auparavant pas toujours de contrats signés, mais qui ont fait l'objet d'un marché public en 2013, avec détail des clauses contractuelles : *CAIRN, Ebsco (Academic Source Premier, Business Source Elite/Premier, EconLit full-text, LISTA full-text), ProQuest¹⁸ ;*
- contrats des produits les plus importants des portefeuilles des institutions non souscrits en consortium BICfB, pour lesquels la licence examinée a le plus souvent été la licence par défaut des éditeurs, consultée soit sur leur site web soit via les clauses mises à disposition par l'agent (ici, Ebsco), soit les contrats disponibles à l'ULg¹⁹ :
 - 15 bouquets de périodiques électroniques : *AIP Publishing, American Mathematical Society Journals, Annual Reviews, ASCE, American Society of Microbiology Journals, BMJ Journals (via Highwire), Walter de Gruyter, Emerald, Informa Healthcare, Institute of Physics Journals, Mary Ann Liebert Journals, Oxford University Press Journals, Royal Society of Chemistry Journals, SIAM, Taylor & Francis Journals) ;*
 - 5 contrats d'agrégateurs de contenu : *Highwire, JSTOR Collections, Persée, Project Muse, Wolters Kluwer Health.*

Afin d'obtenir également des données sur les termes des autres souscriptions, souvent de plus petits éditeurs, moins importants en termes financiers ou en termes de nombre de titres, les informations concernant les clauses d'accès post-abonnement et d'archivage pérenne ont été examinées, à titre indicatif, dans la base de données de l'agent Ebsco pour les titres du portefeuille ULg commandés via ce dernier. Ebsco utilise pour la description des clauses d'accès pérenne dans la base d'administration Ebsconet le vocabulaire mis au point par la Digital Library Federation Initiative (DLF) en 2004²⁰.

L'analyse de ces différents contrats et clauses avait pour but de répondre aux questions suivantes :

- concernant l'accès post-abonnement et l'accès continu :
 - des clauses d'accès post-abonnement sont-elles présentes ?
 - si oui, sous quelles modalités l'accès post-abonnement est-il proposé ? via la plateforme éditeur ? via transfert de fichiers ? via un tiers ?
 - l'accès post-abonnement concerne-t-il l'ensemble des titres souscrits, les différents types de contenus ?
 - les fonctionnalités des contenus et les usages autorisés diffèrent-ils de l'accès pendant l'abonnement ?
 - quels sont les coûts à prévoir pour l'accès post-abonnement ?
 - y a-t-il des modalités prévues et quelles sont-elles en cas de transfert de titre(s) et/ou de cessation de publication ?
- concernant l'archivage pérenne (interne ou externe) :

¹⁸ On utilise ici la notion d'agrégateur de contenu pour désigner un intermédiaire fournissant en majorité des contenus de plusieurs éditeurs indépendants sur une même plateforme. Ebsco et Proquest sont ici des agrégateurs de bases de données bibliographiques avec texte intégral tandis que CAIRN est un agrégateur de périodiques électroniques.

¹⁹ L'analyse a été faite sur les contrats et sur le cadastre ULg par commodité, ces données étant les plus accessibles et étant relativement similaires a priori à celles des autres institutions, en tout cas les autres universités complètes. L'analyse pourrait être effectuée sur les autres cadastres et contrats si souhaité.

²⁰ <http://old.diglib.org/pubs/dlf102/dlf102.htm>, appendice D, « Data element dictionary » proposait déjà 3 champs en 2004 : perpetual access holdings, perpetual access note, perpetual access right (utilisés par Ebscohost).

- les éditeurs mentionnent-ils des garanties d'archivage pérenne ? si oui, lesquelles ?
- les éditeurs autorisent-ils l'archivage pérenne / et, le cas échéant, l'accès continu, via des solutions tierces ?

L'analyse détaillée des clauses contractuelles est disponible dans l'Annexe 1.

2.2.1 Droits d'accès post-abonnement et d'accès continu

En ce qui concerne le contenu de ces clauses d'accès post-abonnement, les résultats de notre analyse sont assez proches de ceux décrits dans la littérature²¹, mais mettent en évidence l'importance de l'analyse de la couverture réelle des clauses post-abonnement, qui n'est pas toujours assez soulignée.

2.2.1.1 Présence de clauses d'accès post-abonnement dans les contrats éditeurs

L'analyse des clauses post-abonnement des contrats BICfB et ULg et des clauses par défaut des éditeurs rejoint les conclusions de l'étude publiée en 2010 par le JISC et le Publishing Research Consortium, qui signalait que, **si la majorité des grands et moyens éditeurs incluaient un droit d'accès contractuel post-abonnement, seuls 55% des petits éditeurs faisaient de même**²².

2.2.1.1.1 Cadastre Ebsco

Les solutions d'accès post-abonnement et continu (via la plateforme éditeur, une plateforme ou une archive tierce, ou la fourniture de fichiers aux bibliothèques) sont couvertes dans le vocabulaire de la DLF utilisé par Ebsconet par la clause de **droit d'accès perpétuel** – définie par 3 champs : *perpetual access holdings*²³, *perpetual access note*²⁴, *perpetual access right*²⁵.

Toutefois, la clause de **droit d'archivage**, définie par 3 champs : *archiving right*²⁶, *archiving format*²⁷, *archiving note*²⁸, qui concerne les modalités d'archivage pérenne (sur le réseau local des bibliothèques, ou via une archive tierce, en particulier LOCKSS, CLOCKSS ou Portico), doit être aussi considérée ici car dans certains cas, elle est la seule présente et l'accord de l'éditeur pour l'archivage via LOCKSS, Portico (avec autorisation d'accès post-abonnement) ou le réseau local des bibliothèques peut constituer une modalité d'accès perpétuel.

²¹ En particulier Beh, Eugenia, et Smith, Jane. Preserving the Scholarly Collection: An Examination of the Perpetual Access Clauses in the Texas A&M University Libraries' Major E-Journal Licenses. *Serials Review* 38, no 4 (décembre 2012): 235-242. doi:10.1016/j.serrev.2012.10.005.

²² E-only scholarly journals: overcoming the barriers | Research Information Network. (2010). Consulté à l'adresse <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/e-only-scholarly-journals-overcoming-barriers>, p.25.

²³ "The dates of coverage for which perpetual rights are available and agreed upon in the legal contract".

²⁴ "Additional information related to perpetual access"

²⁵ "The right to permanently access the licensed materials paid for during the period of the license agreement"

²⁶ "The right to permanently retain an electronic copy of the licensed materials"

²⁷ "The format of the archival content"

²⁸ "Additional information related to archiving rights, product and format"

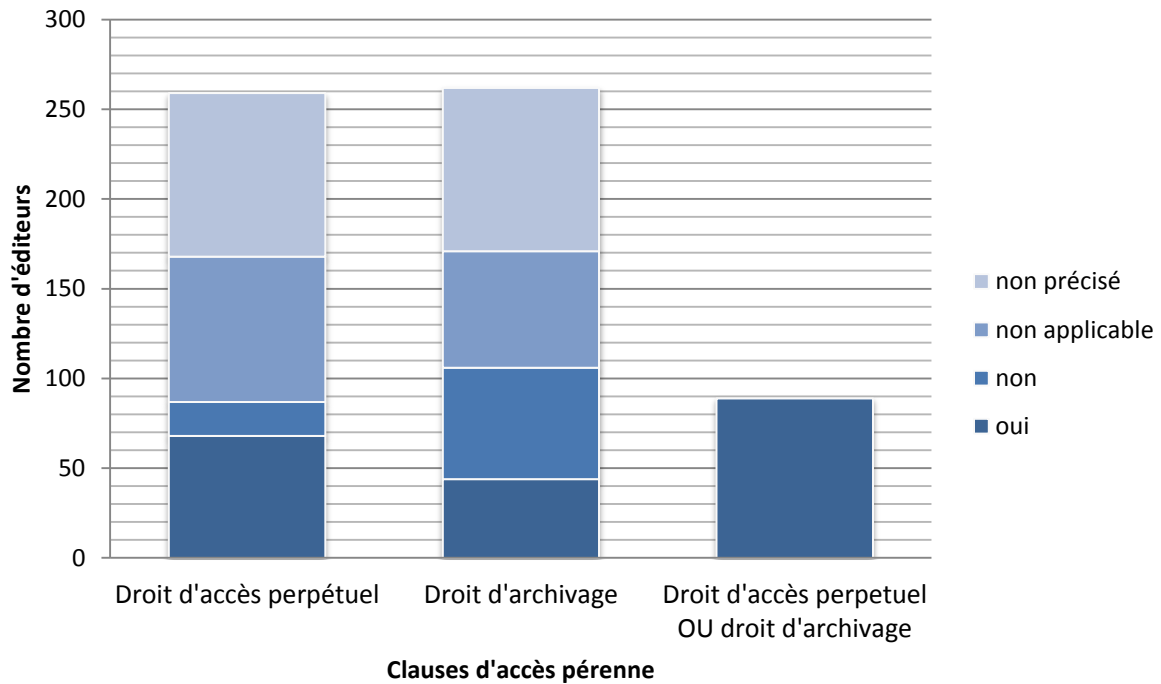


Figure 4 - L'accès continu dans les contrats éditeurs d'après le cadastre Ebsco 2013 de l'ULg disponible sur Ebsconet.

Une analyse rapide effectuée sur le cadastre Ebsco 2013 de l'ULg pour les achats de périodiques électroniques via l'agent montre que la présence de clauses d'accès continu ou post-abonnement n'est pas une évidence pour la majorité des éditeurs.

Sur les 259 éditeurs constituant le cadastre ULg, la base de données Ebsconet mentionne que 68 éditeurs autorisent l'accès post-abonnement, 19 éditeurs ne permettent pas l'accès post-abonnement dans le cadre du contrat courant (même s'il est possible de l'acquérir) et pour 81 éditeurs, l'accès post-abonnement n'est pas applicable²⁹. L'information est absente pour les titres de 91 éditeurs. En ce qui concerne le droit d'archivage, 43 éditeurs l'autorisent, 62 ne l'autorisent pas, l'information est non applicable dans le cas de 65 éditeurs et non précisée pour 91 éditeurs. En définitive, seuls 89 éditeurs, soit 34% de l'ensemble examiné, autorisent explicitement l'une ou l'autre modalité d'accès continu.

Ce chiffre est moins élevé que celui que relevait le JISC, mais une part probablement importante de la différence s'explique sans doute par le nombre important d'éditeurs pour lesquels rien n'est précisé sur Ebsconet sur le plan des droits d'accès continu accordés. En outre, un certain nombre de contrats majeurs analysés ci-dessous sont négociés directement avec l'éditeur et ne sont donc pas enregistrés dans la base Ebsconet.

²⁹ Communication d'Ebsco, 2 août 2013 : « L'accès pérenne n'est pas possible dans tous les abonnements "online" ce qui est précisé via la notion de "not applicable", lorsque la notion "no" est indiquée (sic) c'est que vous ne l'avez pas dans votre abonnement online mais qu'il est possible de l'acquérir moyennant une modification d'abonnement et éventuellement un coût supplémentaire. »

2.2.1.1.2 Contrats des grands éditeurs et agrégateurs

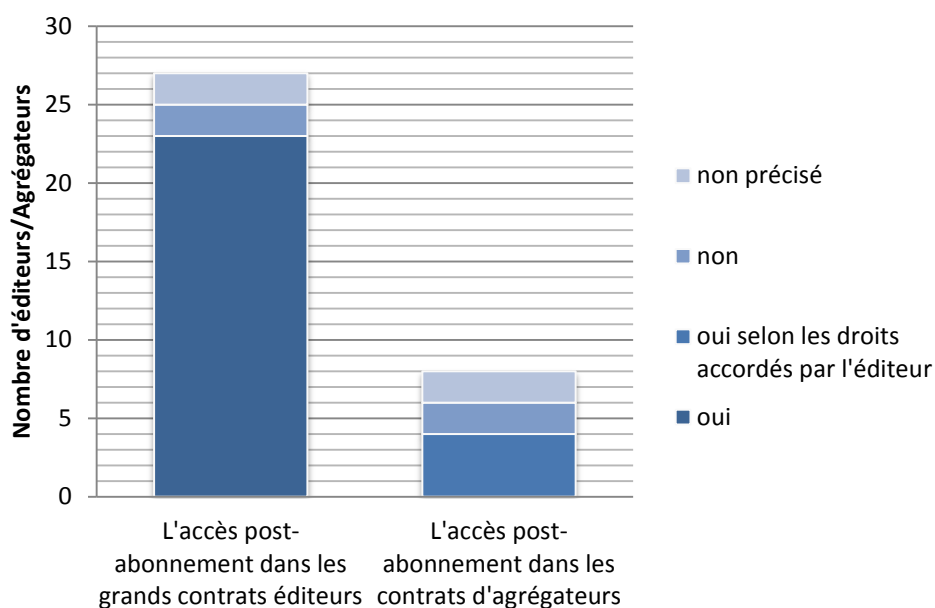


Figure 5 - L'accès post-abonnement dans les 35 grands contrats éditeurs et agrégateurs analysés (les plus importants en nombre de titres et coûts (BICfB))

Sur les 27 contrats des grands éditeurs de bouquets de périodiques (ou titre unique dans le cas de Science et PNAS) analysés (12 négociés en consortium BICfB, 15 non négociés en BICfB³⁰), tous contiennent des clauses d'accès post-abonnement à l'exception de la licence de l'American Mathematical Society et de celle d'Emerald³¹. La majorité (ACS, APA, APS, Annual Reviews, ASCE, ASM, BMJ, de Gruyter, Elsevier, Informa Healthcare, Mary Ann Liebert, Nature, OUP, Palgrave MacMillan, RSC, IOP, SAGE, Springer, Taylor & Francis, Wiley) autorisent l'accès post-abonnement sous l'une ou l'autre forme (accès via la plateforme de l'éditeur ou fourniture de fichiers). Seul AAAS (Science) interdit explicitement l'accès post-abonnement quelle qu'en soit la forme³².

Notons qu'en nombre de titres, les bouquets des grands éditeurs multidisciplinaires souscrits en consortium (Elsevier, Springer, Wiley, SAGE) représentent autour de 25-30% du portefeuille de périodiques électroniques (pour les grandes institutions) et jusqu'à 65% du portefeuille pour les plus petites institutions. En matière de prix, ces bouquets représentent une proportion encore plus élevée du portefeuille.

Les agrégateurs de contenu payant³³, n'étant pas propriétaires des contenus, ne disposent généralement pas des droits sur ceux-ci. Ils ne proposent donc souvent qu'une « clause de bonne volonté », qui accorde un droit d'accès post-abonnement en fonction des droits accordés ou non par

³⁰ Sur ces 27 contrats, 14 sont négociés directement avec l'éditeur, sans passage par l'agent. Parmi les autres, ce sont parfois les conditions standard de l'éditeur ou les informations publiées sur son site web, qui ont servi de base à l'analyse, à défaut d'un contrat en bonne et due forme.

³¹ Ces éditeurs participent pourtant à Portico (avec accès post-abonnement) et à LOCKSS.

³² L'accès post-abonnement n'est pas permis non plus dans le cas de la SIAM d'après les informations disponibles sur Ebsconet, mais ces informations ne semblent pas correctes puisque l'éditeur participe à Portico (avec accès post-abonnement) et à LOCKSS.

³³ La distinction entre éditeur et agrégateur n'est en réalité pas aussi nette : plusieurs grandes sociétés éditrices ont des contrats spécifiques avec des éditeurs associés, qu'ils publient sur leur plateforme (ex. Elsevier, Wolters Kluwer...); nous établissons ici la distinction à des fins d'analyse.

l'éditeur. C'est le cas pour Wolters Kluwer Health, ProQuest, Highwire et CAIRN³⁴. Ebsco ne propose aucune clause d'accès post-abonnement³⁵, et JSTOR exclut explicitement l'accès post-abonnement.

L'appel d'offres organisé en 2013 par la BICfB pour les bases de données à renouveler en 2014 a confirmé cette position des fournisseurs de contenu Ebsco et ProQuest. Il s'agit là d'un élément, avec la généralisation des outils Discovery, et la technologie d'OpenLinking, qui accentue la différence entre l'achat d'un accès au contenu full-text auprès de l'éditeur et l'achat d'un accès au contenu dans le cadre de données agrégées, ces dernières constituant par ailleurs un outil de recherche de qualité.

Pour SAGE, Wiley, la clause d'accès post-abonnement du contrat BICfB est plus permissive que les conditions mentionnées dans le cadastre Ebsco. Pour les autres, là où la comparaison est possible, les clauses correspondent.

Les 27 contrats des grands éditeurs (23 contrats) et agrégateurs (4 contrats) autorisant l'accès post-abonnement ont été analysés en détails dans les deux points suivants (modalités et coûts, et couverture de l'accès post-abonnement), afin d'illustrer les problématiques, qui sont les mêmes dans le cas des autres contrats d'après les clauses présentes dans Ebsconet.

2.2.1.2 Modalités et coûts d'accès post-abonnement

Pour les éditeurs qui l'autorisent, **les modalités et les coûts de l'accès post-abonnement**, lorsqu'elles sont précisées, peuvent être assez différents en fonction du contrat.

Accès post-abonnement via la plateforme de l'éditeur	
gratuit sans conditions	Annual Reviews, Oxford University Press, SAGE, Taylor et Francis ³⁶
gratuit s'il existe une souscription courante à au moins un titre du bouquet, payant sinon (sous la forme de frais de maintenance annuelle)	Elsevier, Springer, Wiley, AIP, Mary Ann Liebert, AIP
possible uniquement moyennant des frais de maintenance annuelle	ACS, APA, NPG
Accès post-abonnement via une plateforme d'archives (avec ou sans <i>moving wall</i>)	
gratuit	PNAS (via Pubmed Central)
moyennant souscription à la plateforme d'archives	PNAS (via JSTOR) ; tous les éditeurs autorisant l'accès post-abonnement via LOCKSS et Portico)
Fourniture des fichiers, sur demande, à la fin de la période de souscription (généralement sur CD-Roms ou DVD-Roms)	
gratuit	RSC, APA
payant	APS, ASCE
coût non précisé	Elsevier, Nature, Wiley, ACS, de Gruyter, NPG

Tableau 1 - Synthèse des modalités post-abonnement dans les grands contrats des institutions BICfB

Malheureusement, lorsque des frais d'accès post-abonnement via la plateforme de l'éditeur sont demandés, ils sont très rarement définis explicitement, ou, ce qui revient au même, ils sont définis comme « raisonnables » et/ou basés sur le recouvrement des coûts du fournisseur, ou à convenir, ou

³⁴ Le contrat CAIRN 2011-2013 n'en contenait aucune, mais le fournisseur a accepté d'inclure un accord de principe pour le contrat 2014-2016.

³⁵ Sur base des informations disponibles en ligne, aucun contrat n'ayant été signé avec l'ULg.

³⁶ Aucun coût n'est précisé.

bien incalculables précisément sur base des données actuelles voire totalement prohibitifs³⁷. Le problème se pose aussi lorsque la fourniture de fichiers proposée est payante, le coût étant souvent non précisé ou aussi peu clair que dans le cas de l'accès sur la plateforme de l'éditeur³⁸.

L'accès post-abonnement est par ailleurs généralement soumis aux restrictions suivantes :

- l'accès post-abonnement n'est pas permis en cas de rupture du contrat (soit de manière générale, soit limitée à la rupture de contrat par l'éditeur) ;
- l'accès post-abonnement est conditionné, comme chez les agrégateurs, au fait que l'éditeur dispose des droits pour les titres concernés (Elsevier, Nature, Wiley, ACS), ou à ses possibilités techniques (ACS, Annual Reviews) ; dans le cas contraire, l'éditeur propose une copie via support physique ; la fourniture des données sur support physique est parfois la seule possibilité (APS, ASCE). Dans ces deux cas, le format des données, les conditions exactes de fourniture (support physique, mode de fourniture...) et les coûts ne sont généralement pas précisés ou difficilement calculables (formulés dans les mêmes termes que pour la fourniture sur la plateforme) ;
- le format est parfois restreint : perte des fonctionnalités d'hyperliens ou autre, accès uniquement aux pdf (ACS).

En général, les clauses de post-abonnement précisent explicitement que les droits et obligations des parties survivent à la fin du contrat et les droits d'usage des contenus après la fin de l'abonnement sont les mêmes que les droits d'usage spécifiés pour le contrat dans son ensemble.

Certains fournisseurs donnent un accord de principe, mais sans fournir aucune information sur les modalités (ASM) et coûts (Annual Reviews). On suppose que l'accès est fourni gratuitement, sur la plateforme éditeur, mais sans garanties.

Les fournisseurs ne précisent généralement pas dans la licence l'existence d'accords avec des solutions tierces d'archivage telles que Portico ou LOCKSS pour l'accès post-abonnement (à l'exception de SAGE et Springer, suite à la demande de la BICfB, et de l'ASM), ce qui montre, au-delà du fait que les licences sont parfois anciennes, qu'un travail de négociation est encore à accomplir sur ce point puisque, dans le cas de Portico en particulier, mais également dans le cas de LOCKSS, pour garantir la pérennité du « bon vouloir technologique », une clause contractuelle entre l'éditeur et la bibliothèque est fortement recommandée pour garantir l'accès post-abonnement (cf. infra).

2.2.1.3 Couverture

La présence d'une clause post-abonnement dans les licences ne garantit pas que tous les titres du contrat soient couverts par cette clause. Dans le cas de certains *big deals* (Elsevier ScienceDirect et Wiley Online Library), seuls les titres de la « **core collection** » (contenus effectivement souscrits historiquement, appelés aussi *holdings*) sont couverts par la clause post-abonnement, à l'exclusion des autres titres du bouquet (SAGE, Springer accordent par contre l'accès post-abonnement pour tous les titres du bouquet).

Dans la plupart des cas également (AIP, APS, APA, Annual Reviews, ASCE, BMJ, de Gruyter, Elsevier, Informa Healthcare, Mary Ann Liebert, Project Muse, OUP, Palgrave, RSC, SAGE, Springer, Taylor &

³⁷ A titre d'exemple : Elsevier demande dans ce cas 0,63 € par demande de téléchargement dans l'année écoulée, indexés annuellement, ce qui donnerait un montant pour l'ULg pour 2013, pour les 327 titres appartenant à la liste des holdings et ayant un ISSN, (soit les seuls concernés par l'accès post-abonnement, voir plus bas), de 153.713 demandes de téléchargement, soit 9.683,92 € de fees. Cette option est donc clairement un incitant à maintenir au moins une souscription sur la plateforme de l'éditeur.

³⁸ APS propose un tarif consortial pour l'achat des CD-Roms. Le prix proposé en 2013 pour l'achat des 30 CD-Roms contenant les publications de l'année 2011 des 7 titres APS était de 1.535\$ (offre consortiale BICfB).

Francis, Wolters Kluwer Health), l'accès post-abonnement couvre **uniquement les années effectivement souscrites**, à l'exclusion des années antérieures généralement disponibles jusque 1997 sur la plateforme dans le cadre d'une souscription courante (Wiley, qui inclut les années antérieures à l'abonnement, est une exception remarquable).

Par conséquent, même lorsqu'existent des garanties d'accès pérenne, le portefeuille concerné est notablement plus réduit que l'accès courant. Si l'on veut garantir une couverture pérenne plus proche de l'accès des contrats courants, il faut envisager à la fois l'extension du portefeuille de *holdings* (et donc un surcoût majeur) et l'achat d'archives, tout en notant que 1) les années situées entre le début de l'électronique (généralement 1997) et le début de la souscription électronique par l'institution ne sont pas toujours proposées à l'achat et 2) l'achat d'archives électroniques ne garantit aucunement, en soi, un accès pérenne aux collections (cf. infra 2.2.3).

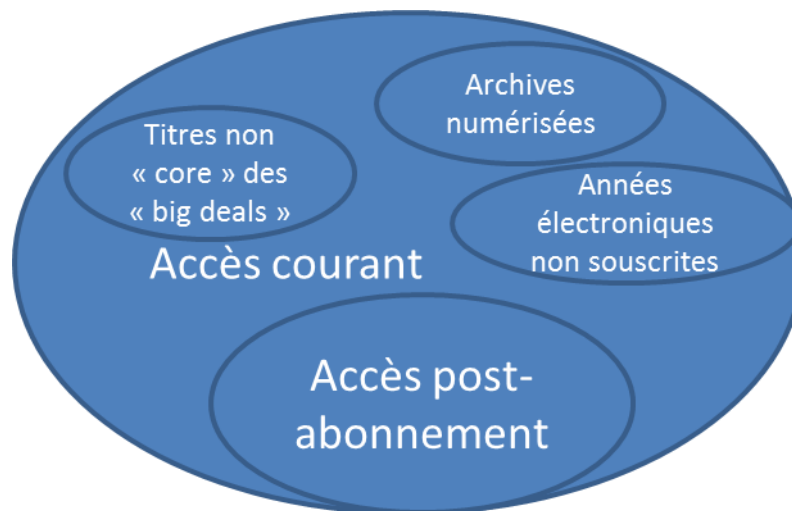


Figure 6 - Les limitations de l'accès post-abonnement contractuel par rapport à l'accès courant

Cette restriction des clauses d'accès et d'archivage pérenne au portefeuille effectivement souscrit en électronique réduit aussi considérablement l'utilisation de l'archivage électronique comme alternative à l'archivage papier et aux projets de conservation partagée des périodiques papier. En effet, en l'état actuel des choses, sans achat d'archives ou extension des droits des licences, une institution ne pourrait se défaire des collections papier que pour les titres et années souscrites en P+E (pour lesquelles les garanties d'accès et d'archivage pérenne sont suffisantes).

Cette différence entre accès courant et archivage autorisé ne fait dans le cas des grands bouquets de périodiques, que refléter la sélection effectuée parmi les contenus. Cette sélection ne se fait pas a posteriori lors de l'archivage des contenus, mais a priori, lors de la sélection des *holdings*³⁹.

A l'inverse, pour les éditeurs ou les plateformes qui proposent un accès « ouvert » avec **barrière mobile** (*moving wall*) (PNAS, CAIRN, ASM), la question de l'accès post-abonnement se pose avec une acuité moindre puisqu'elle ne concerne jamais que les toutes dernières années ou les tous derniers mois souscrits. La question de l'archivage pérenne des contenus continue toutefois, même dans ces cas, à garder toute son importance.

³⁹ Des modèles alternatifs se mettent en place pour viser plutôt une sélection a posteriori pour l'acquisition pérenne, en particulier dans les modèles d'achat d'e-books ; citons par exemple le modèle « EBS – Evidence-Based model » d'Elsevier, ou les modèles « PDA – Patron Driven Acquisition » où la sélection des *holdings* pour l'achat pérenne se fait au terme de l'année, en fonction par exemple des statistiques d'usage. La fonction du *swapping* des titres dans les contrats de périodiques basés sur les abonnements historiques est également une façon de se rapprocher de ce mécanisme de sélection a posteriori.

2.2.1.4 Clauses de transfert et *Transfer Code of Practice*

Le transfert de titres d'un éditeur à un autre est l'un des risques de perte d'accès (« événement déclencheur »), pendant ou après la durée de l'abonnement. En effet, dans ce cas, les clauses des contrats en cours avec les abonnés ne sont pas garanties d'office légalement par le nouvel éditeur, mais dépendent des modalités de transfert ou de rachat négociées entre les éditeurs. C'est la raison qui a motivé l'initiative du TRANSFER Code of Practice⁴⁰, qui propose des recommandations de garanties, de format d'échange de données et de procédure lors des cas de transfert. Ce code a été signé par la majorité des grands éditeurs internationaux et par ALPSP⁴¹. Même si le Transfer Code of Practice n'est pas juridiquement contraignant, il reste un signe de bonne volonté et il est important d'inclure une clause à ce sujet dans la négociation des licences pour les éditeurs qui ne seraient pas signataires⁴². L'information sur les conditions de transfert des titres (contractuelles ou via signature du TRANSFER Code of Practice) sont reprises pour chacun des grands éditeurs examinés dans l'Annexe 1.

2.2.2 Garanties d'archivage pérenne

Outre les conditions d'accès post-abonnement, un certain nombre d'éditeurs garantissent explicitement dans les licences, dans une certaine mesure, l'archivage pérenne, généralement via une formule incluant la mention d'« efforts (commerciallement) raisonnables pour maintenir la préservation des contenus dans une archive » (AAAS, ACS, APS, JSTOR, OUP) soit sans précision (AAAS), soit via des sites miroirs, soit via des systèmes tiers d'archivage.

Dans le cadastre Ebsco (cf. supra l'analyse de la présence de la clause de droit d'archivage, fig. 4), 43 éditeurs autorisent l'archivage, 62 ne l'autorisent pas, l'information est non applicable dans le cas de 65 éditeurs et non précisée pour 91 d'entre eux.

Cependant, la notion même d'« efforts raisonnables » (voire commercialement raisonnables) est en contradiction avec les exigences de l'archivage pérenne (cf. infra 4.1), et ces clauses ne peuvent, en tant que telles, à l'exception de celles qui concernent les systèmes tiers, être considérées comme des garanties suffisantes.

Pour pallier à ce manque, la BICfB a, en 2012, décidé d'inclure dans son modèle de licence utilisé pour les négociations avec les éditeurs, une clause intitulée "Archival rights and long-term preservation" : *"The Publisher undertakes to provide, or to make arrangements for a third party to provide, an archive of the Licensed Materials for the purposes of long term preservation of the Licensed Materials, and to permit Authorized Users to access such archive after termination of this License or in the event the Publisher ceases to publish a part or parts of the Licensed Materials, as defined in Schedule 2; the Publisher will be able to provide the Licensee on request with a list of the Licensed Materials available in archives, complying with the KBART recommendation (published by the UK Serials Group (UKSG))."*⁴³ Cependant, aucun des éditeurs approchés à ce jour qui ne participait pas déjà à une solution d'archivage pérenne externe (APS, CAIRN⁴⁴) n'a accepté d'inclure cette clause

⁴⁰ <http://www.uksg.org/Transfer/Code>

⁴¹ http://www.uksg.org/Transfer/Transfer_Publishers

⁴² L'Annexe 1 reprend également l'information pour les éditeurs avec lesquels les institutions de la BICfB ont un contrat.

⁴³ La Norme NISO RP-7-2012, SERU: A Shared Electronic Resource Understanding. Consulté à l'adresse http://www.niso.org/publications/rp/RP-7-2012_SERU.pdf, qui propose un modèle de licence au bénéfice à la fois des éditeurs et des bibliothèques inclut une clause d'accès pérenne, mais dont les modalités ne sont pas, logiquement, explicitement définies, et sont à préciser au cas par cas.

⁴⁴ En décembre 2014, dans le cadre du renouvellement du contrat d'abonnement des institutions de la BICfB, CAIRN n'a pas fermé la porte à une participation éventuelle à la solution choisie par la BICfB pour la préservation des titres électroniques.

dans le contrat. En outre, en pratique, comme le constatait déjà l'étude du JISC (Morrow, Beagrie, & Jones, 2008), l'existence d'une clause d'archivage pérenne sans mention de solution préférée spécifique n'est en soi pas suffisante : toutes les solutions ne sont pas équivalentes en matière d'accès et d'archivage pérenne, et elles n'offrent généralement pas (à l'exception de CLOCKSS) des garanties d'accès sans conditions – les institutions de la BICfB ne bénéficient d'ailleurs pas, actuellement, d'un accès à ces archives via les solutions tierces LOCKSS ou Portico, auxquelles elles ne souscrivent pas.

2.2.3 Les bouquets d'archives

L'accès en ligne aux archives des éditeurs, c'est-à-dire aux anciens numéros des périodiques scientifiques (souvent les numéros antérieurs au milieu ou à la fin des années 1990) est proposé aux bibliothèques depuis plusieurs années selon plusieurs modalités :

- l'achat proprement dit des archives, la bibliothèque acquérant une garantie d'accès aux numéros concernés sur la plateforme de l'éditeur (Springer OJA, Science, APS PROLA...) après un paiement unique (*one-shot*) ;
- la souscription à des archives sur la plateforme éditeur pour une durée contractuelle ; elle est parfois gratuite pour la durée du contrat d'accès aux titres courants (APS PROLA, Annual Reviews Back Volumes) ;
- la souscription à des bouquets d'archives proposés via d'autres plateformes (JSTOR), la bibliothèque accédant aux archives via la plateforme pendant la durée de sa souscription au produit.

La désignation « archives » peut induire à penser, erronément, que l'accès pérenne est garanti. Or, ce n'est pas le cas : la souscription à des collections d'archives équivaut, dans les termes, à la souscription au contenu courant sur les plateformes éditeurs.

- dans le cas de plateforme d'archives payantes, l'accès post-abonnement n'est pas nécessairement garanti (cf. JSTOR), ni l'archivage à long terme (cf. CAIRN)⁴⁵. Le cas de JSTOR est particulier puisqu'en réalité, on peut le considérer comme un achat (via l'Archive Capital Fee), suivi du paiement d'une souscription annuelle pour l'accès « post-achat » ;
- dans le cas d'achat d'archives (Springer OJA), l'accès post-abonnement / post-achat est souvent garanti soit via la plateforme de l'éditeur, soit via transfert de fichiers, mais cet accès n'est pas nécessairement gratuit ;
- dans le cas de bouquets d'archives accessibles en complément dans le cadre d'une souscription aux titres courants (APS PROLA, Annual Reviews Back Volumes, Science Classic), l'accès post-abonnement aux archives n'est pas garanti. Ces archives sont disponibles à l'achat, avec des clauses d'accès à long terme.

2.3 Conclusions et recommandations

2.3.1 Négociations avec les éditeurs

Les licences actuelles souffrent généralement, quand les clauses d'accès pérenne existent (pour la majorité des grands éditeurs), d'un flou en termes de conditions et/ou de coûts difficile à concilier avec les exigences juridiques et la planification de la préservation à long terme. Il est donc important :

⁴⁵ JSTOR vise à appliquer le modèle de préservation pérenne de Portico à sa plateforme – ce qui ne présenterait pas encore toutes les garanties organisationnelles nécessaires.

- de prévoir, dès le début des contacts avec les éditeurs, de négocier des clauses d'accès pérenne précises en termes de couverture, de modalités d'accès, de formats, de coûts, de conditions d'usage, de garanties d'accès à long terme – ou de clarifier les conditions existantes – afin d'éviter tout problème en cas de contestation ;
- de préciser ou de clarifier, 1) en cas de possibilités multiples (accès sur la plateforme éditeur ou fourniture de fichiers, par exemple), qui décide de la modalité effective : l'éditeur ou l'institution abonnée et 2) les délais et modalités de demande et de fourniture ou d'accès.

Au sujet de la couverture, les années situées entre environ 1995-1997 (fin de la couverture des bouquets d'archives électroniques vendus par les éditeurs (Springer, Elsevier, APS, Wiley) et le début de la souscription au format électronique (qui peut être 2002 par exemple) posent problème : en effet, elles sont, dans la plupart des contrats des grands éditeurs, accessibles online dans le cadre des souscriptions courantes, mais ne sont pas couvertes par les clauses post-abonnement puisqu'elles n'ont été ni souscrites, ni achetées. Elles ne font pas l'objet par ailleurs d'offres spécifiques d'achat. Il est primordial, si les universités souhaitent s'orienter vers une politique de collections e-only continue dans le temps, de clarifier l'offre et le statut de ces titres.

Il est bien entendu toujours préférable d'avoir une licence signée, les conditions générales des éditeurs pouvant changer à tout moment unilatéralement. En outre, un nombre important de plus petits éditeurs et de bases de données full-text ne communiquent pas d'informations sur l'accès pérenne aux ressources, et sont plus difficiles à approcher au cas par cas. Il serait donc utile :

- que les institutions négocient une licence avec clauses d'accès pérenne au moins pour les titres qu'elles jugent les plus importants (en terme d'usage, de coût...) si aucune licence n'a été signée (Taylor et Francis, RSC, American Society of Civil Engineers, American Geophysical Union, American Society for Biochemistry and Molecular Biology, IOP, SIAM...)⁴⁶ et en particulier si aucune information n'est disponible au sujet de la politique pérenne de l'éditeur.
- que les institutions ou la BICfB, éventuellement en partenariat avec d'autres institutions, négocient, au moins, la définition de bonnes pratiques et recommandations pour les éditeurs auprès des associations d'éditeurs et des agrégateurs de contenus. L'approche d'ALPSP (The Association of Learned and Professional Society Publishers) par exemple pourrait être un début⁴⁷.

Toutes les solutions d'accès pérenne et d'archivage à long terme ne se valent pas, comme nous le verrons plus loin. La seule présence de clauses à ce sujet n'est donc pas suffisante, il s'agit d'établir une hiérarchie en fonction des préférences de la BICfB (et d'adapter le modèle de licence en conséquence) :

⁴⁶ Les éditeurs listés ici sont les éditeurs dont les titres sont les plus chers, le risque étant calculé sur base de l'investissement étant donné l'importance de l'accès post-abonnement. Le risque est évalué différemment lorsque l'on envisage la question non pas de l'accès post-abonnement mais de l'archivage à long terme des titres. De ce point de vue, les titres les plus à risque sont souvent les titres d'éditeurs plus petits, moyens, et/ou Open Access.

⁴⁷ L'association a lancé en 2008 un sondage auprès de ses membres pour savoir ce qu'ils pensaient de la question de l'archivage à long terme des publications électroniques. 90% ont estimé que c'était une question cruciale.

- négocier un accès post-abonnement sur la plateforme éditeur totalement gratuit ou gratuit en cas de maintien d'au moins une souscription, ou éventuellement pour un très faible coût ; ceci est particulièrement important si la BICfB s'oriente vers le choix de CLOCKSS et LOCKSS⁴⁸ ;
- négocier, en fonction du choix éventuel d'une ou plusieurs solutions d'archivage tierces par les universités de la BICfB, ou de scénarios de conservation particuliers (cf. infra 7.2), des clauses d'archivage dans la solution choisie ;
- négocier l'augmentation de la couverture d'accès pérenne : en termes de titres et d'années couvertes (couverture des titres achetés en papier et accessibles dans le contrat courant en version électronique, abandon des *big deals* et accès post-abonnement à l'ensemble des titres inclus dans les *big deals*...).

Afin de simplifier la gestion de l'historique des contrats, d'une part, et de clarifier, voire d'étendre les états de collections couverts par l'accès post-abonnement, une orientation idéale serait de **négocier avec les éditeurs une couverture et des conditions rétrospectives** dans le cadre des nouveaux contrats.

Plus largement se pose la question du poids des institutions ou de la BICfB dans les négociations avec les éditeurs. Si cette problématique peut être abordée en soi, par la mise en place de licences nationales ou l'intervention de groupements européens ou internationaux, elle est aussi l'une des forces des solutions d'archivage tierces des périodiques électroniques (Portico, LOCKSS et CLOCKSS), qui peuvent négocier l'archivage pérenne des contenus auprès des éditeurs au nom de tous leurs membres (cf. infra 3.3).

2.3.2 Gestion de l'information

Indépendamment du flou des clauses contractuelles, l'identification des garanties d'accès pérenne titre par titre n'est pas un processus aisé, car :

- l'information n'est actuellement pas centralisée, en tous cas à l'ULg qui ne dispose pas de système de gestion du portefeuille globalisé (ERM), mais répartie entre la base de données de l'agent – où l'information peut être, selon les cas, précisée par titre ou par bouquet – et les licences négociées avec les éditeurs accompagnées de leur(s) liste(s) de titres respective(s), dans le meilleur des cas en format excel compatible avec la norme KBART ;
- l'information n'est pas normalisée : malgré l'existence de dictionnaires de données (Digital Library Federation Initiative, 2004⁴⁹), et celle de formats de licence « machine readable » (ONIX⁵⁰), la terminologie employée par les éditeurs dans les licences n'est pas uniforme ni structurée⁵¹ ;
- l'information n'est pas toujours complète, en particulier sur les années d'accès, liées à l'historique des contrats – ou bien l'information n'est pas mise à jour⁵². Le problème est crucial pour les titres souscrits dans des contrats antérieurs mais plus dans le contrat actuel.

⁴⁸ Cf. chapitre 7. Important si le moissonnage des contenus n'a pas été effectué pendant la durée de l'abonnement ; important pour les éditeurs non couverts par ces solutions.

⁴⁹ <http://old.diglib.org/pubs/dlf102/dlf102.htm>, appendice D, « Data element dictionary » proposait déjà 3 champs en 2004: perpetual access holdings, perpetual access note, perpetual access right (utilisés par Ebscohost).

⁵⁰ <http://www.editeur.org/21/ONIX-PL/>. Le schéma de données d'ONIX-PL définit les éléments « ArchiveCopy », « ContinuingAccess » et « NotificationOfDarkArchive » ainsi que plusieurs relations sur « ContinuingAccess » permettant de préciser notamment la durée et la couverture.

⁵¹ Elle l'est peut-être dans les ERM au cas où l'éditeur y intègre une version spécifique du texte de sa licence, mais c'est à vérifier.

⁵² Waller, A., & Bird, G. (2006). « We Own It »: Dealing with "Perpetual Access in Big Deals. Serials Librarian, 50(1/2), 179-196, indique que, lors de la vérification manuelle des holdings effectuée pour 60 bibliothèques

Il est donc difficile et coûteux en ressources humaines, pour un titre donné, de savoir s'il est couvert par une garantie d'accès post-abonnement et de d'archivage à long terme. Or, cette information est nécessaire :

- pour la prise de décision à plusieurs étapes clés du cycle de gestion des collections : négociation et achat, examen de la possibilité de passage à l'e-only au lieu de print + électronique, arrêt ou renouvellement de la souscription ;
- pour entreprendre les actions assurant effectivement l'accès pérenne au titre en fonction des scénarios possibles : envoi d'une réclamation à Portico⁵³, ...

Le JISC Collections⁵⁴ faisait en 2010 des constatations similaires, ce qui l'a amené à lancer le projet Post-Cancellation Entitlement Registry Scoping Project (finalisé en mai 2012)⁵⁵, afin de récolter les données d'accès pérenne pour les institutions partenaires, d'une part, et d'établir les coûts et workflow associés, en particulier pour la mise à jour annuelle de ces données. Si les coûts et workflow précis n'ont pas été établis, le rapport a mis en évidence la lourdeur du problème et la pluralité des stratégies à mettre en place. Un projet parallèle lancé par EDINA, le centre national de données pour la communauté de l'éducation supérieure au Royaume-Uni, désigné par le JISC⁵⁶, fin 2011 (PECAN 2) vise quant à lui à mettre en place l'application (Entitlement Registry) destinée à stocker ces données⁵⁷ et à gérer les droits d'accès post-abonnement aux contenus (interface nationale établissant le lien entre les droits d'accès au contenu et les solutions d'accès à ce contenu), mais également une *dark archive*⁵⁸.

A l'échelle de la BICfB, un projet similaire à celui du JISC demanderait des moyens dédiés. Les solutions les plus rationnelles, étant donné la taille du consortium et la diversité des systèmes de gestion, nous semblent devoir être envisagées à l'échelle des institutions :

- voir avec les agences d'abonnement si, là où ce n'est pas le cas, les conditions des licences négociées peuvent être intégrées dans leur système (elles fournissent déjà par exemple, la gestion titre à titre des bouquets). Même dans le meilleur des cas cependant (clauses contractuelles titre à titre, historique des données), le problème reste que seuls les titres achetés via l'agent sont inclus dans le système, ce qui ne permet pas une gestion globale de la problématique ;
- si possible, mais cela a un coût, utiliser un système d'ERM et y intégrer les données concernant l'accès pérenne (par exemple dans le cadre de l'achat d'un nouveau système de gestion intégrée des collections).

Quelle que soit la solution choisie, la vérification des clauses contractuelles des contrats négociés en consortium peut continuer à être gérée de manière centralisée à l'échelle du consortium afin d'économiser les dépenses en ressources humaines. Toutefois, en l'absence d'une solution centralisée, leur intégration dans un ERM doit être effectuée au niveau institutionnel, ou par l'agent.

canadiennes pour 6 contrats d'abonnement, l'information à jour était, si pas une exception, loin d'être la norme.

⁵³ Pour activer le titre dans la LOCKSS Box, il n'y a pas a priori besoin de liste indépendante, les titres activables sont visibles dans la LOCKSS box directement (même principe que SFX d'ExLibris par exemple).

⁵⁴ <http://www.jisc-collections.ac.uk>. JISC Collections est l'organisme national de support de la fourniture de contenu numérique pour l'éducation et la recherche au Royaume-Uni.

⁵⁵ Voir <http://www.jisc-collections.ac.uk/Reports/entitlementregistryreport/> et <http://www.jisc-collections.ac.uk/Entitlement-Registry-Project/>

⁵⁶ <http://edina.ac.uk/>. EDINA est hébergé par l'Université d'Edimbourg en Ecosse.

⁵⁷ <http://edina.ac.uk/projects/pecan/> et http://edina.ac.uk/projects/pecan2_summary.html

⁵⁸ C'est en tous cas ce que précisent les objectifs initiaux du projet, mais aucune information sur cette *dark archive* n'a pu être obtenue avant la finalisation du présent rapport.

Dans tous les cas, il est crucial pour l'accès post-abonnement :

- de **maintenir l'historique et la mise à jour des données** ;
- d'intensifier la vérification et l'harmonisation des données (manuellement, ou avec l'aide d'outils informatiques) issues des contrats éditeurs et des contrats des institutions⁵⁹ ;
- d'harmoniser les formats de données en vérifiant et encourageant l'utilisation des standards (KBART, ONIX) par les systèmes de gestion des ressources électroniques, et d'enregistrer également les informations du TRANSFER code of Practice.

ONIX_PL (<http://www.editeur.org/21/ONIX-PL/>) (ONIX for Publications Licence)

Le vocabulaire xml ONIX-PL est destiné à enregistrer l'information de licence des publications électroniques (principalement les e-journaux, mais il est adapté aussi aux autres types de publications (e-books...), afin que cette information puisse être transmise et intégrée dans les différents systèmes de gestion des ressources (entre éditeur et bibliothèque, entre éditeur et agent, entre agent et bibliothèque). La version actuelle du vocabulaire est la version 1.0, publiée en novembre 2008 et restée stable depuis lors. Le schema de données d'ONIX-PL définit les éléments « ArchiveCopy », « ContinuingAccess » et « NotificationOfDarkArchive » ainsi que plusieurs relations sur « ContinuingAccess » permettant de préciser notamment la durée et la couverture. Afin de permettre la création, l'édition et la maintenance des expressions ONIX_PL, EDITEUR a développé un logiciel Open Source, OPLE (<http://www.editeur.org/22/OPLE-Software/>).

ONIX_PH (<http://www.editeur.org/127/ONIX-PH/>) (ONIX for Preservation Holdings)

Le vocabulaire xml ONIX-PH est une spécialisation du vocabulaire ONIX Serials, destinée à enregistrer l'information de préservation des holdings. La version actuelle du draft (version 0.21) a été publiée en mars 2012. A l'origine, ONIX-PH a été développé pour assurer la transmission de données entre les solutions tierces d'archivage et le Keeper's Registry (cf. infra), mais des extensions du vocabulaire peuvent être envisagées pour (1) couvrir la communication des *holdings* archivés entre l'éditeur et d'autres partenaires et (2) couvrir l'archivage papier des *holdings*. Le vocabulaire contient une section destinée à couvrir les termes de licence, et qui pourrait être étendue pour inclure des données structurées sur les informations de licence.

En l'absence d'une de ces solutions, la gestion des clauses d'accès pérenne, en particulier des clauses post-abonnement, ne pourra peut-être, pour des raisons de coûts, être réalisée dans un premier temps que pour les titres jugés prioritaires par l'institution.

3 Les solutions d'archivage à long terme des e-journaux⁶⁰

Les enjeux et problématiques de la conservation pérenne des e-journaux sont établis depuis le début ou le milieu des années 2000. La question a suscité un grand intérêt et plusieurs publications dans les années 2006-2010⁶¹, examinant en particulier les solutions tierces d'archivage existantes, mais peu

⁵⁹ La qualité des métadonnées contractuelles est en effet une condition sine qua non de l'accès pérenne aux contenus, tout comme la qualité des métadonnées techniques l'est pour la préservation.

⁶⁰ Nous remercions Fanny Corbel, stagiaire de l'ENSSIB à l'Université de Liège en janvier 2012, qui a réalisé la recherche bibliographique et l'étude préliminaire concernant ce chapitre, et en particulier les solutions LOCKSS, CLOCKSS et Portico.

⁶¹ Plusieurs études (dont un certain nombre par Charles Beagrie Ltd), en particulier celles du JISC (Morrow, Beagrie, & Jones, 2008) et du CLIR (Kenney, Entlich, & Hirtle, 2006), dans le monde anglo-saxon, mais aussi en France et en Allemagne (Beagrie, 2010), ont posé la problématique, défini des critères d'évaluation et esquissé des pistes de solution.

de nouveautés sont apparues depuis dans ce cadre général. Toute la question réside dans la décision pour les bibliothèques de s'impliquer ou non dans une ou plusieurs initiatives existantes, en fonction de l'évaluation des priorités, des garanties et services proposés et des simulations de coûts – questions centrales dans la réflexion actuelle⁶².

Les solutions d'archivage pérenne des e-journaux doivent naturellement présenter les fonctionnalités requises pour tout système d'archivage numérique, mais également avoir un mandat spécifiquement dédié et des fonctionnalités suffisantes pour répondre aux problématiques spécifiques d'accès pérenne définies en points 1.2 et 1.3. Ce critère permettra de distinguer les solutions spécifiquement dédiées (Portico, LOCKSS, CLOCKSS, e-Depot) d'autres solutions intégrant l'archivage pérenne de la documentation électronique, et souvent associées à la problématique, mais qui ne peuvent, en tant que telles, apporter une solution suffisante aux bibliothèques : les dépôts institutionnels, les plateformes d'archives, les logiciels de préservation pérenne.

Etant donné la complexité de la problématique et la diversité des éditeurs et contenus, aucune solution tierce existante à l'heure actuelle ne constitue une panacée. L'objectif est donc plutôt d'identifier les forces et les faiblesses de chacune des solutions et leur articulation entre elles et avec les points d'action possibles auprès des éditeurs, afin de proposer des scénarios destinés à permettre à la BICfB de se rapprocher de l'objectif que formulait en 2011 le groupe de travail Journal Archiving Implementation Group du JISC : *“Perhaps not to get everything archived but to make sure there is accommodation for everything”*⁶³.

Dans cette perspective, seules les solutions tierces dédiées spécifiquement aux e-journaux ont fait l'objet d'un examen approfondi, et d'une comparaison détaillée (chapitre 4). Les solutions d'archivage non dédiées sont abordées plus brièvement, pour poser des balises et identifier leur utilité spécifique éventuelle dans un scénario global. Finalement, nous présentons également brièvement pour compléter le panorama quelques initiatives supplémentaires mises en place dans les pays voisins liées à l'archivage pérenne de la documentation électronique (chapitre 5).

3.1 Critères minimaux et typologie

Les systèmes d'archivage numérique pérenne ont été définis fonctionnellement par l'Open Archival Information System (OAIS)⁶⁴, publié en 2002 par le Data Archiving and Ingest Working Group du Consultative Committee for Space Data Systems (CCSDS), sous la coupole de la NASA, qui offre un référentiel commun désormais généralement accepté pour décrire un système d'archivage numérique. Depuis, un travail important a été réalisé aux Etats-Unis et en Europe pour arriver à des processus de validation et de certification de ces systèmes, concrétisé en 2010 via le « Trusted Digital Repository.eu », en un système de certification à plusieurs niveaux, intégrant les différentes initiatives antérieures : le Data Seal of Approval (DSA)⁶⁵, un système de certification par auto-évaluation et peer-reviewing, sur base de 16 guidelines, mis au point par le Data Archiving and Networked Services (DANS)⁶⁶ néerlandais ; le RAC (ISO 16363)⁶⁷, basé sur le travail du Center for

⁶² Une analyse risque-bénéfice de l'investissement dans les solutions d'archivage des e-journaux en fonction du statut des titres a été proposée par le JISC [2008], sous la forme de 4 scénarios en fonction de la probabilité des événements déclencheurs et de l'impact de la perte d'accès au titre selon que le document fait partie ou non de la « core collection » de l'institution et qu'un accès courant est requis ou non.

⁶³ En particulier la « long tail » des e-journaux numériques de petits éditeurs non participant aux grandes solutions d'archivage et des contenus Open Access.

⁶⁴ <http://public.ccsds.org/publications/archive/650x0m2.pdf>. Dans l'OAIS, un système d'archivage est considéré comme l'articulation de personnes, de fonctions, de technologies et d'objets « qui a accepté la responsabilité de préserver l'information numérique à long terme et de la rendre accessible à une communauté définie »

⁶⁵ <http://datasealofapproval.org>

⁶⁶ <http://www.dans.knaw.nl/>

Research Libraries (CRL) et le rapport « Trustworthy Repositories Audit & Certification : Criteria and Checklist » (TRAC)⁶⁸ publié en février 2007 et qui proposait un ensemble de 14 critères, et la norme DIN 31644⁶⁹ (34 critères d'évaluation), basée sur le travail de nestor et support au processus de certification du nestor Seal of Approval⁷⁰.

Tous ces référentiels intègrent les aspects organisationnels, techniques et juridiques d'un système d'archivage pérenne, et sont utiles dans l'évaluation des solutions (cf. point 4.2.1.1).

Les systèmes d'archivage pérenne des e-journaux constituent une instanciation spécifique de systèmes d'archivage numérique, et doivent présenter, sur tous ces critères, des droits et fonctionnalités adaptés à leur mission, et en particulier répondre aux problématiques d'accès pérenne explicitées en point 1.3. Un ensemble de 7 critères spécifiques pour les solutions d'archivage à long terme dédiées à la préservation des publications scientifiques numériques, adaptés aux besoins des bibliothèques, des éditeurs et de la communauté académique, a été défini en 2006 par le rapport « E-Journal Archiving Metes and Bounds : A Survey of the Landscape »⁷¹, publié par le Council on Library and Information Resources. Ces critères incluent avant tout la définition claire d'un mandat d'archivage pérenne des e-journaux et la possession des droits pour le faire, mais aussi la possibilité d'accès – au moins dans certains cas – au contenu archivé pour les bibliothèques, la participation à un réseau et la définition d'une couverture suffisante des contenus.

Nous avons retenu quatre critères spécifiques caractérisant des solutions d'archivage pérenne des e-journaux avec mandat spécifiquement dédié par rapport à d'autres solutions d'archivage pérenne de ressources électroniques :

- le fait qu'elles disposent d'un mandat clair et d'une mission dédiée à l'archivage des e-journaux ;
- le fait qu'elles soient le résultat d'une collaboration entre les éditeurs et les bibliothèques pour la mutualisation des ressources et des responsabilités, et la définition de droits d'archivage au bénéfice de tous ;
- le fait qu'elles proposent des conditions d'accès suffisantes, c'est-à-dire accessibles en ligne aux usagers des bibliothèques, idéalement via les systèmes de gestion de l'information existants (résolveurs de liens), et sans restriction d'embargo ;
- le fait qu'elles aient une vocation à une couverture large, multidisciplinaire.

3.2 Les solutions sans mandat spécifiquement dédié

Plusieurs solutions intégrant l'archivage pérenne de la documentation électronique, et souvent associées à la problématique, indépendamment de leurs qualités intrinsèques, n'ont pas pour objectif premier la préservation des périodiques électroniques, et ne remplissent pas – actuellement en tous cas – l'ensemble des critères que nous avons considérés ci-dessus comme minimaux pour

⁶⁷Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510

⁶⁸Dale, R. L., et Bruce Ambacher. *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, 2007. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf. Le TRAC a été utilisé pour évaluer LOCKSS et Portico (cf. plus bas).

⁶⁹Information und Dokumentation - Kriterien für vertrauenswürdige digitale Langzeitarchive <http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907&languageid=de&bcrumblevel=3>.

⁷⁰http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html

⁷¹Kenney, Anne R., Richard Entlich, et Peter B. Hirtle. *E-Journal Archiving Metes and Bounds: A Survey of the Landscape*, septembre 2006. <http://www.clir.org/pubs/reports/pub138/pub138.pdf>. Cet ensemble de critères se base sur le travail réalisé en 2000 par la Digital Library Federation (<http://www.diglib.org>).

une solution d'archivage des périodiques électroniques pour les bibliothèques universitaires : soit en termes d'accès, soit en termes de couverture, soit en termes de mandat, voire, dans certains cas, en termes de garanties d'archivage pérenne (définies plus en détail en 4.1.1). Certaines de ces solutions sont toutefois intéressantes comme modèles ou pistes de collaboration pour la BICfB.

3.2.1 Les dépôts institutionnels

Les dépôts institutionnels répertorient et archivent l'ensemble des publications des membres d'une institution, y compris, pour une part en tous cas, le texte intégral de ces dernières. Toutefois, ils ne peuvent, à l'heure actuelle, constituer une réponse suffisante à la problématique de la préservation pérenne des e-journaux :

- en raison de leur couverture : ils visent les contenus publiés au sein des institutions, non les contenus utiles pour, voire cités par, les publications des institutions ; seuls, ils ne peuvent concurrencer les collections des bibliothèques ;
- parce qu'ils constituent en majorité des *light archives*, qui ne présentent peut-être pas (encore) toutes les garanties d'archivage à long terme ;
- parce que, même en imaginant que la couverture de l'ensemble des dépôts existants au niveau national, institutionnel et international, soit actuellement suffisante pour concurrencer les collections des bibliothèques (ce qui est plutôt un objectif), tous les contenus ne sont pas en accès ouvert – et la perspective n'est pas proche. Les dépôts institutionnels ont toutefois le mérite de tenter de modifier le modèle de contrôle d'accès éditeurs.

3.2.2 Les logiciels de préservation numérique (commerciaux et Open Source)

Les logiciels de préservation numérique propriétaires existent depuis plusieurs années et ont contribué à forger le champ de la préservation numérique. DIAS (IBM) a été utilisé dès 2003 par e-Depot et KOPAL (cf. infra 5.2 et 5.3), la solution de Tessella est utilisée par les Archives nationales du Royaume-Uni, Rosetta (ExLibris) a été développé en collaboration avec la Bibliothèque nationale de Nouvelle-Zélande et mis en production en 2009. Il est utilisé par des bibliothèques, archives et d'autres grandes organisations en Australie, Europe (notamment la Bibliothèque Nationale de Bavière, ETH-Zürich, la KULeuven) et Amérique du Nord.

Les logiciels de ce type proposent des fonctionnalités larges et largement paramétrables, destinées à préserver de grandes collections d'objets numériques de différents formats et structures⁷², et sont dès lors très adaptés pour des organisations telles que les bibliothèques nationales, ou des collections numérisées. Cette robustesse demande en contrepartie un investissement financier non négligeable⁷³. Par ailleurs, il s'agit d'outsourcing à un logiciel propriétaire, ce qui, dans le cas de la préservation pérenne, constitue un risque. Finalement, en termes d'accès, de droits et de couverture, une solution telle que Rosetta, destinée principalement aux contenus numériques détenus et hébergés par les institutions est moins adaptée aux e-journaux natifs numériques souscrits par licence auprès des éditeurs. En effet, tout le bénéfice des solutions collaboratives en termes de négociation de droits et d'ingestion est absent, ces étapes étant du ressort des bibliothèques, ce qui place les bibliothèques dans une position de faiblesse et augmente la charge de travail.

⁷² Rosetta a fait l'objet d'un test de charge avec la Church of Jesus Christ of Latter-day Saints dont les résultats sont publiquement disponibles.

⁷³ Nous ne disposons pas d'information précise sur le coût de souscription, mais il est sans commune mesure avec le coût de LOCKSS ou Portico.

Très récemment se sont développés également des logiciels de préservation numérique Open Source tels Archivematica (release 1.0 le 22 janvier 2014) et DAITSS. Archivematica est développé par Artefactual System, en collaboration avec le sous-comité Technologie de l'Unesco, et différentes grandes organisations canadiennes et américaines du monde des bibliothèques, archives et musées (dont le MoMa et plusieurs bibliothèques universitaires). Le logiciel utilise une méthodologie de développement agile, et propose un résultat très modulaire, sous la forme d'une suite de briques logicielles permettant tous les processus de l'archivage numérique, de l'ingestion à l'accès, en passant par la migration). DAITSS⁷⁴ est développé par le Florida Center for Library Automation (FCLA) et utilisée par la Florida Digital Archive (FDA), un dépôt numérique partagé par 11 universités publiques de Floride. En décembre 2010, après 5 années de production, le logiciel a été réécrit sous la forme d'un ensemble de Webservices RESTful et est devenu Open Source (DAITSS2). DAITSS est adapté aux *dark archives*, mais ne permet pas, en tant que tel, l'accès en ligne ou public ; il ne gère pas non plus la création des *archival packages*. Ces logiciels ont l'avantage d'être Open Source et gratuits, mais, seuls, souffrent de la même limitation que les logiciels propriétaires : il s'agit de solutions avant tout techniques qui n'incluent pas de services d'ingestion et de négociation des droits spécifiques aux e-journaux.

3.2.3 Les plateformes d'archives

Les plateformes d'archives Open Access, telles PubMed Central⁷⁵, ArXiv⁷⁶ ou Persée⁷⁷ sont des initiatives intéressantes en termes d'accès (ouvert soit dès le dépôt, soit après une période d'embargo n'excédant généralement pas 12 mois). Elles proposent également certaines garanties d'archivage à long terme : serveurs et localisations multiples, missions établies dans la durée (PubMed Central), financement national (PubMed Central) archivage pérenne garanti (Persée, dont l'archivage pérenne est garanti par le CINES – qui est aussi chargé du volet « archivage pérenne » de licencesnationales.fr, cf. infra 5.1). Cependant, même dans les meilleures conditions, elles ont certaines limites :

- en termes de couverture : elles ont pour mission de préserver la littérature scientifique d'un domaine en particulier, mais ce contenu ne correspond pas nécessairement aux contenus des contrats des institutions, ce qui complexifie la gestion (en particulier dans le cas d'Arxiv, où la gestion se fait au niveau de l'article et non du périodique) ;
- en termes de contrôle : leurs missions ne sont pas définies par rapport aux bibliothèques, et celles-ci n'ont pas de poids dans la gouvernance ou les décisions de l'archive.

Ces solutions sont donc plutôt des solutions d'accès alternatives, par exemple pour assurer l'accès post-abonnement à des contenus archivés dans une *dark archive*, sans qu'elles puissent remplacer une solution plus large pour l'archivage pérenne des e-journaux.

Les plateformes thématiques qui proposent un accès ouvert et des garanties d'archivage pérenne suffisantes pourraient aussi intervenir en complément des solutions dédiées (cf. point 7.1) pour les titres non couverts par ces solutions.

C'est le cas par exemple pour l'Archeology Data Service⁷⁸, plateforme fondée en 1996 par un consortium de départements d'archéologie d'universités et le Conseil de l'archéologie anglais (BCA) dans le but de fournir des données numériques de haute qualité et pérennes pour la recherche en

⁷⁴ <http://daitss.fcla.edu/>

⁷⁵ <http://www.ncbi.nlm.nih.gov/pmc/>

⁷⁶ <http://arxiv.org>

⁷⁷ <http://www.persee.fr/>

⁷⁸ <http://archaeologydataservice.ac.uk/>

archéologie, en tant que l'une des branches thématiques de l'Arts and Humanities Data Service (AHDS)⁷⁹.

Ceci suppose qu'elles soient enregistrées dans le Keeper's Registry (cf. point 4.2.3.1) ou un système équivalent, afin de pouvoir les identifier.

3.3 Les solutions avec mandat spécifiquement dédié

Trois solutions d'archivage (LOCKSS, CLOCKSS et Portico) ont vu le jour pour répondre spécifiquement à la problématique de la préservation pérenne des e-journaux. Elles sont toutes trois issues des résultats du programme « E-journals archiving » de la A. Mellon Foundation⁸⁰ qui, au début des années 2000, finança plusieurs recherches destinées à explorer les possibilités techniques et économiques de créer une archive numérique viable.

Ces solutions sont basées sur la collaboration, une définition claire des rôles et responsabilités entre les éditeurs et les bibliothèques et la mutualisation des ressources, et ont vocation à archiver tous les périodiques électroniques scientifiques (même si en pratique, ce n'est pas – et ne sera probablement jamais – le cas), payants et Open Access. Portico et LOCKSS ont étendu récemment leur couverture à la préservation des e-books, et permettent aussi, selon des modalités un peu différentes, la préservation d'autres contenus numériques détenus par les bibliothèques (via les D-collections dans Portico, et les réseaux privés LOCKSS (PLN)).

Bien qu'elles répondent toutes trois à l'objectif défini, elles sont très différentes dans leur philosophie et leur technologie (LOCKSS et CLOCKSS sont basées sur la préservation distribuée alors que Portico est une archive centralisée). Elles se distinguent également par leur portée en matière d'accès : Portico et LOCKSS proposent un accès post-abonnement alors que CLOCKSS se limite à l'archivage pérenne.

On présentera ici aussi e-Depot, le système d'archivage pérenne mis en place par la Bibliothèque nationale des Pays-Bas (KB), à l'origine en lien avec le dépôt légal. E-Depot est un système centralisé dont la gouvernance n'est pas partagée (à la différence des trois autres solutions), mais qui propose un archivage pérenne des périodiques dont les conditions d'accès sont similaires à celles de CLOCKSS.

Les différentes initiatives sont présentées brièvement ci-dessous ; une présentation plus détaillée et en parallèle de LOCKSS, CLOCKSS et Portico, touchant les différents aspects des systèmes d'archivage pérenne (aspects organisationnels, techniques, juridiques et fonctionnement) se trouve dans l'Annexe 2.

3.3.1 LOCKSS⁸¹

LOCKSS (« Lots of Copies Keep Stuff Safe ») est un logiciel Open Source créé en 2000⁸² par les Bibliothèques de l'Université de Stanford, afin de permettre aux bibliothèques de garantir, à faible coût, l'archivage pérenne et l'accès perpétuel à leurs contenus électroniques (e-journaux, e-books,

⁷⁹ Il est à noter toutefois que l'AHDS ne reçoit plus de financement au niveau national depuis 2008, le financement de sa maintenance ayant été repris par le JISC.

⁸⁰ <http://www.diglib.org/preserve/ejp.htm>

⁸¹ Nous remercions ici Victoria Reich, directrice exécutive de LOCKSS, qui a montré une disponibilité continue et sans faille pour répondre à toutes nos questions concernant LOCKSS.

⁸² Développement initié en 1998, version alpha en 2000, version beta entre 2002 et 2004 testée avec 50 bibliothèques, mise en production en avril 2004.

collections numérisées, DI...) sous la forme d'un système de préservation distribuée⁸³. Il est géré depuis 2005 par la LOCKSS Alliance⁸⁴, organisation sans but lucratif composée des bibliothèques membres du Global LOCKSS Network (cf. infra), qui finance l'équipe technique basée à Stanford (en 2012, 10 personnes + 5 chercheurs associés (HP Palo Alto, Intel Berkeley, Harvard, Sun)⁸⁵). Celle-ci apporte une assistance technique pour l'installation et l'utilisation de LOCKSS (y compris le développement de plug-ins), assure la maintenance et le développement du logiciel, et promeut le projet auprès des éditeurs et des bibliothèques. Les bibliothèques payent une souscription annuelle pour la participation à la LOCKSS Alliance, la participation des éditeurs est gratuite.

La philosophie du projet est avant tout de redonner aux bibliothèques leur rôle de conservation en leur garantissant la présence « en local » du contenu qu'elles ont acheté, pour des coûts abordables, comme dans le modèle de l'imprimé, mais LOCKSS offre aussi aux éditeurs un moyen de préserver leurs contenus sous leur forme publiée, tout en garantissant aux lecteurs un accès continu : "Libraries keep what they buy, publishers preserve what they publish, readers have continual access"⁸⁶.

LOCKSS s'assure contre les risques touchant la préservation à long terme **via une stratégie collaborative (distribution, duplication et collaboration, monitoring permanent par les pairs)**, et un principe d'**économie** s'appuyant sur le plus de **simplification et d'automatisation** possible des processus (ingestion, migration, accès). L'objectif est de limiter les coûts d'une part, et d'assurer la qualité des archives par un processus de monitoring et de réparation permanent.

L'utilisation du logiciel LOCKSS nécessite toujours un réseau (serveurs informatiques et institutions partenaires) mais peut s'implémenter sous deux modalités différentes (et distinctes) :

- via le réseau global LOCKSS (Global LOCKSS Network – GLN) pour préserver les publications électroniques scientifiques (e-journaux, e-books) qui intéressent suffisamment de participants (les « collections générales » des bibliothèques) ;
- via un réseau privé (Private LOCKSS Network – PLN), pour les contenus numérisés ou natifs qui appartiennent à une communauté spécifique (collections audio, images, databases, documents...) ou pour les contenus éditeurs qu'une communauté souhaite préserver indépendamment du Global LOCKSS Network (pour des raisons de contrôle, de propriété des données ou de spécialisation). Dans le cas d'un PLN, les responsabilités économique, de gouvernance, voir les négociations et certains développements, sont assurés par la communauté privée. 12 PLN existent actuellement (juillet 2013)⁸⁷, régionaux, nationaux ou internationaux, thématiques ou généralistes. Certains de ces réseaux sont gérés par la LOCKSS Alliance (ADPN, COPPUL, PeDaLs), ou indépendamment (LUKII, MetaArchive, Datapass). L'intérêt de ces réseaux est multiple du point de vue de la BICfB : au-delà des partenariats possibles, ils proposent des solutions techniques et organisationnelles dont la

⁸³ Présentation sur le site de LOCKSS (consulté le 21 juin 2013) : "The LOCKSS Program, based at Stanford University Libraries, provides libraries and publishers with award-winning, low-cost, open source digital preservation tools to preserve and provide access to persistent and authoritative digital content."

⁸⁴ La Grande-Bretagne a fondé la UK LOCKSS Alliance afin de gérer leur Réseau national LOCKSS (mais les contenus sont dans le GLN et l'assistance technique n'est qu'en partie délocalisée).

⁸⁵ Source : F. Lecocq (2013), *Cours sur les plateformes logicielles*.

⁸⁶ Source : Page d'accueil de www.lockss.org.

⁸⁷ [Alabama Digital Preservation Network \(ADPN\)](#), [CLOCKSS Archive](#), [Council of Prairie and Pacific University Libraries \(COPPUL\) Consortium](#), [Data Preservation Alliance for the Social Sciences \(Data-PASS\)](#), [Digital Commons](#), [Digital Federal Depository Library Program](#), PNL de l'[Instituto Brasileiro de Informaçõem Ciência e Tecnologia \(IBICT\)](#), [LukII](#) (LOCKSS und KOPAL Infrastruktur und Interoperabilitat), [MetaArchive Cooperative](#), [PNAS Proceedings of the National Academy of Sciences](#), [Persistent Digital Archives and Library System – PeDALs](#), [Synergies](#).

BICfB pourrait s'inspirer. Certains projets ont implémenté des couches logicielles ou composants supplémentaires à LOCKSS, en particulier pour la migration planifiée des données, absente dans LOCKSS. Metaarchive utilise l'outil Conspectus⁸⁸ pour gérer la description des collections par les membres, LUKII a intégré KoLiBRI avec LOCKSS (cf. infra), l'intégration de LOCKSS et Archivemata est également testée⁸⁹ ;

- LOCKSS étudie actuellement⁹⁰ l'intérêt d'utiliser des PLN pour conserver des contenus « généraux » actuellement gérés uniquement au sein du GLN. Cette piste est intéressante pour grouper des synergies tout en offrant plus de souplesse que le GLN, en particulier en matière de choix de contenus archivés.

Le Global LOCKSS Network préserve également des e-books.

L'archivage avec LOCKSS⁹¹

- L'éditeur autorise l'accès à son contenu en publiant un « manifeste » d'autorisation sur sa plateforme (une formule d'autorisation explicite sur une page html)⁹², pour chaque contenu autorisé pour l'archivage (en pratique, un volume ou une année de périodique) ; c'est généralement la LOCKSS Alliance qui négocie cet accès avec les éditeurs, mais une bibliothèque peut s'en charger.
- La bibliothèque (adhère à la LOCKSS Alliance)⁹³ installe et configure le logiciel LOCKSS sur un serveur dédié (sous Linux, un PC milieu de gamme suffit en tant que serveur⁹⁴, un disque RAID est recommandé pour le stockage⁹⁵, actuellement, la LOCKSS Alliance recommande 4 Tb) pour créer une « LOCKSS Box ».
- La bibliothèque active dans la LOCKSS Box le moissonnage des contenus auxquels elle a droit sur les plateformes des fournisseurs pour lesquels des plug-ins d'ingestion⁹⁶ sont déjà disponibles dans LOCKSS.
- Pour les nouveaux contenus, elle doit obtenir au moins six « votes » au sein de la LOCKSS Alliance – c'est-à-dire qu'au moins six autres institutions doivent se déclarer prêtes à préserver le contenu – afin qu'il y ait suffisamment de copies pour en assurer la préservation. Si l'éditeur accepte d'autoriser l'accès à ces contenus via le manifeste, l'équipe technique de LOCKSS établit alors un plan de préservation et écrit les plug-ins nécessaires (la bibliothèque peut également le faire, et bénéficier ainsi, en pratique, d'une réduction de frais de souscription).

⁸⁸ <http://metaarchive.org/metawiki/index.php/Conspectus>

⁸⁹ Le sujet a été abordé lors du 2013 Private LOCKSS Networks Community Meeting (Alabama, octobre 2013), notamment par le Dr. Anthony Leroy (ULB), que nous remercions chaleureusement par ailleurs pour son ouverture à la collaboration et pour le temps qu'il a bien voulu nous consacrer pour nous parler du projet en cours de réseau privé LOCKSS initié par l'ULB en 2012 ainsi que de son expérience avec LOCKSS.

⁹⁰ Communication téléphonique de Victoria Reich, directeur exécutif de LOCKSS, en juillet 2013.

⁹¹ Un Guide d'installation est disponible sur le site web de LOCKSS: <http://www.lockss.org/locksswp/wp-content/uploads/2011/12/LOCKSS-Linux-Install1.pdf>; un guide rapide d'utilisation est disponible sur le site de la UK LOCKSS Alliance http://www.lockssalliance.ac.uk/files/2012/01/LOCKSS_Quick_Start_Reference_v1.2.pdf

⁹² Par exemple "LOCKSS system has permission to collect, preserve, and serve this Archival Unit" ou une licence Creative Commons.

⁹³ L'adhésion à la LOCKSS Alliance est nécessaire pour avoir accès aux contenus non Open Access disponibles dans le Global LOCKSS Network. En pratique, l'équipe de LOCKSS accepte que l'adhésion soit payée à la fin de la première année après avoir testé le système.

⁹⁴ La configuration minimale recommandée dans le guide d'installation de LOCKSS est la suivante: CPU 1GHz compatible Intel x86, 1 Gb de mémoire (2 sont recommandés), un lecteur CD ou DVD, un contrôleur SATA.

⁹⁵ L'utilisation d'un disque sans la technologie RAID est possible, mais nécessitera un temps de reconstruction du contenu de la LOCKSS Box plus important (via le réseau de pairs) en cas de crash du disque.

⁹⁶ Un plug-in est un fichier xml qui indique à LOCKSS comment et où il doit harvester les contenus, en fonction de la structure des urls. Il faut un plug-in pour chaque groupe de publications partageant la même structure. Si elles le souhaitent, les bibliothèques peuvent aussi écrire des plug-ins, un générateur écrit en Java est disponible. Le paramétrage de l'ingestion des contenus comprend aussi la définition de la fréquence d'ingestion.

- Le logiciel utilise le moissonnage (web harvesting) (avec vérification des autorisations via IP par l'éditeur⁹⁷) pour parcourir et collecter le contenu, qu'il monitore et répare régulièrement en fonction des informations disponibles dans les autres LOCKSS boxes du Réseau.
- Attention, l'ingestion dans la LOCKSS box de l'institution doit être effectuée au moment où l'accès au contenu est autorisé pour les institutions, donc dans certains cas avant la fin de l'abonnement.
- En cas d'inaccessibilité du contenu sur la plateforme de l'éditeur, quelle qu'en soit la raison (interruption temporaire ou momentanée du serveur, abonnement expiré...), l'accès est fourni par la LOCKSS box, soit via proxy, soit via caching, soit via intégration dans le résolveur de liens (automatiquement et de manière transparente pour l'utilisateur, en préservant toutes les informations originales, y compris le contexte d'affichage).

Editeurs participants	Plus de 600 éditeurs
Bibliothèques participantes	Environ 150 ⁹⁸
Contenu archivé	Plus de 10.000 titres de périodiques e-books (pas de données sur le nombre)
Souscription annuelle	De \$2.300 à \$11.515 par an selon la taille de l'institution ⁹⁹ pour les bibliothèques, gratuit pour les éditeurs

Tableau 2 - LOCKSS en chiffres, février 2014

3.3.2 Portico¹⁰⁰

Portico est une plateforme d'archives électroniques pérennes fondée en 2002 par JSTOR, transférée en 2004 à l'organisation sans but lucratif Ithaka (qui gère aussi JSTOR), et lancée en 2005. Portico formule sa mission de la manière suivante : « préserver les e-journaux, e-books, et autres contenus électroniques scientifiques pour assurer que les chercheurs et étudiants y auront accès dans le futur »¹⁰¹.

Portico est gérée par le conseil d'administration d'Ithaka, composé de membres académiques, éditeurs et bibliothécaires, et par un comité de conseil spécifique composé de 13 éditeurs (5) et bibliothécaires d'institutions importantes (8) mais très majoritairement américaines (California Digital Library, University of California, Cornell University, New-York University, Harvard University, Lesley University, Wiley, ACM, University of Chicago Press, Cengage Learning), à la notable exception des Pays-Bas (Bibliothèque nationale et Wolters-Kluwer) et de l'Irlande (Trinity College, Dublin).

Portico a été conçu spécialement comme une archive pérenne « tierce », en réponse aux problématiques spécifiques adressées au début des années 2000¹⁰², et vise la préservation des contenus via une gestion commune et équilibrée éditeurs – bibliothèques (qui contribuent tous financièrement sous la forme de souscriptions annuelles), un modèle économique viable et d'une infrastructure sécurisée mais centralisée. L'approche technique respecte les normes et standards courants et est documentée. La gestion est entièrement centralisée : les éditeurs versent les

⁹⁷ La LOCKSS Box doit donc avoir une adresse IP statique.

⁹⁸ L'information n'est pas publiée sur le site web de LOCKSS, mais mentionnée dans la littérature.

⁹⁹ Tarifs pour les institutions américaines, publiés sur <http://www.lockss.org/join/>

¹⁰⁰ Nous remercions Catherine Kosturski, Coordinatrice commerciale pour l'Europe, et Javanica Curry, International Participation Director, qui ont aimablement répondu à l'ensemble de nos questions sur Portico.

¹⁰¹ Page d'accueil de www.portico.org.

¹⁰² Notamment par le Research Libraries Group. (2002). Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report. Mountain View, CA. Consulté à l'adresse <http://oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>

contenus à l'archive, celle-ci effectue les opérations techniques assurant leur préservation, et gère les droits d'accès. L'accès aux contenus préservés est lié à cette philosophie top-down et n'est possible qu'en cas d'événement entraînant la disparition définitive du contenu sur d'autres plateformes, sauf si l'éditeur autorise l'accès post-abonnement (ce qui est le cas, en juin 2013, pour plus de 80% des titres).

Pour les bibliothèques et les éditeurs, Portico fonctionne selon le même principe qu'une assurance¹⁰³ : tant qu'elle souscrit, une bibliothèque peut accéder aux contenus auxquels la bibliothèque a droit (collections souscrites antérieurement en cas de désabonnement ou collections souscrites en cours en cas d'autre événement déclencheur).

Portico a lancé également récemment des services D-Collection pour préserver les journaux, images et autres sources primaires, qui fonctionnent de la même manière que le service de préservation des e-journaux et des e-books.

L'archivage avec Portico

- Un contrat est établi entre Portico et l'éditeur.
- Portico établit un plan de préservation et développe les outils nécessaires à l'ingestion et à la migration.
- L'éditeur transfère [périodiquement] le contenu concerné vers Portico via ftp ou support physique, OAI-PMH ou logiciel développé par Portico et Portico l'ingère selon le plan de préservation, en extrayant les métadonnées nécessaires et en effectuant les migrations et reconditionnements nécessaires.
- Portico monitore et gère le contenu, sur base quotidienne (vérification d'intégrité et réparations du contenu) et annuelle (migrations si nécessaire, vérification des plans de préservation) ; les bibliothèques ont aussi un accès à l'archive à des fins d'audit et de monitoring (schéma dans Kirchhoff, 2006).
- Si le contenu n'est plus accessible ni chez l'éditeur, ni ailleurs, Portico rend, sur sa plateforme ou via les résolveurs de liens, le contenu accessible aux bibliothèques membres. Si l'éditeur l'autorise, le contenu est accessible aussi pour l'accès post-abonnement uniquement aux bibliothèques autorisées.
- La bibliothèque peut devenir membre au moment où elle le souhaite, l'accès au contenu de Portico est rétroactif, sur base des autorisations confirmées par les éditeurs. C'est fort utile en cas de désabonnement massif (au contenu d'un grand éditeur par exemple), ou en cas d'interruption de la souscription à Portico pour une ou quelques années, mais la durée de rétroaction est limitée en pratique par la durée de conservation des contrats.

Editeurs participants	236 éditeurs (représentant plus de 2.000 sociétés et associations)
Bibliothèques participantes	Plus de 900
Contenu archivé	Plus de 17.000 titres de périodiques Plus de 220.000 e-books
Souscription annuelle	De moins de \$1.000 à \$24.720 selon le budget en acquisitions documentaires de l'institution ¹⁰⁴ , entre \$250 et \$81.960 pour les éditeurs selon leur chiffre d'affaires ¹⁰⁵ . Un supplément de

¹⁰³ La comparaison est de Portico.

¹⁰⁴ Tarif officiel, publié sur <http://www.portico.org/digital-preservation/join-portico/for-libraries>. Pour les institutions avec un budget annuel d'acquisitions documentaires de moins de 150.000\$, la souscription équivaut à 1,03% du budget en acquisitions documentaires.

¹⁰⁵ Tarif officiel, publié sur <http://www.portico.org/digital-preservation/join-portico/for-publishers>

souscription est nécessaire pour bénéficier du service d'archivage des e-books.

Tableau 3 - Portico en chiffres, février 2014

3.3.3 CLOCKSS

CLOCKSS (« Controlled LOts of Copies Keep Stuff Safe ») est une initiative d'éditeurs académiques (AMA, Elsevier, American Physiological Society, IOP, Nature, SAGE, OUP, SIAM, Springer, Taylor & Francis, Wiley) et de bibliothèques (Allemagne, Australie, USA, Italie, Canada, Ecosse, Hong Kong) académiques majeurs fondée en 2006 et destinée à construire une archive viable, géographiquement distribuée, qui assure la préservation à long terme des publications scientifiques électroniques pour le bénéfice de la communauté scientifique globale. Les contenus préservés dans CLOCKSS (d'éditeurs commerciaux ou Open Access) qui cessent d'être accessibles ailleurs sont mis en accès ouvert, gratuitement, sous licence Creative commons, sur le portail de CLOCKSS ou via un résolveur de liens.

CLOCKSS, comme son nom l'indique, utilise le logiciel LOCKSS pour préserver les contenus – il s'agit en fait d'un Private LOCKSS Network, distribué géographiquement entre 12 bibliothèques importantes dans le monde entier (USA, Canada, Hong Kong, Japon, Australie, Ecosse, Berlin, Italie). Le Réseau est fermé mais les participants participent au CLOCKSS Advisory Council, qui se réunit au moins une fois par an, virtuellement. CLOCKSS ne gère pas l'accès post-abonnement.

8 titres ont déjà été mis en accès ouvert sur les serveurs CLOCKSS¹⁰⁶.

L'archivage avec CLOCKSS

- Comme avec LOCKSS, l'éditeur autorise l'accès à son contenu en publiant un « manifeste » d'autorisation sur sa plateforme, pour chaque contenu autorisé pour l'archivage (en pratique, un volume ou une année de périodique), ou bien il transmet les fichiers source via ftp.
- 3 CLOCKSS Boxes (Indiana University et Stanford Library) ingèrent le contenu et établissent la version d'autorité en comparant les 3 copies ingérées.
- Les Preservation LOCKSS Boxes (les serveurs des 9 autres participants) ingèrent le contenu depuis les CLOCKSS Boxes d'ingestion.
- Le contenu est préservé grâce à un système d'audit et réparation automatique entre les CLOCKSS Boxes, mais le contrôle d'authenticité entre les LOCKSS Boxes s'effectue via SSL.
- Si un contenu archivé cesse d'être accessible ailleurs, et que le comité CLOCKSS prend la décision de le rendre accessible, le contenu est automatiquement migré vers le format le plus récent et copié sur deux serveurs accessibles publiquement (EDINA data Center et Stanford University), où il est accessible directement ou via des résolveurs de liens (CrossRef, résolveurs locaux des bibliothèques ou liste en ligne).

Editeurs participants	180 éditeurs
Bibliothèques participantes	246 institutions
Contenu archivé	Plus de 9.000 titres de périodiques
Souscription annuelle	De \$450 à \$15.000 selon le budget en acquisitions documentaires de l'institution, entre \$200 et \$25.000 pour les éditeurs selon leur chiffre d'affaires.

Tableau 4 - CLOCKSS en chiffres, février 2014

¹⁰⁶ Liste disponible sur : http://www.clockss.org/clockss/Triggered_Content. Tarif sur http://clockss.org/clockss/Contribute_to_CLOCKSS

3.3.4 e-Depot

Après un travail initié dans les années 90, la Bibliothèque nationale des Pays-Bas (KB) a mis en place en 2003 e-Depot, un système de préservation numérique basé sur le logiciel DIAS d'IBM et la norme OAIS, afin de préserver les publications électroniques soumises au dépôt légal sur le long terme. Pour répondre à la problématique de la difficulté de situer géographiquement les publications électroniques, le service International e-Depot avait l'objectif de proposer un archivage pérenne pour les publications scientifiques d'éditeurs internationaux¹⁰⁷. La recherche et les projets liés à e-Depot ont régulièrement bénéficié du soutien du gouvernement des Pays-Bas. Le site web de la KB indique aussi que "In the coming years the KB intends to develop a sustainable business model for the e-Depot which will reflect both public and private responsibility for our digital scholarly and cultural heritage."

L'accès aux publications archivées dans e-Depot n'est pas défini génériquement, mais est déterminé par les accords de chacun des éditeurs avec la KB. De manière générale toutefois, l'accès aux contenus archivés qui sont encore disponibles sur d'autres plateformes (éditeur ou autre) est limité à la consultation sur place, à l'exclusion de l'accès en ligne, sauf dans le cas des publications Open Access¹⁰⁸. Lorsqu'un contenu n'est plus disponible sur une autre plateforme, le contenu est rendu accessible sur le site web d'e-Depot. En 2013, 2 publications avaient été « triggered » et rendues accessibles via la page d'e-Depot.

En avril 2009, e-Depot et le DOAJ annonçaient le lancement d'un projet pilote destiné à préserver tous les journaux du DOAJ. Un accord a ensuite été signé entre le DOAJ et e-Depot, et en février 2013, e-Depot préservait "92.000 Open Access articles from 650 publishers, supplying 900 titles"¹⁰⁹, soit un tout petit peu plus de 10% des titres présents dans e-Depot.

En 2013, e-Depot était en train de modifier son système d'archivage (et d'abandonner DIAS), afin de répondre à l'exigence d'un archivage à plus grande échelle et d'objets numériques plus variés qu'auparavant, tout en adaptant en continu le système aux « best practices » de la préservation numérique¹¹⁰. En février 2014, l'adresse du site web d'e-Dépôt n'était pas disponible. Les listes de titres ne sont malheureusement pas à jour sur le site de la KB ni dans le Keeper's Registry (cf. point 4.2.3.1), ce qui est dommage car il serait intéressant de voir dans quelle mesure e-Depot est complémentaire des autres solutions tierces d'archivage, en particulier au sujet des titres Open Access.

4 LOCKSS, CLOCKSS et Portico : évaluation

Un ensemble limité de six critères d'évaluation des solutions d'archivage a été défini, sur base de ceux décrits – de manière souvent très détaillée – dans la littérature et dans les normes et initiatives d'évaluation des systèmes d'archivage pérenne (généraux ou spécifiquement consacrés aux e-journaux) mais en incluant aussi les aspects spécifiques à la prise de décision que sont la question du

¹⁰⁷ 14 éditeurs + plateformes participaient à l'initiative en février 2014 : Elsevier, BioMed Central, Nederlands Tijdschrift voor Geneeskunde, Blackwell Publishing, Oxford University Press, Taylor & Francis, Sage Publications, Springer, Atlantis Press, Koninklijke Brill, Directory of Open Access Journals, International Union of Crystallography, IOS Press, Mary Ann Liebert.

¹⁰⁸ <http://kb.nl/en/expertise/long-term-usability-of-digital-resources/information-for-international-publishers>

¹⁰⁹ Commentaire de Marcel Ras, e-Depot Manager, sur le blog de D. Rosenthal : <http://blog.dshr.org/2013/02/rothenberg-still-wrong.html>.

¹¹⁰ L'url renseignée jusqu'en 2012 (<http://www.kb.nl/hrd/dd/index-en.html>) n'est plus accessible en février 2014, sans url de remplacement.

contrôle effectif sur la solution par les bibliothèques et la question du coût qui à la fois en indiquent la faisabilité et en mesurent le risque.

L'objectif de l'examen des différentes solutions dédiées à la lumière de ces critères est de répondre aux questions suivantes : quelle solution / ou combinaison de solution(s) présente les plus grandes garanties ? Ces garanties sont-elles suffisantes pour investir dans ces solutions pour les titres électroniques ? Quels seraient les coûts ? Ces coûts sont-ils supportables par les institutions de la BICfB ? Quelle part du portefeuille des institutions BICfB pourrait être garantie par les solutions tierces ?

En fonction de ces réponses, des points d'action pourront être proposés à la BICfB (chapitre 7).

4.1 Critères d'évaluation

4.1.1 Les garanties d'archivage pérenne

Comme le soulignait déjà l'étude Metes and Bounds¹¹¹, et Rosenthal e.a. en 2005¹¹², il y a un large consensus sur la nature des risques menaçant les archives pérennes : **risques physiques** : pannes de l'infrastructure (hardware, media, logiciels, réseaux) et catastrophes naturelles, **risques humains** : erreur ou attaque intentionnelle (interne ou externe), **risques techniques** : obsolescence des formats et des media, **risques organisationnels** : faillites organisationnelles (financière¹¹³, légale, managériale, sociétale).

Une série de stratégies globales sont considérées comme des garanties minimales face à ces risques et doivent donc être présentes dans les différentes solutions : **stratégies techniques et de gestion** : redondance géographique (réplication) et sécurisation du système (authentification), monitoring des media et des données, contrôle qualité (lors des différentes étapes du processus d'archivage), utilisation de formats non-propriétaires et stratégies de migration (de formats et media) ou équivalentes ; **stratégies organisationnelles** : indépendance administrative et organisationnelle, droits correspondant aux missions et objectifs assignés ; gouvernance équilibrée ; viabilité du *business model*.

D. Rosenthal et l'équipe de LOCKSS¹¹⁴ résumaient fort bien les stratégies de préservation pérenne comme des principes généraux s'appliquant à tous les aspects des systèmes (software, hardware, métadonnées, organisation, processus, responsabilité) : la duplication (géographique, organisationnelle), la transparence (de gestion, du logiciel, des formats, de l'organisation), la diversité (hardware, software, fundings...), l'audit/contrôle (authenticité, complétude...), l'économie (dans les processus d'ingestion, migration, accès, tant au niveau de la gestion que de la technique).

Dans l'évaluation des différentes solutions dédiées, nous distinguerons, pour faciliter l'analyse, les garanties d'archivage pérenne organisationnelles et financières, d'une part, et les garanties techniques et de gestion, de l'autre.

¹¹¹ Kenney, Entlich, & Hirtle, 2006, Metes and bounds, p. 35.

¹¹² Rosenthal, D., Th. Robertson, T. Lipkis, S. Morabito, et V. Reich. « Requirements for Digital Preservation Systems. A bottom-up approach ». D-Lib Magazine 11, n° 11 (2005). <http://lockss.org/locksswiki/files/DlibRequirements.pdf>.

¹¹³ L'un des exemples de perte de données pour manque de moyens est précisément celui de l'archive numérique CICNet Journals, créée en 1991 dans le but d'archiver les e-journaux, et qui cessa ses opérations en 1997 par manque de financement (voir : <http://www.atlasofdigitaldamages.info/v1/stories/cicnet-e-journal-archive/>).

¹¹⁴ Rosenthal, Robertson, Lipkis, Reich, & Morabito, 2005.

En ce qui concerne les garanties techniques, il n'existe pas à l'heure actuelle, à ma connaissance, de large consensus sur les meilleures solutions d'implémentation¹¹⁵. En particulier, pour les systèmes d'archivage pérenne des e-journaux :

- la redondance géographique peut être réalisée notamment via des sites miroirs ou via un réseau peer-to-peer, utilisant différentes technologies hardware (disques durs locaux ou en réseau, cloud...);
- le monitoring des media et des données et le contrôle qualité et les stratégies de réparation peuvent être manuels ou automatiques ou une combinaison des deux, réguliers ou non, et utiliser différentes procédures;
- les objets peuvent être convertis, avant d'être stockés, en un format standard (par exemple PDF/A) ou stockés dans leur format original; ils peuvent être ingérés dans le système sous la forme de « fichiers sources » (complets) ou de fichiers de présentation (correspondant aux fichiers publiés) – chaque méthode présentant des avantages et des inconvénients;
- la migration peut avoir lieu sur base temporelle régulière ou sur événement déclencheur (accès, check...), manuellement ou automatiquement, ou présenter une combinaison des différentes stratégies selon les contenus. En pratique, au moins deux approches et philosophies concurrentes existent: celle, la plus répandue, qui prend des mesures prophylactiques – et coûteuses – pour contrer l'obsolescence prévue des formats (migration à intervalles de temps réguliers, vérification de la validité des formats lors de l'ingestion dans l'archive...) et la philosophie minimaliste de *migration on access* que défend D. Rosenthal (LOCKSS), qui relativise le danger d'obsolescence des formats et estime que les mesures de prévention ne sont pas pertinentes en termes d'analyse coût-bénéfice¹¹⁶. Certaines solutions, notamment e-Depot, croient en une stratégie d'émulation¹¹⁷.

En principe, les différentes solutions sont conformes à l'OAIS; elles proposent des stratégies pour couvrir les risques techniques et sont suffisamment robustes, mais l'audit des systèmes constitue une source utile pour l'évaluation de la qualité ou de certains points faibles des solutions. Au-delà de la robustesse, une question importante reste l'évaluation de la performance des systèmes (lors de la migration, du stockage et de l'accès). Comme le souligne M. Seadle, "Librarians should not, for example, judge the quality and reliability of an archive's migration services merely on the fact that they exist, but on how effectively they function at the right time"¹¹⁸. Malheureusement, à l'heure actuelle, il n'existe pas de métrique partagée ni de tests standardisés¹¹⁹ et les audits des solutions sont ponctuels et basés principalement sur l'examen de la documentation et des informations fournies par les systèmes.

¹¹⁵ Rosenthal, Robertson, Lipkis, Reich, & Morabito, 2005, soulignaient déjà, dans leur approche illustrant différentes techniques utilisées pour adresser différentes problématiques de la préservation pérenne, que tous les systèmes cités satisfaisaient à l'OAIS.

¹¹⁶ Voir Rosenthal, D. S. H. (2010). Format obsolescence: assessing the threat and the defenses. *Library Hi Tech*, 28(2), 195-210. doi:10.1108/07378831011047613, et les discussions sur le blog de D. Rosenthal.

¹¹⁷ La KB a mené en 2005-2007 le projet Dioscuri – emulation (<http://www.kb.nl/en/expertise/e-depot-and-digital-preservation/emulation/project-emulation-dioscuri>).

¹¹⁸ Seadle, M. « Archiving in the networked world: metrics for testing ». *Library Hi Tech* 29 (2011): 557-564, p. 8.

¹¹⁹ Pour une explication de la problématique et des recommandations, voir Michael Seadle, Andreas Rauber, Adam Rusbridge, Sabine Schrimpf, Matt Schultz, « Technical Alignment » in *Aligning National Approaches to Digital Preservation*, McGovern, Nancy Y, Katherine Skinner. Atlanta, Ga.: Educopia Institute Publications, 2012. http://www.educopia.org/sites/default/files/ANADP_Educopia_2012.pdf, p. 169 et ss. L'article définit 5 critères minimaux pour la mise en place de tests valides dans la communauté de la préservation digitale: des objectifs clairs, des données de test suffisamment complexes pour refléter la réalité, des échelles de mesure, une base de connaissance des résultats des tests.

Aucun test n'a été effectué dans le cadre de la présente étude, et l'évaluation technique des solutions est fondée sur l'examen de la documentation fournie par les systèmes et de la littérature, en particulier les rapports d'activités et d'audit¹²⁰. Elle privilégie la description globale des mécanismes d'archivage, en utilisant la terminologie du modèle OAIS (Open Archival Information System), et la mise en évidence de certains points faibles identifiés dans la littérature, sans viser à une analyse exhaustive de la technologie des solutions.

En fonction des éléments à sa disposition, la BICfB pourra décider, dans une seconde phase, d'effectuer des tests d'un ou plusieurs systèmes ou une analyse technique plus approfondie.

4.1.2 L'accès

Si les modalités d'accès sont toujours partie intégrante des solutions d'archivage pérenne (l'accès devant toujours, à un moment ou à un autre, être fourni sous une forme ou sous une autre à une communauté donnée), elles sont particulièrement importantes dans le cadre de la préservation des e-journaux, dans la mesure où, d'une part, les modalités d'accès courant constituent la référence pour les usagers dans le cas de la littérature scientifique¹²¹ et où, d'autre part, l'accès est soumis aux conditions spécifiées dans la licence éditeur.

La définition et l'évaluation de l'accès dans les solutions d'archivage pérenne des e-journaux comprennent plusieurs éléments, qui couvrent à la fois des aspects techniques et juridiques :

- en premier lieu, les événements déclencheurs autorisant l'accès (en particulier l'autorisation ou non de l'accès post-abonnement) – ceux-ci doivent être négociés par la solution d'archivage auprès des éditeurs – et l'(im)édiateté de l'accès en cas d'événement déclencheur (délai et nécessité d'interventions humaines) ; rappelons que les solutions d'archivage pérenne peuvent fonctionner comme des *dark archives* (et proposer un accès uniquement si les contenus ne sont plus disponibles ailleurs) ou comme des *light archives*, et gérer à la fois la préservation et l'accès perpétuel ;
- la richesse du contenu fourni aux usagers (présentation originale ou non, incluant ou non les fonctionnalités hypertexte, citations...) ;
- les droits d'usage associés (impression et téléchargement, réserve de cours, mais aussi ILL) – semblables à ceux de la licence accordée par l'éditeur ou restreints par la solution ;
- l'intégration des solutions dans les systèmes d'accès propres aux bibliothèques (catalogues, résolveurs de liens...), la transparence d'accès et la performance technique.

4.1.3 La couverture

La couverture des solutions, c'est-à-dire les contenus qui y sont potentiellement archivables (et accessibles) par les institutions, est un critère important, qui distingue les solutions d'archivage pérenne dédiés aux e-journaux des solutions d'archivage pérenne de contenus électroniques non spécifiquement dédiés aux e-journaux. En effet, la couverture n'est pas, en premier lieu, une question de capacité de stockage, mais de droits d'ingestion des contenus négociés par les solutions

¹²⁰ LOCKSS Audit Report, November 2007 Altman, M., & Crabtree, J. (2011). TRAC-Based Auditing of LOCKSS: Using the SAFE-Archive System - [archiving2011_altman_crabtree.pdf](http://thedata.org/files/thedata_new2/files/archiving2011_altman_crabtree.pdf). Présenté à Archiving 2011. Consulté à l'adresse http://thedata.org/files/thedata_new2/files/archiving2011_altman_crabtree.pdf (basé sur la version précédente des critères TRAC) ; CRL Report on Portico Audit 2010. (2010). Consulté à l'adresse <http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf> ; Reilly Jr., B. F. (2008). Summary of the Test Audits Portico and LOCKSS. *Charleston Advisor*, 9(3), 61-62.

¹²¹ Ce qui est renforcé par le fait que la littérature scientifique sous format électronique est en majorité récente. Difficile pour l'utilisateur d'accepter des restrictions sur la consultation d'articles publiés seulement quelques années auparavant.

(indépendamment ou via les institutions) auprès des éditeurs. Ces derniers sont plus ou moins enclins à travailler avec une solution en fonction des facilités (techniques et financières) et des garanties (de contrôle d'accès et de préservation) qu'elle lui offre.

Du point de vue des institutions, une solution d'archivage pérenne des e-journaux doit soit pouvoir couvrir un périmètre suffisant des collections électroniques (présentes ou futures) des bibliothèques, en termes de titres et d'années, soit pouvoir s'intégrer harmonieusement dans un scénario global. Le périmètre s'évalue dans l'absolu en fonction du nombre de titres et d'éditeurs ou producteurs différents couverts par la solution, mais aussi, en particulier, en fonction de l'adéquation de cette couverture avec le portefeuille des institutions et l'échelle des contenus prioritaires définie par l'institution.

Dans le cadre de cette étude, nous avons distingué la couverture des contenus payants de celle des contenus Open Access, et, au sein des premiers, les périodiques ou bouquets de périodiques souscrits en tant que sources primaires auprès des éditeurs ou d'agrégateurs, et les périodiques disponibles via des bases de données full-text. Au sein des périodiques payants souscrits auprès des éditeurs ou agrégateurs, nous avons également distingué les titres courants des bouquets d'archives, car ils constituent des achats distincts et posent par là des questions différentes.

L'évaluation de la couverture des solutions implique aussi des questions de gestion et de publicité de l'information et des accès : l'accès à la liste des contenus archivés par des solutions tierces, la qualité de celle-ci et les possibilités de comparaison de cette liste avec celles des contenus d'autres solutions d'archivage, d'une part, l'intégration avec les résolveurs de liens et les outils de gestion des abonnements des institutions, d'autre part, sont fondamentaux.

4.1.4 Le degré de contrôle

Nombre de bibliothèques, et les institutions de la BICfB ne font pas exception, ont déjà pu expérimenter le hiatus existant entre les garanties contractuelles et les garanties réelles, et les effets collatéraux de l'outsourcing (augmentation non contrôlée des prix, temps d'implémentation de nouvelles fonctionnalités ou de résolution des problèmes...).

Ceci étant, plus les institutions ont un contrôle effectif sur la solution d'archivage, plus les garanties de conservation pérenne sont élevées. Evidemment, le contrôle a un coût, soit matériel (hardware, software), soit en ressources humaines (développements, gestion, gouvernance...), mais l'absence totale de contrôle est, à notre avis, un risque à ne pas courir.

La transparence et la documentation sont les premières conditions de contrôle. Le contrôle peut s'effectuer, ensuite, selon les cas :

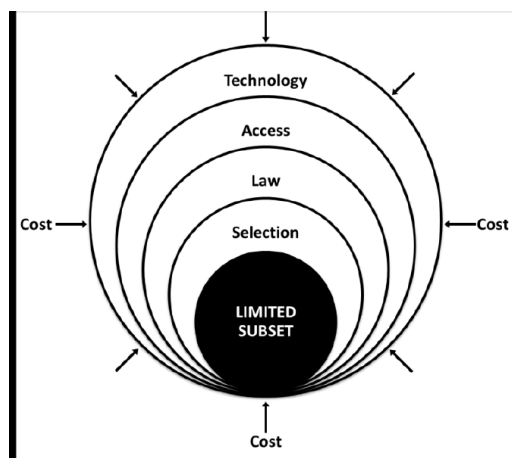
- sur les données physiques : via l'hébergement local ou le monitoring à distance ;
- sur les aspects logiciels : avec un logiciel Open Source et/ou la participation au développement ;
- sur la gouvernance et les aspects juridiques : via la participation aux organes décisionnels.

4.1.5 Les coûts

Les coûts sont malheureusement, actuellement, la cause principale expliquant l'absence de participation ou d'initiatives des bibliothèques universitaires en matière de préservation des collections électroniques¹²². On peut avancer deux éléments d'explication : (1) l'investissement dans des solutions tierces est jugé trop important, voire inutile par les bibliothèques, dans un contexte de

¹²² Voir entre autres Kastlelec, M. (2012). Practical Limits to the Scope of Digital Preservation. Information Technology & Libraries, 31(2), 63-71.

restriction budgétaire¹²³, et (2) la planification des coûts à long terme en matière d'archivage pérenne est extrêmement difficile – le risque économique est à la base de tous les autres.



(Kastellec, 2012) **Figure 1. Nested Model of Limitations**

Or, comme le soulignait déjà Stephen, 1994¹²⁴ *“Librarians need to be equally aware of the costs of caring for an item during its useful life when assessing its potential value for their collections, and need to accept the responsibility for the inevitable cost implications of acquisition.”* Une estimation des coûts la plus précise possible des coûts de l'archivage pérenne permet d'évaluer cette responsabilité.

Dans le cas des solutions tierces, lorsqu'elles ne sont pas gratuites, l'évaluation des coûts est largement facilitée puisque la souscription annuelle couvre une partie importante des coûts globaux de la solution, et une large part des activités techniques. Pour comparer les coûts des solutions, nous devons d'abord : 1) identifier, pour chaque activité de l'archivage pérenne des e-journaux, les coûts couverts et non couverts par la souscription à la solution et 2) tenter d'évaluer les coûts non couverts par la souscription.

Les initiatives sur le sujet de l'évaluation des coûts pour la conservation pérenne des données numériques, basées sur le principe de cycle de vie du document¹²⁵, sont nombreuses depuis le milieu de la dernière décennie¹²⁶. Toutefois, aucun modèle ne s'est encore imposé comme un standard. C'est la raison pour laquelle le projet 4C¹²⁷ (2012-2013) tente de faire le point et de fournir des outils utiles aux différentes communautés pour l'estimation des coûts de l'archivage numérique¹²⁸.

¹²³ “Libraries will need to be convinced that the threat is sufficiently real over the medium to long term to justify an annual expenditure – even a relatively modest one, from stringent budgets.” (Morrow, Beagrie, & Jones, 2008).

¹²⁴ Stephen Andy (1994), *The Application of Life Cycle Costing in Libraries: A Case Study Based on Acquisition and Retention of Library Materials in the British Library*, IFLA Journal, 20(2), 1994, 130-140.

¹²⁵ Etabli par Stephen (1994) pour la première fois dans le monde des bibliothèques.

¹²⁶ Pour un panorama actuel des initiatives, voir le rapport du projet 4C. Pour une revue de la littérature plus ancienne, voir la revue de la littérature du projet LIFE, très complète.

¹²⁷ <http://4cproject.eu/>

¹²⁸ <http://4cproject.eu/news-and-comment/4c-blog/44-4c-s-cost-model-evaluation> cite quelques aspects négatifs des modèles qui ont pu empêcher leur adoption large : 1) manque de guides pratiques d'utilisation (bien utiles étant donné la multiplicité des paramètres), 2) les modèles existants ne sont pas facilement combinables ni interopérables entre eux, 3) manque d'information sur la manière de récolter certains paramètres financiers nécessaires aux modèles.

Les projets d'évaluation des coûts de l'archivage pérenne, même s'ils englobent un champ bien plus vaste que les coûts liés aux solutions tierces d'archivage des e-journaux, sont utiles dans notre contexte également car ils permettent :

- d'aider à modéliser les activités (les modélisations se basent généralement sur un modèle OAIS adapté) ;
- d'aider à définir les coûts – en fournissant notamment des étalons, formules et paramètres financiers – pour le stockage physique des données et les temps d'activité pour chacun des aspects.

Actuellement, comme le souligne le projet 4C, il y a encore un long chemin à parcourir avant de disposer d'un modèle et de données partageable car standard. Dans le cadre de la présente étude, nous nous sommes limités à les consulter plutôt comme « garde-fous » à notre propre estimation des coûts, basée sur un schéma d'activités simple.

Des modèles de coûts partagés et plus détaillés pourraient être utiles à la BICfB dans un second temps pour :

- mettre en perspective les coûts des solutions tierces avec les coûts supposés de mise en place d'une solution entièrement *in house* pour la BICfB ;
- définir les alternatives techniques (stockage des données, préservation logique...) les moins onéreuses à mettre en place¹²⁹ ;
- estimer les coûts liés à des perfectionnements des solutions tierces (dans le cas d'ajout de couches logicielles supplémentaires à LOCKSS par exemple).

Si les perspectives à long terme sont incertaines, on peut espérer que la mise en commun des ressources et la diffusion de bonnes pratiques au niveau international devraient peu à peu réduire les coûts en ressources humaines (temps d'analyse, temps de développement...) des activités impliquées dans la préservation pérenne des contenus numériques.

4.2 Evaluation et comparaison de LOCKSS, CLOCKSS et Portico

Les trois solutions internationales d'archivage électronique dédiées aux e-journaux sont évaluées en fonction de chacun des critères définis ci-dessus, afin d'évaluer dans quelle mesure elles peuvent être adaptées aux besoins de la BICfB, de définir leurs forces et leurs points faibles, et de définir des points d'action.

4.2.1 Garanties d'archivage pérenne

4.2.1.1 Retours d'expérience et audits

LOCKSS et Portico ont été auditées par le CRL il y a déjà plusieurs années (LOCKSS en 2007¹³⁰, Portico en 2010)¹³¹ et ont été considérées globalement comme des solutions acceptables pour la préservation à long terme des journaux scientifiques électroniques. L'audit mettait cependant en évidence certains points faibles de chacune des solutions, au niveau technique et organisationnel, ainsi que des recommandations associées.

¹²⁹ Cf. Rosenthal et al., 2012, pour une étude brillante des coûts du stockage des données.

¹³⁰ L'audit a eu lieu en septembre - octobre 2006, avec une brève mise à jour en janvier 2007, soit pendant la phase de test de la procédure d'audit par le CRL, ce qui explique la systématisation moindre des critères que pour l'audit de Portico, réalisé entre avril et octobre 2009, le rapport ayant été finalisé en janvier 2010.

¹³¹ Un audit de CLOCKSS est en cours, et devrait être terminé en mai 2014 (source : <http://www.crl.edu/news/9550>).

Dans le cas de LOCKSS, les questions concernaient la viabilité du modèle économique – tout en soulignant que LOCKSS affirmait avoir trouvé un équilibre financier depuis 2007 sur base de la participation des bibliothèques à la LOCKSS Alliance, le manque de présence d’une communauté de développement du logiciel Open Source suffisante au moment de l’audit, et quelques inquiétudes sur la *migration-on-access*, qui n’avait encore, au moment de l’audit, été testée que de manière assez limitée, sans que la solution ait implémenté des liens vers des dépôts de convertisseurs de formats. Les autres retours d’expérience disponibles confirment que LOCKSS fonctionne à tous niveaux comme prévu¹³², mais certaines critiques¹³³ concernent le manque de *user-friendliness* de l’interface de gestion, qui liste basiquement les unités d’archivage par éditeur sans hiérarchie supplémentaire décrivant les collections.

Dans le cas de Portico, les questions concernaient la problématique du plan de succession – aucune organisation n’étant désignée explicitement comme successeur officiel – le manque de transparence (documentation trop faible¹³⁴, notamment au niveau des responsabilités et procédures, mises à jour du logiciel ; impossibilité d’auditer techniquement la solution), la capacité technique du serveur à gérer des accès en masse en cas de *trigger event* important (mais Portico affirmait que l’utilisation prévue de la plateforme JSTOR devait régler ce problème), ainsi que la couverture de la solution¹³⁵ et l’outil de comparaison des *holdings*, le processus d’ingestion (impliquant des opérations de normalisation de la part de Portico), le problème de délai d’accès en cas de *trigger event*, et les relations entre Portico et JSTOR, considérées comme représentant à la fois un risque et une opportunité.

Même si les solutions ont évolué et répondu à certaines recommandations de l’audit (notamment une publication de documentation et d’une possibilité d’audit minimale dans le cas de Portico), les problèmes principaux signalés en 2007 et 2010 restent d’actualité.

Les questions de viabilité économique et du plan de succession seront abordées dans le point 4.2.1.3. Nous abordons brièvement ci-dessous la question de la performance de l’ingestion et de la migration, points techniques sensibles et représentatifs, liés aux choix philosophiques et fonctionnels des solutions.

4.2.1.2 Garanties techniques

La description point par point des choix techniques des différentes solutions est disponible dans l’Annexe 2, et une description plus détaillée de leur fonctionnement – ou de certains aspects de

¹³² “In all instances known to the auditor, the LOCKSS boxes performed as expected and access to content was restored” (CRL Audit, 2007), cité par Final Report of the 2CUL LOCKSS Assessment Team, 2011. <http://2cul.org/sites/default/files/2CULLOCKSSFinalReport.pdf>, p. 10. Voir aussi Dalton, P., & Conyers, A. (2008). Evaluation of the JISC UK LOCKSS Pilot. Consulté à l’adresse <http://www.era.lib.ed.ac.uk/handle/1842/3336>

¹³³ Dalton, P., & Conyers, A. (2008) et entretien téléphonique avec Fabrice Lecocq (Directeur Délégué aux Projets, INIST-CNRS) en février 2014. Nous en profitons pour remercier Fabrice Lecocq pour le temps qu’il a accepté de nous consacrer et les informations très utiles qu’il nous a transmises, en particulier sur la distribution actuelle des responsabilités de l’archivage pérenne en France, sur le groupe PIN (Pérennisation des Informations Numériques) et ses travaux, et sur sa longue expérience avec LOCKSS.

¹³⁴ Portico a mis en ligne en 2009 une série de «preservation policies» sur des points techniques et organisationnels spécifiques (<http://www.portico.org/digital-preservation/about-us/portico-resource>). Cependant, ces documents restent souvent au niveau de la déclaration de principe et n’ont pas été mis à jour depuis 2009.

¹³⁵ Portico comprenant à l’époque 83 éditeurs et 10.461 titres de périodiques, ce qui représentait à peine 50% du contenu disponible dans CrossRef.

celui-ci – est accessible dans la littérature¹³⁶. De manière générale, LOCKSS met l'accent sur l'automatisation maximale des processus afin de réduire les risques humains, sur le monitoring permanent et sur la diversité technologique (format, hardware), et sur une architecture logicielle et une infrastructure délibérément simples et légères – donc économiques, et ce sont là ses principaux points forts. Une autre force du logiciel, puisqu'il est Open Source¹³⁷, est la possibilité d'utiliser d'autres couches logicielles pour des fonctionnalités spécifiques (migration ou gestion logique). Portico, à l'inverse, met en avant les principes de centralisation et de gestion globale et régulière dans un respect des principes et standards (technologie, métadonnées) de la préservation pérenne, de définition claire des responsabilités et de participation aux réflexions institutionnalisées de la communauté internationale¹³⁸. L'utilisation de formats ouverts est privilégiée, ainsi que l'utilisation, pour la migration, des plug-ins Open Source dans la mesure du possible, mais la technologie reste propriétaire.

LOCKSS moissonne les contenus sur le site web de l'éditeur et les préserve tels quels (*presentation file*), alors que Portico effectue une ingestion manuelle, contrôlée et normalisée de fichiers fournis par l'éditeur (*source file*) dans des formats limités et prédéfinis.

Portico normalise les fichiers reçus et envisage une migration régulière et planifiée de l'archive (*batch migration*) à des fins de préservation, alors que LOCKSS privilégie une migration lors de l'accès (*migration on access*). L'investissement humain et matériel important nécessaire pour l'ingestion et la migration dans le cas de Portico (normalisation des fichiers, vérification et correction des formats, traitements par lot) est justifié par l'idée qu'un contrôle des formats est nécessaire pour la préservation pérenne : des formats (de documents et de métadonnées) non valides ou trop anciens mettent en péril la pérennité des ressources puisqu'ils risquent de n'être plus lisibles à long terme. Un autre avantage, mais qui va à l'encontre du principe de diversité, est la relative uniformisation des formats de l'archive, qui garantit également une limitation des développements pour les migrations futures. Un risque de cette méthode est la performance des migrations, qui sont des migrations à grande échelle, avec tous les coûts et les risques liés¹³⁹.

Or, dans le cas des formats des ressources en ligne, cette idée d'obsolescence des formats est mise en cause par D. Rosenthal et l'équipe de développement de LOCKSS, qui soulignent au contraire « la mort lente des formats web »¹⁴⁰. En pratique, dans la *migration on access*, la migration est postposée

¹³⁶ LOCKSS: Maniatis, Petros, Mema Roussopoulos, Thomas J. Giuli, David SH Rosenthal, et Mary Baker. « The LOCKSS peer-to-peer digital preservation system ». ACM Transactions on Computer Systems (TOCS) 23, no 1 (2005): 2–50; Pour une rapide introduction: Reich, V., et D. Rosenthal. « Distributed digital preservation: Lots of copies keep stuff safe ». In Proceedings Indo-US Workshop on International trends in Digital Preservation March, 24–25, 2009. <http://humgroup.lockss.org/locksswiki/files/ReichIndiaFinal.pdf>. Portico: Morrissey, Sheila M., Matthew Stoeffler, Vinay Cheruku, William J. Howard, John Meyer, et Suresh Kadirvel. « Migration at Scale: A Case Study » (2012). <http://www.portico.org/digital-preservation/wp-content/uploads/2012/11/MigrationAtScale.pdf>.

¹³⁷ Disponible sur SourceForge : <http://sourceforge.net/projects/lockss/>. Toutes les versions du code sont préservées en cvs.

¹³⁸ <http://www.portico.org/digital-preservation/about-us/affiliations-memberships> ;
<http://www.portico.org/digital-preservation/about-us/community-involvement> .

¹³⁹ Voir pour une illustration de ces risques et coûts l'article Morrissey, S. M., Stoeffler, M., Cheruku, V., Howard, W. J., Meyer, J., & Kadirvel, S. (2012). Migration at Scale : A Case Study. Consulté à l'adresse <http://www.portico.org/digital-preservation/wp-content/uploads/2012/11/MigrationAtScale.pdf>, écrit par des membres de l'équipe de développement de Portico. Ils mentionnent notamment que la réalisation effective de la migration a duré plus d'un an (3 mois de planification, 9 mois de réalisation), (l'équipe ne travaillait pas à temps plein sur le projet).

¹⁴⁰ Rosenthal, D. S. H., Lipkis, T., Robertson, T. S., & Morabito, S. (2005). Transparent Format Migration of Preserved Web Content. D-Lib Magazine, 11(01). doi:10.1045/january2005-rosenthal et Rosenthal, D. S. H. (2010). Format obsolescence: assessing the threat and the defenses. Library Hi Tech, 28(2), 195-210.

jusqu'à ce qu'un utilisateur demande l'accès au contenu ; l'outil de migration est intégré dans le pipeline de dissémination et non appliqué au contenu original ; la migration se fait donc en temps réel. Ce système a l'avantage de limiter le nombre et la fréquence des opérations de migration, d'éviter un accroissement de l'archive lié au maintien de fichiers migrés pour l'ensemble du contenu de l'archive et d'éviter les pertes d'information liées à des migrations, et de proposer la technologie de migration la plus récente au moment de l'accès. Le système a été testé en 2005 pour la conversion du format GIF vers le format PNG. LOCKSS prévoyait d'intégrer une API pour des plug-ins de format-convertors, un process de matching entre les types MIME acceptés par les browsers et celui du document, afin d'utiliser un plug-in de conversion si nécessaire ; un registre distribué de convertors, similaire à celui des plug-ins de harvesting déjà existants. Le site web de LOCKSS mentionne l'étude de N. Jackson (British Library) d'octobre 2012, qui, après une analyse de 2,5 billions de fichiers datant d'entre 1996-2010, soutient la thèse de D. Rosenthal¹⁴¹.

Dans les deux cas, tant LOCKSS que Portico, la performance réelle de la migration (consommation des ressources, temps nécessaire, fiabilité) reste une question ouverte, car aucun test d'envergure suffisante ou migration réelle n'ont été effectués à notre connaissance¹⁴². Par ailleurs, et c'est un argument en faveur de la philosophie de LOCKSS, aucun format n'a été identifié comme obsolète en 2006 dans le cadre du projet LIFE¹⁴³, ni aucun format web jusqu'à présent. Il est donc difficile d'affirmer que le principe de migration périodique semble plus sécurisant pour l'avenir que le principe de *migration on access*.

Dans le cas de l'ingestion, la question de la performance n'est pas cruciale¹⁴⁴, tant que l'ingestion peut être effectuée dans des délais raisonnables (un an maximum dans le cas de LOCKSS puisque les contenus concernés sont des contenus à souscription au minimum annuelle). L'ingestion dépend à la fois du temps matériel d'ingestion des fichiers dans le système, mais également, pour les contenus de nouveaux éditeurs, du temps de planification et de développement nécessaire pour préparer les fichiers à l'import (création de la *manifest page* et développement des plug-ins dans le cas de LOCKSS, réception et normalisation des fichiers dans le cas de Portico¹⁴⁵). Nous n'avons toutefois pas de données précises sur ces différents délais¹⁴⁶, mais il faut signaler que, dans le cas de LOCKSS, il

doi:10.1108/07378831011047613 ; <http://www.lockss.org/news-media/blog/> pour un récapitulatif complet des billets de D. Rosenthal consacrés au thème de l'obsolescence des formats. D. Rosenthal souligne également dans le même contexte l'inutilité des outils de vérification de formats.

¹⁴¹ <http://www.lockss.org/news-media/blog/> : "In October 2012, Andrew N. Jackson, British Library reported on an analysis of 2.5 billion files dated between 1996 – 2010. His conclusion, "Our initial analysis supports Rosenthal's position; that most formats last much longer than five years, that network effects to appear to stabilize formats, and that new formats appear at a modest, manageable rate." L'étude originale est disponible sur Arxiv.org: <http://arxiv.org/abs/1210.1714> .

¹⁴² Pour LOCKSS, les tests ont été réalisés pour un seul format de fichier (Rosenthal, Lipkis, Robertson, & Morabito, 2005), pour Portico, un seul test est décrit dans la littérature (Morrissey et al., 2012) et concernait la migration du format de métadonnées.

¹⁴³ McLeod, R.; Wheatley, P.; Ayris, P., 2006.

¹⁴⁴ Rosenthal, Lipkis, Robertson, & Morabito, 2005, soulignent que le principe de lenteur (sloth) est un attribut important des systèmes de préservation numérique. LOCKSS implémente un temps d'ingestion long (avec des limites imposées) pour ne pas être en compétition avec les users réels et pour éviter qu'une attaque ne puisse modifier beaucoup de copies en peu de temps. V. Reich a confirmé par mail que le temps d'ingestion du contenu dans le cadre de l'installation d'une LOCKSS box était « conséquent ». Logiquement, le temps de restauration du système devrait lui aussi être conséquent, mais peut être évité en utilisant des disques RAID.

¹⁴⁵ Le processus CronPrep, responsable de l'ajout de métadonnées aux fichiers ingérés a une capacité de traitement, en 2012, de 24 millions d'articles et 240 millions de fichiers par an (avec utilisation de 50 à 75% des capacités maximum) (Morrissey et al., 2012).

¹⁴⁶ Sauf en ce qui concerne l'écriture des plug-ins : un outil de génération de plug-ins (fichiers xml) est disponible, et, sauf difficulté particulière de structure du site de l'éditeur, l'écriture d'un plug-in peut ne demander que quelques heures.

semble que les éditeurs ferment parfois la possibilité de moissonnage après un certain temps, voire une fois l'accès courant terminé. Il est donc primordial d'archiver les contenus au plus tôt.

Nous n'avons pas trouvé d'information mesurable dans la littérature sur la performance d'accès des systèmes en cas d'événement déclencheur pour une quantité importante de contenus et/ou des contenus très consultés. Dans le cas de LOCKSS, celle-ci dépend en partie du matériel et de la configuration du serveur local.

Dans la mesure où aucun test des systèmes n'a été effectué dans le cadre de cette étude, où tant LOCKSS que Portico ont été certifiés par le CRL (TRAC), et où nous ne disposons pas des compétences requises pour juger des choix technologiques effectués par les deux solutions, nous n'avons pas souhaité les départager du point de vue technique.

Garanties techniques	LOCKSS	CLOCKSS	Portico
Philosophie et principes généraux	Logiciel Open Source, automatisation maximale des procédures, principe d'économie et de monitoring permanent ¹⁴⁷		Logiciel propriétaire (mais respect des standards), gestion centralisée et planification, monitoring ponctuel
Performance	<p>Aucun retour négatif de la communauté sur le fonctionnement (monitoring, ingestion, accès automatiques), sauf en termes de difficulté de configuration (interface basique, aide de Stanford souvent nécessaire pour la configuration)</p> <p>Temps d'ingestion semble long ; pas d'informations sur la durée de restauration en cas de crash de la LOCKSS box locale ; capacité de fourniture d'accès dépend de la capacité et de la configuration du serveur local</p>		Nombreuses opérations manuelles (normalisation lors de l'ingestion, migrations planifiées, ouverture de l'accès), qui constituent un risque, en matière de coûts et de délais
	Tests publiés très limités : temps d'ingestion semble long, pas de résultat de test de charge sur la fourniture d'accès (mais dans le cas de LOCKSS, c'est défini en interne)		
Points d'action	<ul style="list-style-type: none"> • Tester la performance du système en termes de temps d'ingestion et de restauration, de fourniture 		

¹⁴⁷ Signalons aussi que le logiciel SAFE permet un audit automatique de LOCKSS. Voir Micah Altman. (2012, décembre 25). Auditing Distributed Preservation Networks. Consulté à l'adresse <http://fr.slideshare.net/drmaltman/auditing-distributed-preservation-networks> et Altman, M., & Crabtree, J. (2011). TRAC-Based Auditing of LOCKSS: Using the SAFE-Archive System - archiving2011_altman_crabtree.pdf. Présenté à Archiving 2011. Consulté à l'adresse http://thedata.org/files/thedata_new2/files/archiving2011_altman_crabtree.pdf

	<p>d'accès, ainsi que l'installation et l'interface de gestion des contenus</p> <ul style="list-style-type: none">• Envisager le développement ou l'identification de modules pour la <i>migration on - access</i>, ou l'ajout d'autres couches logicielles si une migration périodique est souhaitée
--	---

Tableau 5 - Synthèse de la comparaison des garanties techniques de LOCKSS, CLOCKSS et Portico

4.2.1.3 Garanties organisationnelles et stabilité financière

Les risques organisationnels (équilibre et stabilité de la gouvernance, viabilité du *business model*) sont des risques importants pour la préservation pérenne, même si les différentes solutions existent désormais depuis plusieurs années. Il y a à notre avis trois facteurs de risque majeurs : l'absence de solution alternative en cas de scénario du pire, la croissance des coûts et le déséquilibre des acteurs.

4.2.1.3.1 Gouvernance

Portico est géré avant tout par le Conseil d'administration d'Ithaka, assisté par un comité de conseil spécifique composé d'éditeurs et de bibliothécaires d'institutions importantes, majoritairement américaines. Cette dépendance à l'égard d'Ithaka, qui est l'organisation responsable de JSTOR et entretient dans ce cadre d'autres relations avec les éditeurs et les bibliothèques, peut être considérée comme un risque. De même, l'importance des grands éditeurs dans la gestion est un facteur d'équilibre selon Portico, mais pourrait, du point de vue des bibliothèques, mener au déséquilibre.

LOCKSS est gérée par l'Alliance, l'association des membres. Dans les faits, la gestion et le développement de LOCKSS sont de facto dépendants de Stanford, mais la volonté de **LOCKSS** d'éviter tout *single point of failure* se matérialise dans son organisation : utilisation d'un logiciel Open Source, gestion des LOCKSS boxes dans des organisations indépendantes, création de PLN gérés indépendamment. Le problème est la mise en pratique de ces principes : la LOCKSS Alliance existe maintenant depuis 8 ans, il n'y a actuellement pas de réelle communauté de développement active hors Stanford, même si des collaborations techniques sont établies avec l'Angleterre (UK LOCKSS) et l'Allemagne (LUKII).

CLOCKSS est un PLN LOCKSS dont la gouvernance est partagée par les institutions participantes (grandes bibliothèques nationales et de recherche et grands éditeurs scientifiques, sur le plan international). Les mêmes restrictions s'appliquent que dans le cas de Portico, si ce n'est que les bibliothèques possèdent le contenu en local, ce qui, étant donné la nature de ces institutions, présente des garanties suffisantes à notre avis en relation avec l'objectif de CLOCKSS (préservation sans accès post-abonnement). En outre, toutes les institutions participantes peuvent envoyer un représentant au « CLOCKSS Advisory Council », qui conseille le « CLOCKSS Board » en matière de politiques et pratiques de préservation et d'archivage numérique.

4.2.1.3.2 Business model

Tant LOCKSS que Portico ont reçu des bourses et des financements publics ponctuels pour les développements initiaux mais basent actuellement leur *business model* sur les rentrées liées aux souscriptions à la solution, qui utilisent d'ailleurs la même classification et dont le tarif est à peu de choses près équivalentes pour les bibliothèques (cf. point 4.2.5.2.1). Les deux solutions affirment

avoir atteint l'équilibre financier depuis plusieurs années. Alors que LOCKSS favorise l'accès des (petits) éditeurs à la solution, en leur offrant une participation gratuite, les éditeurs sont, pour Portico, des soutiens financiers au même titre que les bibliothèques¹⁴⁸. Les deux solutions visent explicitement la stabilité des frais de souscription dans le temps¹⁴⁹.

L'évolution des partenariats des différentes solutions pose une question fondamentale de viabilité et/ou d'orientation des modèles. Pour LOCKSS, il y avait en janvier 2013 environ 150 bibliothèques participantes¹⁵⁰, seules à payer une participation. Dans le cas de Portico, le nombre de souscriptions a crû de manière très importante tant pour les bibliothèques que pour les éditeurs¹⁵¹, qui contribuent tous financièrement. Les revenus de Portico via la participation des membres sont donc bien plus élevés que ceux de LOCKSS (même en comptant les 13 PLN), et augmentent sensiblement plus rapidement. Ceci est rassurant pour la viabilité de l'organisation, mais pose aussi question : soit LOCKSS dispose de financements ponctuels nettement plus importants, soit les coûts de traitement et/ou de stockage du contenu sont fort importants pour Portico, soit Portico réinvestit massivement dans d'autres voies. Pourtant, Portico facture de manière indépendante la préservation des e-journaux et des autres contenus (e-books, D-Collections), ce qui n'est pas le cas de LOCKSS.

La politique de responsabilité financière de l'ensemble des acteurs de Portico – qui assure une diversité des sources et un certain degré de revenus, peut aussi, peut-être, exclure les contenus, particulièrement à risque, de petits éditeurs, si ceux-ci ne peuvent ou ne souhaitent pas participer financièrement.

Le modèle financier de CLOCKSS est le plus intéressant : les contributions des participants, fort faibles (entre \$450 et \$15.000, avec, pour les bibliothèques de la BICfB, un coût maximal annuel de \$1.800), sont fixées sur 5 ans et ont pour but d'aider CLOCKSS à construire une dotation, de manière à pouvoir réduire les coûts de participation au fil du temps¹⁵².

4.2.1.3.3 Plan de succession

Une difficulté majeure du modèle de **Portico** est l'absence de plan de succession satisfaisant. Portico ne désigne en effet aucune organisation mais prévoit seulement, si Portico devait disparaître, de tenter de trouver une association non-profit qui pourrait lui succéder. Si cette démarche

¹⁴⁸ Les tarifs sont calculés en fonction du chiffre d'affaires des publications des éditeurs et varient entre 250 et 81.960\$ (voir <http://www.portico.org/digital-preservation/join-portico/for-publishers>).

¹⁴⁹ Pour LOCKSS, les fees n'ont pas évolué depuis l'origine pour les membres souscrivant (cf. mail V. Reich + chiffres donnés dans le rapport en 2007, effectivement identiques). Pour Portico, les fees semblent également avoir très peu évolué : elles variaient pour les e-journaux pour les éditeurs entre \$250 et \$75.000 en 2006, et atteignent aujourd'hui un maximum de \$81.960 ; pour les bibliothèques, elles variaient en 2006 entre \$1.500 et \$24.000, et atteignent aujourd'hui un maximum de \$24.720 (ce qui représente une augmentation de 3% en 6 ans).

¹⁵⁰ Rosenthal, D. S., & Vargas, D. L. (2013). Distributed digital preservation in the cloud. Présenté à 8th International Digital Curation Conference, Amsterdam. Consulté à l'adresse <http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf>.

¹⁵¹ 742 bibliothèques participantes en mai 2012, 910 en décembre 2013 ; 13 éditeurs participants en 2006, 83 en 2009 et 189 en décembre 2013.

¹⁵² "Libraries and publishers around the world are invited to help us build an endowment, which is intended to lower participatory fees over time" ([http://clockss.org/clockss/Contribute to CLOCKSS](http://clockss.org/clockss/Contribute%20to%20CLOCKSS)). Voir au sujet de la dotation comme financement des solutions d'archivage pérenne l'article Rosenthal, David S.H., Daniel C. Rosenthal, Ethan L. Miller, Ian F. Adams, Mark W. Storer, et Erez Zadok. « The economics of long-term digital storage ». In Memory of the World in the Digital Age Conference, Vancouver, BC. Retrieved from <http://www.lockss.org/locksswp/wp-content/uploads/2012/09/unesco2012.pdf>, 2012. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Rosenthal_et_al_27_B_1330.pdf

n’aboutissait pas, “Portico will provide a copy of the archival versions of the content back to the content licensor and an appropriate not for-profit institution, such as the Library of Congress, and assign preservation rights to such institution ». Si les éditeurs reçoivent les fichiers, rien n’est donc prévu pour assurer, en cas de force majeure, l’accès des bibliothèques aux contenus souscrits, ni leur récupération.

Dans le cas de **LOCKSS**, il n’existe pas non plus de plan de succession, mais l’existence de PLN et le fait que le logiciel soit Open Source mitigent le risque. Par ailleurs, étant donné la structure de la solution, l’échec de la solution ne mettrait pas en cause l’accès aux contenus déjà archivés, qui resteraient en la possession des institutions. Le même raisonnement est valable pour **CLOCKSS** ; en outre, même si le projet venait à échouer, l’archive serait en possession de 12 bibliothèques ou institutions indépendantes.

4.2.1.3.4 Synthèse

Garanties organisationnelles et financières	LOCKSS	CLOCKSS	Portico
Gouvernance	Gouvernance partagée, mais dépendance forte de Stanford	Gouvernance équilibrée entre grands acteurs éditeurs et bibliothèques	Gouvernance partagée, mais dépendance forte d’Ithaka
Business model / risque économique	Equilibre financier, vise la stabilité des coûts de contribution, mais viabilité à long terme pose question, pas ou peu de croissance du nombre de participants ces dernières années	<i>Business model</i> intéressant : limitation des coûts et modèle financier sur le principe de la dotation de manière à pouvoir réduire les coûts au fil du temps	Equilibre financier, vise la stabilité des coûts, croissance continue du nombre de membres, mais question du coût de l’archive à long terme ¹⁵³ et de l’évolution de la souscription
Plan de succession	Pas de plan de succession, mais données en local et logiciel Open Source, existence de réseaux PLN, mais pas de réelle communauté de développement	Pas de plan de succession, mais données hébergées en local par 12 institutions	Pas de plan de succession suffisant, pas de possibilité de récupération des données par les bibliothèques ; collaboration avec e-Depot assure une certaine sécurité

Tableau 6 - Synthèse de la comparaison des garanties organisationnelles et financières de **LOCKSS**, **CLOCKSS** et **Portico**

4.2.2 Accès

Les conditions et modalités d’accès aux contenus archivés dans les solutions tierces sont extrêmement différentes, dans leurs principes généraux, mais aussi dans leurs modalités pratiques, qui valent la peine d’être examinées en détails.

¹⁵³ Le modèle technique, impliquant de lourdes opérations manuelles, est beaucoup moins économique que celui de **LOCKSS** ou **CLOCKSS**, et les coûts de l’archivage pérenne centralisé sont difficiles à estimer sur le long terme (cf. point 4.2.1.5).

L'accès est, avec le contrôle, le fer de lance de **LOCKSS** et n'est limité que par la volonté des éditeurs (qui doivent donner l'autorisation de moissonnage¹⁵⁴) et les possibilités techniques (durée d'ingestion et écriture des plug-ins, capacité du serveur d'accès). Aucune limitation n'existe en matière de conditions d'accès, qu'il s'agisse des événements déclencheurs, de richesse du contenu présenté ou de droits d'usage associés. En pratique, quelle que soit la raison qui rend le contenu non accessible sur le serveur de l'éditeur (qu'il s'agisse d'une brève indisponibilité technique, d'une situation post-abonnement, de transfert de titre ou de la disparition pure et simple de l'éditeur), la LOCKSS box de l'institution prend le relais et livre instantanément – dans les limites de la rapidité du réseau et du serveur – le contenu, de manière transparente et sous sa forme originale (à l'exclusion des contenus dynamiques comme les publicités), à l'utilisateur. Les droits d'usage du contenu ne sont pas gérés ni limités par la solution et sont, en principe, ceux qui se trouvent dans la licence originelle de souscription au contenu via l'éditeur¹⁵⁵. L'accès correspond donc entièrement au niveau élevé de la demande des usagers et des bibliothèques. En outre, l'activation de l'accès en cas d'indisponibilité du serveur de l'éditeur est entièrement automatique et ne nécessite aucune intervention de la part de l'institution. Il existe une autre limitation potentielle, mais qui reste théorique, d'ordre technique, qui devrait faire l'objet de tests, et éventuellement d'adaptations en termes de choix et de configuration du serveur : la capacité de la LOCKSS Box de l'institution à supporter la charge nécessaire en cas d'indisponibilité de contenus de plateformes très consultées.

L'audit du CRL de 2007 identifiait une faiblesse potentielle dans le fait que les droits étaient cédés par l'éditeur à la bibliothèque et non à la LOCKSS Alliance, celle-ci ne pouvant donc intervenir pour faire globalement respecter les droits des bibliothèques participantes auprès d'un éditeur en défaut¹⁵⁶.

CLOCKSS ne gérant que l'archivage pérenne, la limitation est forte en ce qui concerne les événements déclencheurs : l'accès post-abonnement n'est pas couvert et la décision d'ouverture d'accès pour un titre est prise par le conseil de CLOCKSS, de manière collaborative (ce qui peut entraîner un certain délai). Cependant, passée cette limite, l'accès est extrêmement ouvert puisqu'il est sous licence Creative Commons (les droits d'usage peuvent donc être plus larges que dans le cas des licences éditeurs originales) : il est ouvert à tous et n'est pas soumis au paiement d'une affiliation à CLOCKSS. La plateforme CLOCKSS est aussi compatible avec l'*open url*. Une autre restriction concerne le transfert des titres d'un éditeur à un autre, puisque cet événement n'est pas pris en compte dans la décision d'ouverture d'accès au titre par CLOCKSS.

La politique d'accès de **Portico** est quant à elle assez large en termes d'événements déclencheurs (même si elle l'est moins que LOCKSS) : l'accès post-abonnement est autorisé pour 89%¹⁵⁷ des titres et les autres événements déclencheurs couvrent la faillite de l'éditeur ou la cessation de publication. Toutefois, en pratique, l'accès est soumis à une procédure de demande et de vérification auprès de l'éditeur concerné qui peut durer de 90 à 120 jours en cas de cessation de disponibilité du contenu, et 30 jours en cas d'accès post-abonnement. Par ailleurs, les cas de transferts de titres entre éditeurs sont explicitement exclus, et l'accès via Portico est autorisé « jusqu'à ce que la publication soit à nouveau accessible chez l'éditeur concerné ou un successeur qui

¹⁵⁴ Autorisation technique (*manifest page*) et autorisation contractuelle (cf. le site de LOCKSS pour des suggestions de clauses à intégrer dans les licences). Le moissonnage du contenu n'a donc lieu qu'une fois l'autorisation explicitement accordée, pour chaque unité de contenu.

¹⁵⁵ Les clauses d'usage ont généralement, dans les contrats éditeurs, une portée explicitement non limitée à la durée de la souscription. Des propositions de clauses contractuelles entre éditeur et bibliothèques sont disponibles sur le site de LOCKSS.

¹⁵⁶ CRL Audit 2007, p. 17.

¹⁵⁷ En juin 2013.

aurait hérité des droits »¹⁵⁸. La plateforme est compatible avec les résolveurs de liens, mais l'accès est également limité en termes de présentation du contenu (puisque le contenu archivé ne présente pas toutes les fonctionnalités du contenu publié originellement sur le portail de l'éditeur, les hyperliens notamment peuvent être perdus) et, dans certains cas, en termes de droits d'usage. Les droits d'usage de la licence de souscription originelle sont conservés, mais s'y superposent des limitations liées à la plateforme Portico : il semble, à la lecture des licences¹⁵⁹, que le PIB est interdit dans tous les cas, les téléchargements et copies sont limités, et, en cas d'accès post-abonnement, le dépôt dans des réserves de cours, les dépôts institutionnels et l'intranet est explicitement interdit¹⁶⁰. La simple définition des droits et processus d'accès via Portico laisse donc déjà présager un risque de difficultés et des négociations.

ACTEURS	EVENEMENTS	ACCES aux contenus archivés		
		LOCKSS	CLOCKSS	Portico
Editeur	Cessation de publication	Oui, immédiat	Oui, avec délai, après décision du comité	Oui, avec délai
	Transfert de titre	Oui, immédiat	Non	Non
	Crash serveur, faillite...	Oui, immédiat	Oui, avec délai, après décision du comité	Oui, avec délai
Institution	Fin d'abonnement au titre auprès de l'éditeur	Oui, immédiat	Non	Oui si accès post-abonnement prévu dans le contrat entre l'éditeur et Portico, après demande de l'institution et validation de la demande auprès de l'éditeur.
	Fin de souscription à la solution d'archivage	Oui, immédiat	Oui	Non
Solution d'archivage	Faillite ou crash	Oui, immédiat	Probable	Incertain

Tableau 7 - Synthèse des accès en cas d'événement déclencheur via LOCKSS, CLOCKSS et Portico

Deux points restent à clarifier auprès de Portico en matière d'accès. En premier lieu, le statut des titres Open Access archivés n'est pas clair : normalement, en cas d'événement déclencheur autre que la fin de l'abonnement, les contenus archivés sont rendus accessibles uniquement aux institutions souscrivant à Portico. Il s'agit de vérifier que cette limitation ne s'applique pas aux titres Open

¹⁵⁸ Cf. Portico Publication Licence Agreement (E-journals), v. 3.5, p. 5 (section « Trigger events ») – licence éditeur.

¹⁵⁹ La question de l'application exacte de ces limitations de droits n'est pas claire, il est possible que seuls les contenus ayant fait l'objet d'événements déclencheurs autres que la fin de l'abonnement soient concernés par l'interdiction de prêt interbibliothèque et les limitations en termes de téléchargements et de copies.

¹⁶⁰ Ces limitations ne sont présentes que dans les licences éditeurs, pas dans les licences bibliothèques. En ce qui concerne le dépôt dans les dépôts institutionnels, ce n'est pas fort problématique puisqu'a priori, il aura été fait au moment de la publication. Par contre, l'interdiction du PIB est plus gênante même si elle est compréhensible.

Access ; dans le cas contraire en effet, l'effet pervers serait important puisqu'un titre Open Access pourrait passer à une forme de souscription payante. Une seconde question est liée à la vérification des contrats qu'effectue Portico auprès des éditeurs : si une institution souscrit à Portico longtemps après la fin de son abonnement à certains contenus, il s'agit de s'assurer que les éditeurs conserveront les archives le temps nécessaire.

Tant dans le cas de LOCKSS GLN que dans celui de Portico, le paiement de la souscription annuelle est une condition nécessaire à l'utilisation de la solution (et donc à l'accès). Toutefois, dans le cas de LOCKSS, les données étant hébergées en local dans le cas du GLN, cette condition ne concerne que l'archivage des nouveaux contenus, les contenus archivés restant accessibles et théoriquement transférables dans une autre solution d'archivage.

En termes d'accès, donc, on peut distinguer, comme c'est souvent le cas dans la littérature, entre les solutions qui permettent l'accès post-abonnement (LOCKSS, Portico) et celles qui ne le permettent pas (CLOCKSS). Mais on peut aussi distinguer les solutions qui ont une philosophie d'accès ouvert (LOCKSS, CLOCKSS) et celles qui ont une philosophie d'accès contrôlé, juridiquement et commercialement (Portico). Soulignons aussi que seule la solution LOCKSS garantit un accès pérenne même en cas de transfert de titres d'un éditeur à un autre. Dans le cas de CLOCKSS et Portico, cet accès n'est garanti que dans les limites de la décision de l'éditeur. Il est donc important d'insister sur la signature par l'éditeur du TRANSFER Code of Practice dans le contexte de l'archivage pérenne, si l'éditeur n'accepte pas l'archivage dans LOCKSS.

Accès	LOCKSS	CLOCKSS	Portico
Événements déclencheurs, délais d'accès	<ul style="list-style-type: none"> Tous (y compris transfert de titres, indisponibilité temporaire du serveur) Pas de délais, d'intervention humaine ni de décisions, l'accès est immédiat (sauf limitations techniques liées au serveur) 	<ul style="list-style-type: none"> Uniquement les événements entraînant l'indisponibilité du contenu sur la plateforme de l'éditeur (pas d'accès post-abonnement). Délais d'accès liés à la décision du Board (non précisés) L'accès n'est pas soumis à la participation de l'institution à CLOCKSS 	<ul style="list-style-type: none"> Tous, sauf le transfert de titres ; l'accès post-abonnement n'est pas autorisé par tous les éditeurs. Délais d'accès liés à la vérification auprès des éditeurs (jusqu'à 30 jours pour l'accès post-abonnement et jusqu'à 120 jours dans les autres cas) L'accès est soumis au paiement de la souscription
Richesse et fonctionnalités du contenu archivé et droits d'usages associés	Aucune limitation, ni en termes de fonctionnalités, ni en termes de droits d'usage	<ul style="list-style-type: none"> Fonctionnalités peuvent être limitées si l'éditeur a fait le choix de déposer les fichiers sources ; Accès sous licence Creative Commons 	<ul style="list-style-type: none"> Fonctionnalités peuvent être limitées (hyperliens...) Droits d'usage limités (PIB, téléchargement et impression, dépôt dans d'autres répertoires)
Intégration de l'accès dans l'environnement technique	Accès via proxy ou résolveurs de liens, transparent et immédiat, mais limité par les capacités du serveur de la LOCKSS Box	<ul style="list-style-type: none"> Accès via la plateforme de l'archive (CLOCKSS ou Portico) ou via le résolveur de liens Aucune information sur la performance de la plateforme de l'archive n'est disponible 	

Points d'action	Tester la configuration et la performance de la LOCKSS Box ; éventuellement tests de charge en fonction du nombre de connexions envisagé.	Négocier la signature par l'éditeur du TRANSFER Code of Practice <ul style="list-style-type: none">• Vérifier le statut des titres Open Access archivés• Vérifier la question de la durée d'archivage des contrats auprès des éditeurs
------------------------	---	---

Tableau 8 - Synthèse de la comparaison des modalités d'accès de LOCKSS, CLOCKSS et Portico

4.2.3 Couverture

La comparaison des contenus archivés dans les différentes solutions donne une idée de leur profil et de la richesse de leur couverture (point 4.2.3.2), et permet aussi d'identifier les combinaisons potentiellement intéressantes. Toutefois, le nombre de titres archivés dans chacune des solutions ne représente que le nombre maximum de titres potentiellement préservés pour les institutions ; pour obtenir le nombre réel, il faut comparer ces listes aux portefeuilles des institutions (point 4.2.3.3).

4.2.3.1 Méthodologie et difficultés

The Keeper's registry¹⁶¹, base de données en ligne développée par EDINA¹⁶² et le Centre international ISSN de Paris à la suite du projet PEPRS¹⁶³ du JISC, contient la liste des e-journaux archivés dans une série de solutions d'archivage pérenne, dont LOCKSS, CLOCKSS, Portico et e-Depot. Le projet est très intéressant puisqu'il constitue une source normalisée pour identifier les e-journaux préservés. Malheureusement, le timing de cette étude n'a pas permis de se fonder sur cette source pour comparer le contenu des solutions, car le projet est encore en version beta et seule une recherche sur les métadonnées des e-journaux est actuellement disponible, sans possibilité d'export global¹⁶⁴.

La comparaison des listes de titres a donc été effectuée sur base des fichiers xls téléchargés sur les sites des solutions¹⁶⁵ via une base de données Access.

La comparaison n'est pas simple étant donné le manque d'harmonisation des données source (listes de *holdings* et non de titres, nombreux ISSN ou e-ISSN manquants, pas de titres uniformes ni de noms d'éditeurs univoques¹⁶⁶, format de liste différent pour distinguer les « preserved » et « in

¹⁶¹ <http://thekeepers.org>.

¹⁶² <http://edina.ac.uk/>

¹⁶³ Burnhill, P., & Guy, F. (2010). *Piloting an E-journals Preservation Registry Service (PEPRS)*. *Serials Librarian*, 58(1-4), 117-126. doi:10.1080/03615261003622742

¹⁶⁴ EDINA prévoit de mettre en ligne très prochainement une application permettant aux institutions enregistrées d'uploader des listes de titres pour les comparer avec la base de données, et peut en attendant effectuer la comparaison manuellement sur base de fichiers envoyés par mail. Cependant, d'une part, nous souhaitons, dans le cadre de ce rapport, effectuer différentes analyses complémentaires, qui auraient nécessité plusieurs échanges de fichiers en un processus itératif, d'autre part, nous souhaitons également comparer entre elles les solutions d'archivage.

Tous nos remerciements vont à Fred Guy, Project Manager du Keeper's Registry, pour les informations qu'il nous a transmises et la comparaison de holdings qu'il a bien voulu effectuer.

¹⁶⁵ Cette comparaison est nécessairement ponctuelle et ne reflète pas la réalité exacte des solutions à un temps donné, puisque les fichiers publiés sur les sites des solutions sont le résultat d'exports à des dates différentes. Les fichiers utilisés étaient les fichiers disponibles sur les sites des éditeurs en juillet 2013, soit les données de novembre 2012 pour CLOCKSS, du 25 juin 2013 pour LOCKSS et du 27 juin 2013 pour Portico.

¹⁶⁶ Une part de ces problèmes, en particulier le manque d'harmonisation des noms d'éditeurs, se pose également pour la comparaison avec les portefeuilles de titres.

progress » dans les différentes solutions, information d'années pas systématiquement indiquée en regard des volumes ou l'inverse). Afin de pouvoir comparer les contenus des solutions sur la base la plus large possible, une harmonisation des données sur base de l'ISSN, de l'e-ISSN, puis du titre pour les périodiques sans ISSN et sans e-ISSN a donc été effectuée¹⁶⁷. Les données de holdings (volumes et années préservés et « in progress ») ont ensuite été traitées pour harmoniser leur présentation.

Il existe un biais supplémentaire : les fichiers des solutions listent des records (*holdings*) et non des titres uniques d'e-journaux. Un record est l'association d'un titre à un éditeur, une plateforme et un état de collection, un titre pouvant avoir changé d'éditeur et/ou de plateforme au cours de son histoire, voire être publié sur plusieurs plateformes à un même moment. Par voie de conséquence, non seulement le nombre de records ou *holdings* présents dans les fichiers n'est pas le nombre de titres uniques, mais en outre, comparer les titres et ISSN sans comparer les plateformes et les années ne prend pas en compte ce facteur. A cause de la complexité de la formulation des états de collection, de l'absence et de l'inconsistance des données de plateformes et éditeurs¹⁶⁸, nous n'avons pas pu comparer les années archivées pour chacun des titres dans chacune des solutions.

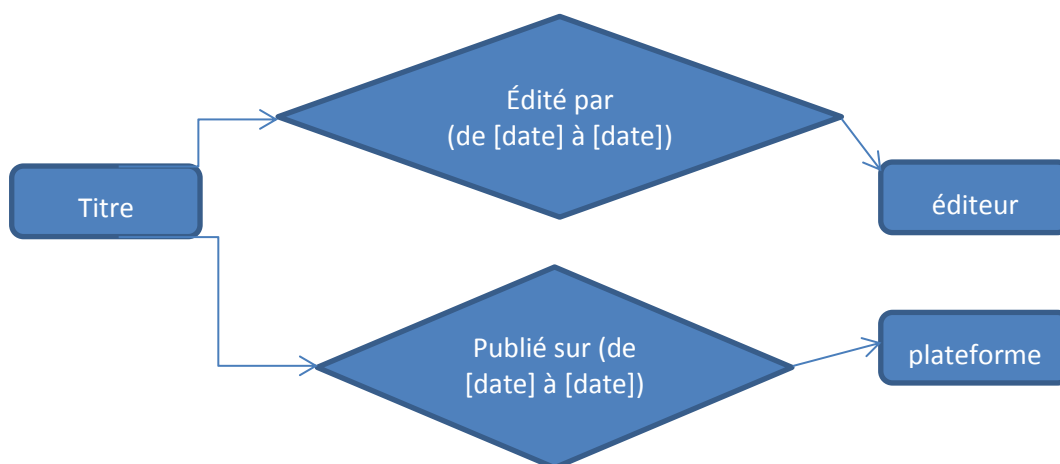


Figure 7 - Modélisation des relations titre – éditeur – plateforme

Pour pallier à ceci dans une certaine mesure, nous avons travaillé avec la notion de seuils pour identifier les éditeurs présents dans les différentes solutions. Les chiffres concernant les nombres de titres restent quant à eux relatifs et ne reflètent pas exactement ce qui pourrait effectivement être archivé par les institutions. Cette remarque est valable pour la comparaison des solutions entre elles, mais aussi pour les portefeuilles des institutions (cf. infra 4.2 .3.3)¹⁶⁹.

¹⁶⁷ Le nombre de titres sans ISSN ni e-ISSN est très faible dans les listes des solutions (0,3% des holdings dans Portico, 0,9% dans CLOCKSS, 1,2% dans LOCKSS). L'amélioration des données, a permis de compléter les ISSN manquants de 0,29% des holdings dans LOCKSS, de 1,84% des holdings dans Portico et de 3,16% des holdings dans CLOCKSS, et de compléter les e-ISSN manquants de 6,60% des holdings dans LOCKSS, de 3,48% des holdings dans Portico et de 5,32% des holdings dans CLOCKSS.

¹⁶⁸ LOCKSS fournit le titre, l'éditeur et les années archivées, mais pas la plateforme. LOCKSS a pris conscience de ce problème et a promis (contact fin août 2013) d'améliorer la liste publiée sur son site web en intégrant cet élément. CLOCKSS fournit le titre et l'éditeur, mais pas les années ni la plateforme. Portico fournit titre, années et éditeur, mais pas la plateforme non plus. Cet élément est cependant moins crucial que pour LOCKSS et CLOCKSS dont le processus d'archivage est lié techniquement à la plateforme. Il faut souligner aussi que les noms d'éditeurs ne sont malheureusement pas unifiés. Le Keeper's Registry n'intègre pas non plus la donnée de plateforme.

¹⁶⁹ Ceci explique aussi par exemple, que, même si LOCKSS n'archive pas de titres de la plateforme Wiley, une analyse de la présence dans les différentes solutions des holdings ULg par cibles sfx donne 96 titres Wiley dans LOCKSS. Ces titres ont en effet été publiés à un moment donné sur une ou plusieurs autres plateformes (ex : Experimental Physiology, Acta Biochimica et Biophysica Sinica). Il faut tenir compte de ces biais dans l'analyse.

Nous avons comparé les listes de titres d'autres plateformes et le portefeuille d'une institution sur la base harmonisée des différentes solutions, en comparant les listes sur l'ISSN et l'e-ISSN. Les titres de ces listes sans ISSN ou e-ISSN ont donc été exclus d'office de la comparaison. N'ont pas pu être pris en compte non plus les titres dont l'information ISSN ou e-ISSN n'était pas complète et ne correspondait pas aux informations disponibles dans la liste harmonisée¹⁷⁰.

4.2.3.2 Comparaison des couvertures des solutions¹⁷¹

La couverture générale des solutions a été analysée pour répondre aux questions suivantes : 1) comparaison générale (combien de titres sont archivés), 2) quels sont les recouvrements et complémentarités, en nombre absolu, en termes de contenus souscrits ou Open Access, en termes de contenus courants ou d'archives ?

4.2.3.2.1 Importance et typologie des contenus archivés

	Dans LOCKSS	Dans Portico	Dans CLOCKSS
Titres (total : 19.567¹⁷²)	10.104 (avec PCA)¹⁷³	15.681 (dont 1.848¹⁷⁴ sans PCA)¹⁷⁵	7.334 (sans PCA)¹⁷⁶
<i>Titres Open Access (source DOAJ – 9.927 titres avec ISSN)</i>	910 (soit 9,2% du DOAJ)	804 (soit 8,1% du DOAJ)	224 (soit 2,3% du DOAJ)
Editeurs	609 (520¹⁷⁷)	180	30
« Petits » éditeurs (< 5 titres)	505	73	5
« Moyens » éditeurs (entre 5 et 30 titres)	74	51	9
« Grands » éditeurs (> 30 titres)	30	56	16
<i>Presses d'universités (*universit*)</i>	131	35	4
<i>Sociétés et collèges américains (*american*)</i>	88	23	5

¹⁷⁰ Par exemple, si un titre dans le portefeuille d'une institution possède une indication d'ISSN mais pas d'e-ISSN, si ce même titre possède dans la liste harmonisée des solutions une information d'e-ISSN mais pas d'ISSN, il ne peut être pris en compte.

¹⁷¹ Seadle, M. (2011). Archiving in the networked world: by the numbers. Library Hi Tech, 29(1), 189-197. doi:10.1108/07378831111117001 a effectué la même analyse mais sur base d'une autre méthode d'harmonisation des données. Globalement, ses conclusions sont similaires aux résultats présentés ici.

¹⁷² Total des titres distincts archivés au moins dans LOCKSS, Portico ou CLOCKSS. Nous n'avons pas, pour des raisons de temps, pu réaliser la comparaison des données chiffrées sur base des listes disponibles après juin 2013.

¹⁷³ En mars 2014 (données du 21 février 2014), la liste des titres archivés dans LOCKSS comprenait 10.599 titres avec ISSN ou e-ISSN, soit une augmentation de 4,9% depuis juin 2013, soit 25 nouveaux éditeurs (dont IOP, ASM, SIAM).

¹⁷⁴ Soit 40 éditeurs, dont 22 avec moins de 8 titres archivés.

¹⁷⁵ En mars 2014 (données du 21 mars), la liste des titres archivés dans Portico comprenait 17.568 titres avec ISSN ou e-ISSN, soit une augmentation de 12% depuis juin 2013, soit 24 éditeurs supplémentaires (dont Erudit, Vittorio Klostermann).

¹⁷⁶ En mars 2014 (données de décembre 2013), la liste publiée sur le site de CLOCKSS comprenait 9.824 titres (avec ISSN ou e-ISSN), soit une augmentation de 29,5%, soit plus de 70 éditeurs supplémentaires (dont Biomed Central, The Company of Biologists, Palgrave Macmillan, Routledge, SIAM).

¹⁷⁷ 520 éditeurs affichés sur le site web, contre 609 sur base de l'analyse du fichier xls, dans lequel les « grands éditeurs » sont parfois distingués en fonction de filiales, par exemple, pour Springer : Springer Publishing Company, SpringerOpen, Springer Business & Media.

Tableau 9 - Couverture des différentes solutions d'archivage en nombre de titres et d'éditeurs, avec indication de l'accès post-abonnement (PCA).

On peut noter pour commencer que le nombre de titres archivés dans l'une ou l'autre solution semble couvrir une partie substantielle du paysage documentaire puisqu'il représente presque 20.000 titres distincts¹⁷⁸.

Portico offre en nombre de titres une couverture beaucoup plus large que LOCKSS (+55%) et CLOCKSS (+114%), même en ne prenant en compte que les titres avec accès post-abonnement (PCA) (+37% par rapport à LOCKSS)¹⁷⁹.

Par contre, si l'on observe le nombre d'éditeurs présents dans chacune des solutions, c'est LOCKSS qui présente le plus de diversité : 520 éditeurs dans LOCKSS pour 180 éditeurs dans Portico. Ceci confirme ce qu'on lit souvent dans la littérature, à savoir que **LOCKSS travaille beaucoup avec de plus petits éditeurs** et notamment des éditeurs académiques¹⁸⁰. Et si l'on définit, faute de mieux, un petit éditeur comme un éditeur ayant moins de 30 titres archivés (cf. M. Seadle [2011]¹⁸¹), les résultats confirment cette analyse, puisque pour LOCKSS, 30 éditeurs ont archivé plus de 30 titres, alors qu'il y en a 56 dans Portico et 16 dans CLOCKSS. Si l'on mesure l'autre bout de l'échelle, on constate que 505 éditeurs sur les 609 ont archivé moins de 5 titres dans LOCKSS, alors qu'il y en a 73 dans Portico et 5 dans CLOCKSS¹⁸².

Comme le soulignait déjà M. Seadle [2011], le nombre élevé de petits éditeurs présents dans LOCKSS et pas dans Portico peut s'expliquer facilement par le coût puisque LOCKSS est gratuit pour les éditeurs : il s'agit avant tout d'un choix de service.

Par contre, en ce qui concerne les titres Open Access, la comparaison des listes des solutions avec la liste des titres du DOAJ indique à peu près le même rapport pour LOCKSS et Portico : sur les 9.227 titres du DOAJ avec ISSN (soit 2.899 éditeurs), 9,2% sont présents dans LOCKSS et 8,1% dans Portico (contre 2,3% dans CLOCKSS), et il s'agit globalement des mêmes car la combinaison de LOCKSS et Portico n'amène le taux qu'à 10,7%. On retrouve notamment Hindawi (244 titres), BiomedCentral (216 titres), MedKnown Publications (77 titres), Copernicus Publications (24 titres), Springer (20 titres), Co-Action Publishing (13 titres), Oxford University Press (4 titres)¹⁸³.

¹⁷⁸ C'est-à-dire à titre illustratif, un nombre de titres semblable au nombre de titres indexés dans Scopus (source : <http://www.elsevier.com/online-tools/scopus/content-overview>, consulté en décembre 2013).

¹⁷⁹ Il faut toutefois noter que sur les 15.681 titres présents dans Portico, seuls 13.390 titres sont effectivement archivés, les autres étant en cours d'intégration. Dans LOCKSS, seuls 8.399 titres sur les 10.104 titres de la liste sont mentionnés comme effectivement archivés, les autres étant « in progress ».

¹⁸⁰ Cette orientation peut en partie s'expliquer par l'origine du projet, la bibliothèque de Stanford ayant aussi fondé la plateforme Highwire, avec qui LOCKSS travaille en étroite collaboration, et qui est une plateforme de petits éditeurs de sociétés savantes.

¹⁸¹ Tout en étant conscient du biais que cela peut représenter si les éditeurs n'archivent qu'une partie de leurs titres dans l'une ou l'autre solution. Toutefois, comme le montre la répartition du nombre de titres pour les grands éditeurs dans les différentes solutions (cf. point 4.2.1.3.3.3), cela ne semble pas être le cas.

¹⁸² Ces différences, si elles restent notables, sont toutefois à relativiser car la définition d'un « éditeur » ne semble pas toujours la même selon les solutions d'archivage et dépend des contrats signés : BioOne est considéré comme un seul éditeur dans Portico, alors que les mêmes titres dans LOCKSS sont listés sous le nom de l'éditeur effectif (soit 91 éditeurs dans LOCKSS pour 1 dans Portico). BioMed Central est repris comme éditeur à la fois dans LOCKSS et Portico, mais pas dans CLOCKSS (où l'éditeur pour les mêmes titres est Springer). Certains des titres repris sous BioMed Central dans LOCKSS et Portico sont également repris dans Springer (ou une filiale associée) dans LOCKSS et Portico, mais pas tous et pas les mêmes – ce qui illustre la complexité des comparaisons basées sur les holdings.

¹⁸³ Éditeurs ayant plus de 10 titres dans le DOAJ et présents dans les deux solutions.

4.2.3.2.2 Recouvrement et complémentarité

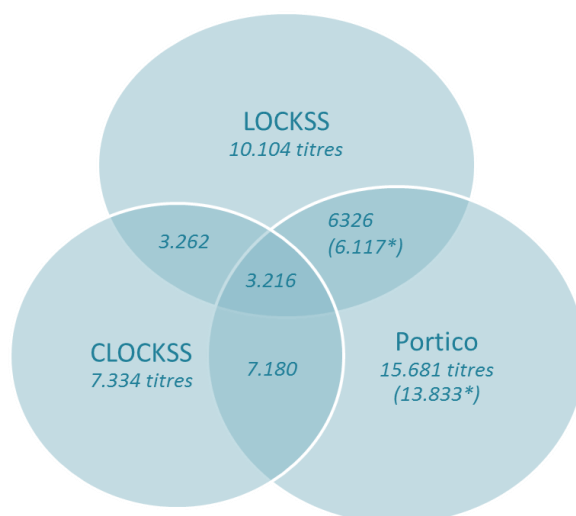


Figure 8 - Recouvrement des solutions en nombre de titres. *[titres avec accès post-abonnement]

Le **recouvrement** entre les solutions n'est que partiel, sauf en ce qui concerne CLOCKSS et Portico, ce qui s'explique relativement bien par les politiques des éditeurs en fonction des garanties proposées par les solutions (cf. point 4.2.3.2.4)

Le recouvrement entre LOCKSS et Portico compte 6.326 titres : 62% des titres de LOCKSS sont présents dans Portico, et presque 40% des titres de Portico se trouvent dans LOCKSS (45% si l'on excepte les titres sans accès post-abonnement) ; chacune des deux solutions présente 35% de titres uniques, non couverts par les deux autres solutions.

Un peu plus de 35% des titres de LOCKSS sont communs à CLOCKSS, et 45% (ou 50%, avec PCA) des titres des Portico se trouvent dans CLOCKSS. Par contre, tous les titres archivés dans CLOCKSS (98%) le sont dans les autres solutions également (43% dans LOCKSS, 97% dans Portico).

Editeur / Nb titres	Portico (*avec PCA)	LOCKSS	CLOCKSS
Elsevier	2737*		1970
Taylor & Francis ¹⁸⁴	2167*	2295	1387
John Wiley & Sons, Inc. ¹⁸⁵	1858*	1**	1388
Springer ¹⁸⁶	1590*	2497 + 122 + 21	1033
IEEE, Inc.	875		
SAGE Publications	703*	763	526
Wolters Kluwer Health	398*		***
Inderscience		397	***
Hindawi Publishing Corporation	351*	296	
Oxford University Press	312*	321	221
Emerald Group Publishing	292*	475	
BioMed Central	269*	281	***
Brill	250		***

¹⁸⁴ Dénomination alternative: Taylor & Francis Group (Portico).

¹⁸⁵ Dénomination alternative: Wiley Blackwell (LOCKSS)

¹⁸⁶ Dénomination alternative: Springer Science+Business Media, Springer Open, Springer Publishing Company (LOCKSS)

Informa Healthcare	218*		
Haworth Press	197*		
Thieme Publishing Group	177*		137
Bentham Open	176*		
BioOne	172*		
Maney Publishing	165*	130	106
Medknow Publications		157	***
IGI Global		152	***
De Gruyter	143*		***
Medknow Publications and Media	132*		***
Karger	123*		
Cambridge University Press	119	3**	
Dove Medical Press		119	***
MDPI		118	
Philosophy Documentation Center	117*		***
World Scientific	117		
Association for Computing Machinery	116*		109 ¹⁸⁷
Fabrizio Serra Editore	115*		115
Franco Angeli Edizioni	102*		

Tableau 10 - Liste des éditeurs ayant plus de 100 titres présents dans au moins une solution.

** [titres transférés].

*** Ces éditeurs participent à CLOCKSS¹⁸⁸ mais aucun de leurs titres n'est listé dans les titres préservés¹⁸⁹, à l'exception de Biomed Central, Inderscience dont des titres ont été inclus entre novembre 2012 et décembre 2013.

La présence de quelques **grands éditeurs** est déterminante pour la couverture de toutes les solutions : les titres des 4 éditeurs les plus importants, Springer, Elsevier, Wiley et Taylor & Francis (qui publient chacun plus de 1.500 titres) comptent pour 53,3% des titres dans Portico (60,4% des titres dans Portico avec accès post-abonnement), 78,8% dans CLOCKSS et 48,9% dans LOCKSS (Elsevier et Wiley n'y étant pas présents).

Une part de recouvrement vient de la participation de quelques grands éditeurs à plusieurs solutions : Springer (environ 1.000 titres), Taylor et Francis (environ 1.170 titres), SAGE (environ 500 titres), OUP (200 titres) rendent compte d'environ 90% du recouvrement entre les trois solutions.

A l'inverse, les titres de l'IEEE (875), certains titres d'Elsevier (près de 800)¹⁹⁰, certains titres de Wiley (près de 500)¹⁹¹ et les titres de Wolters Kluwer (386), rendent compte de la moitié des titres uniques

¹⁸⁷ Association of Computing Machinery dans CLOCKSS.

¹⁸⁸ Source : http://clockss.org/clockss/Participating_Publishers (consulté le 18 février 2012).

¹⁸⁹ Ni dans la liste de novembre 2012 qui a servi de base à la comparaison, ni dans la liste de décembre 2013, la dernière disponible sur le site de l'archive en février 2014.

¹⁹⁰ Les titres Elsevier uniquement préservés dans Portico correspondent à des titres avec des états de collection non courants (soit qu'ils aient changé de nom, qu'ils aient cessé d'être publié, qu'ils aient été transférés...). Sur les 767 titres identifiés, seuls 81 sont indiqués comme actifs. Sur ces 81 titres, 40 sont inclus dans la Freedom Collection ou dans les collections thématiques sur ScienceDirect, et sur ces 40 titres, 38 ont été inclus dans ces collections après 2010 (et 19 après 2011). L'absence de ces titres dans CLOCKSS semble donc liée au moment du moissonnage et/ou du transfert des titres.

¹⁹¹ L'explication est probablement similaire, car aucun de ces titres ne fait partie des collections actives sur la Wiley Online Library en 2014.

dans Portico. Ceux de Brill (238), de Gruyter (143), Bentham Open (176), Cambridge University Press (119), Karger (123) pour presque l'autre moitié¹⁹².

Les titres de Wiley (1.858) et d'Elsevier (2.737), présents dans Portico mais pas dans LOCKSS expliquent une bonne partie de la différence en termes de quantité de contenus couverts entre les deux solutions. Quelques grands éditeurs européens présents dans Portico sont aussi absents de LOCKSS (Brepols, de Gruyter, Brill, Cambridge University Press).

Dans LOCKSS, les **presses d'universités (la majorité américaines¹⁹³), et les sociétés ou colleges¹⁹⁴**, représentent, dans l'ensemble, près de la moitié des éditeurs de titres uniques (176 sur 394). Ces types d'éditeurs comptent aussi pour la moitié des éditeurs de titres uniques dans Portico (50 sur 93). Il y a aussi, dans LOCKSS, beaucoup plus d'éditeurs uniques que dans les autres solutions : **Portico travaille avec environ 90 éditeurs uniques, LOCKSS avec près de 400¹⁹⁵**.

Ces conclusions s'articulent assez facilement avec les éléments concernant la philosophie et la gouvernance des solutions, tout comme avec leur philosophie d'accès. CLOCKSS est un partenariat entre acteurs majeurs des bibliothèques et du monde de l'édition à des fins de préservation contrôlée, Portico a été fondé à l'origine en collaboration avec des éditeurs majeurs, alors que LOCKSS a défini dès le début une orientation marquée vers les bibliothèques et une volonté de préserver les collections les plus en danger, et a établi des partenariats en fonction, notamment via la plateforme Highwire¹⁹⁶ et le projet OpenLOCKSS¹⁹⁷.

M. Seadle concluait en 2011 que les éditeurs préfèrent généralement travailler avec plusieurs solutions pour répartir les risques ; CLOCKSS utilisant le même logiciel que LOCKSS, le choix des éditeurs de l'une ou l'autre solution dépend de ce qu'ils veulent fournir un accès post-abonnement ou pas.

Ces éléments mettent en évidence la **complémentarité des solutions, en particulier de LOCKSS et CLOCKSS**. Cette complémentarité est très visible via la couverture des solutions, comme l'illustre le tableau ci-après. La combinaison des titres présents dans LOCKSS et CLOCKSS, en particulier, donne une couverture de 14.176 titres, ce qui s'approche de la couverture de Portico (15.681 titres).

¹⁹² Or, Brill et de Gruyter ont décidé de participer à CLOCKSS, même si aucun titre de ces éditeurs n'est encore préservé dans l'archive.

¹⁹³ Une requête sur le nom d'éditeur « *universit* » (incluant les presses d'universités étrangères, donne 152 résultats pour l'ensemble des solutions, dont 131 dans LOCKSS et 35 dans Portico et 4 dans CLOCKSS.

¹⁹⁴ Une requête sur le nom d'éditeur « *american* » donne 104 résultats pour l'ensemble des solutions, dont 88 dans LOCKSS contre 23 dans Portico, et 5 dans CLOCKSS. Une requête sur « *universit* or *american* » donne 255 éditeurs au total, dont 218 dans LOCKSS contre 58 dans Portico.

¹⁹⁵ On a effectué pour chaque solution l'identification des éditeurs de titres uniques puis exclu les éditeurs communs à ceux de la liste des titres non uniques.

¹⁹⁶ La bibliothèque de Stanford a fondé la plateforme Highwire, avec qui LOCKSS travaille en étroite collaboration, et qui est une plateforme de petits éditeurs de sociétés savantes. LOCKSS intègre tous les titres de Highwire Press – ce qu'a confirmé notre analyse, qui a trouvé 1.286 titres Highwire intégrés dans LOCKSS sur 1.392, uniquement sur l'ISSN commun. La même requête sur les titres de Portico donne 1.052 titres, et dans CLOCKSS, on a 786 titres.

¹⁹⁷ Ce projet financé par le JISC et mené par l'Université de Glasgow en 2007, avait pour objectif de négocier avec un certain nombre d'éditeurs Open Access du Royaume-Uni l'inclusion de leurs titres dans LOCKSS (<http://www.lib.gla.ac.uk/Research/openlockss/>). 28 éditeurs ont été contactés, ce qui représente 32 titres (Source : UKLA Pilot report, 2008). Le rapport LOCKSS UKLA Pilot mentionne aussi le projet LOCKSS Humanities, au cours duquel les universités américaines participantes ont identifié et ajouté à LOCKSS un certain nombre de titres Open Access prioritaires en Sciences humaines.

Couverture	Portico + LOCKSS + CLOCKSS	LOCKSS + CLOCKSS	Portico + CLOCKSS	Portico + LOCKSS
Titres avec accès post-abonnement via la solution	17.820	10.104	7.095	17.820
Titres sans accès post-abonnement via la solution	1.747	4.072	8.740	1.639
Total	19.567	14.176	15.835	19.459

Tableau 11 - Complémentarité des solutions LOCKSS, CLOCKSS, et Portico en termes de couverture

4.2.3.2.3 Couverture présente et rétrospective, archives

Pour préciser l'évaluation de la couverture des solutions, il est utile d'examiner leur couverture dans le temps et de distinguer : la couverture courante (2013 - présent), la couverture des titres natifs électroniques (1995/1997 - présent) et la couverture des archives numérisées (*backfiles*). La problématique est en effet différente selon les cas :

- la couverture courante permet d'évaluer la réactivité des solutions, ou en d'autres termes, la rapidité d'ingestion des contenus, voire d'identifier le retrait éventuel d'éditeurs si le décalage entre l'année courante et la dernière année archivée est important ;
- la couverture des titres natifs électroniques est une priorité en termes de garanties d'accès aux collections, en particulier dans le cas des collections e-only ; en général, la publication électronique dans les grands bouquets d'éditeurs a lieu vers 1995-1997, cependant, nous avons utilisé l'année 2000 comme limite pour faciliter l'analyse ;
- la couverture des archives, parfois accessibles dans le cadre de contrats courants (ex. PROLA d'APS), mais dont l'acquisition pérenne auprès des éditeurs fait toujours l'objet d'un contrat séparé, est un élément important, soit pour évaluer la couverture des bouquets d'archives déjà achetés pour les institutions (c'est le cas des Archives Springer – OJA via la BICfB, et des archives Elsevier hors BICfB) soit comme élément d'analyse coût-bénéfice de l'achat d'archives électroniques par rapport à la gestion de collections d'archives imprimées.

4.2.3.2.3.1 Couverture présente et rétrospective

Le **début d'archivage** pour chaque titre n'est pas toujours explicite dans les différentes solutions. Le début d'archivage d'un titre dans Portico correspond au minimum à l'année du début du contrat entre l'éditeur et Portico, mais l'objectif explicite est de couvrir la totalité des volumes de la publication. Dans LOCKSS, l'année de début d'archivage correspond à l'accès autorisé par l'éditeur pour le moissonnage. CLOCKSS vise à archiver la totalité de la collection pour chaque titre, mais l'état de collection effectivement archivé n'est pas précisé¹⁹⁸ dans la liste des titres (les analyses d'état de collections n'y ont donc pas été effectuées).

	Total ¹⁹⁹	Année courante (2013 ou "present")	Titres récents (2000-)	Archives (<2000)	Titres avec vol. 1
Titres					
LOCKSS	10.104	4.595	7.146	3.293	4.625
Portico	15.681	2.912	12.242	6.259	3.822

¹⁹⁸ C'était le cas dans la liste utilisée pour l'analyse ; la liste de titres disponible en février 2014 sur le portail clockss.org reprend bien les indications des états de collection, mais nous n'avons pu, pour des raisons de temps, intégrer ces données dans l'analyse.

¹⁹⁹ La comparaison n'est pas exhaustive puisque dans LOCKSS, 2.835 titres n'ont pas d'indication d'année archivée et 3.106 n'ont pas d'indication de volumes archivés. Dans Portico, 3.011 titres n'ont ni indication d'année, ni indication de volume.

Editeurs					
LOCKSS	602	337	565	37	233
Portico	180	47	117	86	89

Tableau 12 - Comparaison de données de couverture courante et rétrospective dans LOCKSS et CLOCKSS

Des deux solutions, c'est LOCKSS qui présente la plus grande part de couverture « active », puisque l'année courante est archivée pour près de la moitié des titres, alors que dans Portico, ce n'est le cas que pour un peu moins de 20% des titres.

Concernant la question de la **couverture récente (à partir de 2000)**, LOCKSS archive 71% des titres de son portefeuille (pour près de 94% des éditeurs couverts) avec une couverture postérieure à l'année 2000, tandis que dans Portico, si une couverture postérieure à 2000 est indiquée pour 78% des titres, ceux-ci ne relèvent que de 65% des éditeurs de son portefeuille.

Pour évaluer la **couverture rétrospective plus ancienne**, on a examiné le nombre de titres indiquant des années de collections antérieures à 2000, et dans quelle mesure les solutions archivaient le volume 1 des publications. Ceci n'indique pas évidemment si le titre existe depuis quelques années ou s'il a une grande histoire de publication en format papier, et, dans ce dernier cas, tous les titres n'ont peut-être pas encore été numérisés. Ces éléments montrent ici encore que la couverture de LOCKSS concerne des publications plus récentes ou à l'histoire moins longue que celle de Portico, puisque davantage de titres remontent jusqu'au volume 1 dans LOCKSS, alors que seulement 32% des titres ont des états de collection indiqués antérieurs à 2000 – ce qui peut s'expliquer assez logiquement par la typologie des éditeurs archivés dans LOCKSS. Ceci montre aussi que la rétro-ingestion prévue dans Portico semble loin d'être achevée actuellement, et c'est aussi le cas dans LOCKSS, dans une moindre mesure.

L'extension de la couverture des solutions à la totalité des années de publication d'un titre reste donc une question à suivre dans tous les cas.

4.2.3.2.3.2 Bouquets d'archives

L'**achat d'archives** électroniques à des fins d'extension de l'accès rétrospectif ne peut se justifier que si les garanties d'archivage pérenne de ces titres sont suffisantes.

La comparaison exacte d'états de collections dans les différentes solutions impliquant des séries de traitements de chaînes de caractères que nous n'avons pas souhaité effectuer dans le cadre de cette étude, nous avons, à titre indicatif, comparé le contenu de trois grands bouquets d'archives avec celui des solutions LOCKSS, CLOCKSS et Portico, en fonction des données dont nous disposons, en approchant les états de collection par paliers de 10 ans sur base de requêtes textuelles²⁰⁰.

SPRINGER	Springer OJA (-1996) [source : liste éditeur]	Dans LOCKSS	Dans CLOCKSS	Dans Portico
Nb de titres total	1015	1009	582	825
<i>Nb de titres sans année indiquée</i>	1	865	(toutes)	0
Collections <1990	770	103	(inconnu)	280
Collections <1980	514	71	(inconnu)	151
ELSEVIER	Elsevier Backfiles (-1994) [source : liste éditeur]	Dans LOCKSS	Dans CLOCKSS	Dans Portico
Nb de titres total	565	(8) ²⁰¹	405	560

²⁰⁰ Par exemple, pour les états de collection antérieurs à 1980, nous avons recherché les années comprenant les chaînes de caractère 18* ou 191* ou 192* ou 193* ou 194* ou 195* ou 196* ou 197*.

²⁰¹ Le chiffre indiqué ici reflète des titres transférés à un moment ou à un autre de ou vers Elsevier, Elsevier ne travaillant pas avec LOCKSS.

Nb de titres sans année indiquée	0	-	(toutes)	0
Collections <1990	461	-	(inconnu)	366
Collections <1980	259	-	(inconnu)	203
SAGE	SAGE Backfiles (-1998) [source : liste éditeur]	Dans LOCKSS	Dans CLOCKSS	Dans Portico
Nb de titres total	463	456	407	440
Nb de titres sans année indiquée	5	5	(toutes)	0
Collections <1990	332	12	(inconnu)	25
Collections <1980	210	10	(inconnu)	17

Tableau 13 - Couverture rétrospective des bouquets d'archives Springer, Elsevier et SAGE dans LOCKSS, CLOCKSS et Portico

Il semble donc que les états de collection présents dans les solutions sont loin de couvrir la totalité de la collection, et que, à l'heure actuelle, ni LOCKSS ni Portico, ni CLOCKSS selon toute probabilité, ne permettent d'archiver les grands bouquets d'archives achetés en tant que collections complètes.

4.2.3.2.4 Dynamisme et capacité d'extension des solutions

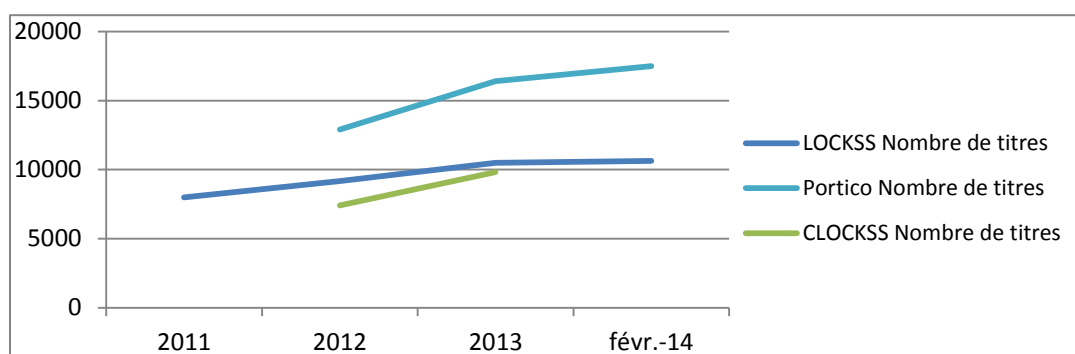


Figure 9 - Evolution du nombre de titres d'e-journaux couverts par les solutions

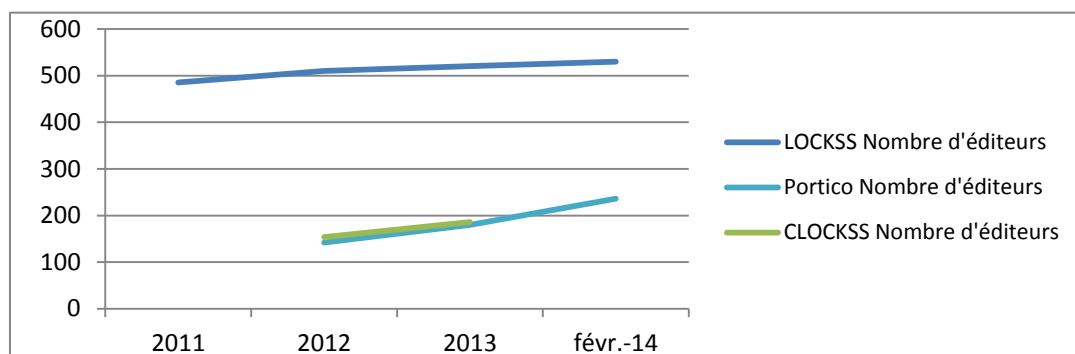


Figure 10 - Evolution du nombre d'éditeurs présents dans les différentes solutions

Remarque préliminaire : il n'est pas facile d'obtenir des données complètes sur la croissance des solutions, et les données ont été glanées ponctuellement sur les sites des solutions et dans la littérature²⁰². Par ailleurs, comme déjà mentionné, les données ne recouvrent pas toujours la même réalité ; l'image est donc partielle et indicative seulement.

²⁰² Les données historiques concernant LOCKSS ne sont disponibles que via les nouvelles publiées sur le site et les billets du blog de du Dr. Rosenthal, par ailleurs, LOCKSS ne publie plus la liste des bibliothèques participantes ; Portico publie une page sur la croissance de l'archive sur son site, mais en indiquant seulement le nombre de titres et de bibliothèques participantes, par le nombre d'éditeurs.

Les trois solutions sont visiblement encore en phase de croissance, non seulement en termes de quantité de contenu archivé, mais également en termes de types de contenus archivés, puisque Portico et LOCKSS, et CLOCKSS ont intégré récemment des e-books (en 2009 pour Portico, en 2013 pour LOCKSS).

Nous avons choisi comme donnée la plus indicative sur la croissance des solutions en termes de couverture le **nombre d'éditeurs participants**, puisqu'il s'agit de l'étape la plus difficile en termes de négociations, et qu'il n'y a pas de difficulté particulière autre que le temps d'ingestion pour étendre la couverture à l'ensemble du portefeuille de l'éditeur. De ce point de vue, la croissance des solutions semble relativement équivalente (cf. fig. 10). Par contre, la croissance de Portico est beaucoup plus rapide en nombre de titres (cf. fig. 11), ce qui peut toutefois s'expliquer par la présence de davantage de grands éditeurs dans Portico que dans LOCKSS. La vitesse de croissance de CLOCKSS en nombre de titres et d'éditeurs est similaire à celle de Portico, même si CLOCKSS comprend moins de titres.

Il est utile de s'interroger sur les **facteurs qui limitent** ou limiteront cette croissance, du point de vue de la participation des éditeurs, dans les trois solutions.

Dans le cas de **CLOCKSS** comme dans le cas de **Portico**, qui offrent des garanties solides aux éditeurs en termes de contrôle de l'archive²⁰³, mais exigent en contrepartie une contribution financière annuelle, il pourrait y avoir une limite financière à la participation des éditeurs. Toutefois, étant donné les montants concernés (même dans le cas des grands éditeurs, dans Portico, où la contribution est la plus élevée, elle n'atteint pas plus de 0,05% du revenu), c'est assez peu probable. Le succès de Portico auprès des grands éditeurs ne se dément d'ailleurs pas depuis plusieurs années, ni, plus récemment, celui de CLOCKSS. On peut toutefois, éventuellement, imaginer un ralentissement possible lié au processus de planification et d'ingestion des contenus dans les deux cas, en particulier dans le cas de Portico qui implique des opérations de transfert plus lourdes de la part de l'éditeur. Dans le cas de Portico, la nécessité de fourniture des fichiers par l'éditeur peut aussi jouer en défaveur de la participation des petits éditeurs, ceux-ci ne disposant pas toujours des capacités techniques nécessaires en interne.

Annual Journals Revenue	Annual Archive Contribution	
	Portico	CLOCKSS
>\$200 millions	\$81.960	\$25.000
\$100-200 millions	\$54.635	\$15.000
\$50-100 millions	\$27.320	\$15.000
\$10-50 millions	\$16.440	\$6.000 - \$14.000
\$5-10 millions	\$5.460	\$4.000
\$1-5 millions	\$2.735	\$2.000
\$500k-1 million	\$1.085	\$1.000
\$250-500k	\$500	\$400
<\$250k	\$250	\$200

²⁰³ Pour Portico : ouverture de l'accès aux bibliothèques après accord de l'éditeur, et, sauf pour l'accès post-abonnement, uniquement si le contenu n'est pas disponible ailleurs, droits d'usage limités, restitution des contenus en cas de faillite de la solution ; pour CLOCKSS : ouverture de l'accès seulement s'il n'est plus disponible sur le site de l'éditeur, après décision du « board ».

Tableau 14 - Contributions officielles des éditeurs pour Portico et CLOCKSS (février 2014)

Dans le cas de **LOCKSS**, le facteur économique ne joue pas puisque la solution est gratuite pour les éditeurs, mais il existe des facteurs limitatifs d'ordre commercial et d'ordre technique. Plusieurs grands éditeurs, approchés en 2007 par NESLI2 dans le cadre du projet pilote de la UK LOCKSS Alliance ont refusé de participer : Blackwell, Elsevier, Springer, Wiley, Science (AAAS), American Institute of Physics, et IEEE. A l'exception de Springer, aucun de ces éditeurs, 5 ans après, ne participe à la solution. L'explication majeure semble avoir été leur préférence pour d'autres solutions, telles Portico et CLOCKSS, qui offrent davantage de contrôle sur les données. Un des éléments d'explication pourrait être que le modèle de LOCKSS n'est pas techniquement adapté au modèle commercial des *big deals* : pour le moissonnage, la distinction entre l'accès courant (collection *big deal*) et l'autorisation d'accès post-abonnement (*core collection*) n'est pas effectuée via la *manifest page* et la reconnaissance IP, et elle semble dépendre uniquement de la gestion des accès autorisés au sein de l'institution²⁰⁴. **L'extension possible de LOCKSS pour couvrir les grands éditeurs semble, au vu de ceci, assez limitée** ; aucun frein majeur n'a par contre été relevé pour les plus petits éditeurs. Des limites existent par contre du côté des bibliothèques : le temps de négociation, ainsi que la nécessité d'un nombre suffisant de bibliothèques souhaitant s'investir pour la préservation de titres définis. Si la collaboration est un avantage car elle permet de mutualiser les ressources, elle montre ici ses limites.

4.2.3.3 Comparaison avec un portefeuille de titres BICfB

4.2.3.3.1 Objectifs et méthodologie

La comparaison du portefeuille total des accès électroniques d'une institution avec la couverture des solutions, au-delà du fait qu'elle est extrêmement décevante, n'a, en soi, pas beaucoup de sens : les solutions ne sont pas universelles (et il est sans doute illusoire de penser qu'elles le seront un jour), et le portefeuille des accès courants d'une institution ne correspond pas nécessairement à sa politique d'archivage. Il est surtout utile d'articuler la comparaison afin de mieux pouvoir identifier les points d'actions, les priorités, et les combinaisons utiles de solutions. La remarque de l'étude du JISC (Morrow, Beagrie, & Jones, 2008) garde en effet toute son actualité : "Finding an arrangement which is realistic, reasonable and equitable is by no means straightforward and much effort has been expended in recent years in developing and testing a range of options" (p. 10). Dans cette optique, nous proposons un essai de typologie du portefeuille afin de préparer la priorisation possible des choix en termes d'accès pérenne.

Périodiques électroniques	UNamur	USL-B	UCL	ULB	ULg	UMONS
Nb titres	18.563	19.098	37.236	23.642	53.881	6.810
Current serials electronic-titles	Non communiqués	16.836	24.827	21.914	Non communiqués	994
Current serials electronic - subscriptions	8.043	7.214	19.626	15.854	15.267	6.725

Tableau 15 – Nombres de titres et souscriptions du portefeuille électronique des institutions de la BICfB (source : Statistiques CIUF, 2011)

Le portefeuille de l'ULg a été choisi comme échantillon pour des raisons de commodité et parce qu'il semblait suffisamment représentatif²⁰⁵, les autres portefeuilles pourraient être analysés ensuite

²⁰⁴ InformaWorld a signalé également lors de la phase test de la UK LOCKSS Alliance, qu'il était difficile d'implémenter des *manifest pages* qui reflètent la nature mouvante des souscriptions courantes et achevées, mais la question a été réglée.

²⁰⁵ L'ULg est l'une des trois institutions complètes de la BICfB – son portefeuille est assez représentatif, et les fichiers sources étaient facilement accessibles par l'auteur. Notons que le portefeuille de l'ULg présente la

selon la même méthode. Le fichier d'export des accès complets aux journaux électroniques de l'ULg a été exporté depuis sfx, et les cibles sfx (plateformes / éditeurs) traitées pour exclure le bruit²⁰⁶ et étiquetées semi-automatiquement pour affiner la catégorisation (selon la typologie définie ci-dessous). Ce fichier a ensuite été comparé avec le fichier consolidé des titres présents dans les différentes solutions. Dans tous les cas ont été exclus de l'analyse, pour des raisons pragmatiques, les titres sans ISSN ni e-ISSN.

Les résultats détaillés pour chaque cible sfx sont disponibles en Annexe 3.

4.2.3.3.2 Typologie du portefeuille

Le portefeuille examiné comprend, en tant que tel, une grande variété d'accès de types différents aux ressources électroniques :

- Accès payants, via différentes modalités déjà présentées plus haut dans l'examen des licences : souscriptions éditeurs ou agent aux titres courants, souscriptions à des bases de données full-text ou bouquets d'agrégateurs, achat ou souscription d'archives²⁰⁷.
- Accès gratuits, soit via des plateformes Open Access, soit via des plateformes d'éditeurs ou agrégateurs payants (publiant certains titres en Open Access), soit via des plateformes de type « bibliothèque numérique » (Gallica, ...).

Il faut ajouter la complexité des états de collection des titres : titre courant ou accès terminé, présence éventuelle d'embargo. Ces éléments balisent les notions de constitution de collection et de fourniture d'accès. La présence d'embargos est souvent signe d'une orientation vers un modèle d'accès au détriment d'un modèle d'achat (bases de données full-text), ou liée au statut Open Access des titres (l'achat étant alors disponible ailleurs).

Au sein de cette multiplicité, nous avons concentré notre analyse sur le portefeuille des accès payants. Ces titres représentent en effet le retour sur investissement des institutions, et leur accès pérenne est lié aux clauses contractuelles des souscriptions. Parmi ces accès, nous avons distingué, autant qu'il était possible²⁰⁸, les accès via l'éditeur des accès via des bases de données full-text, qui constituent des problématiques distinctes en termes d'accès pérenne, de même que l'accès aux bouquets d'archives, traités eux aussi séparément.

Au sein des accès gratuits, il est difficile de faire la part des choses entre les ressources numérisées, souvent au sein de « bibliothèques numériques gratuites » (Gallica, DigiZeitschriften,...) et les titres de périodiques natifs numériques publiés en Open Access, ou accessibles gratuitement. Si les seconds ont bien pour objet d'être archivés par les solutions telles LOCKSS, CLOCKSS ou Portico, ce n'est pas toujours le cas des premières, en tous cas pas selon les mêmes modalités²⁰⁹. Ces ressources

particularité, au sein de la BICfB, d'avoir sélectionné et activé l'accès à un grand nombre de titres Open Access via le résolveur de liens.

²⁰⁶ Les cibles sfx contiennent en très grande majorité des e-journaux, mais également quelques book series, conference proceedings, etc.

²⁰⁷ Les cibles concernant des « ouvrages de références » et des statistiques (Brill Reference Work et OCDE iLibrary) ont été exclues de l'analyse car elles ne concernent pas des périodiques électroniques à proprement parler.

²⁰⁸ La distinction a été établie sur base d'un étiquetage manuel des cibles sfx, dont la qualité n'est certainement pas parfaite.

²⁰⁹ Portico propose la souscription au D-Collections, qui visent spécifiquement ces collections et contenus numérisés, mais payants (partenariats actuellement avec Adam Matthew Digital (<http://www.amdigital.co.uk/>), une division de SAGE et avec Gale (Cengage Learning) ; plusieurs réseaux privés LOCKSS concernent également ce type de contenus, mis à disposition librement ou non, notamment MetaArchive et COPPUL).

numérisées dans le cadre de « bibliothèques numériques » sortent d'ailleurs du champ spécifique de cette étude.

Portefeuille ULg	Total	Courants	Courants sans embargo	Note
Nombre total d'accès électroniques	62.622	47.522	42.227	
Accès payants e-journaux	23.178	16.578	12.946	
> titres d'éditeurs (listes des cibles par nombre)	8.138	6.890	6.858	
> titres d'agrégateurs	12.738	10.966	7.387	2.434 titres sont des doublons avec les accès éditeur, dont 701 sans embargo
> bouquets d'archives (d'éditeur ou agrégateur) ²¹⁰	2.302	-	-	
Accès gratuits (y compris « bibliothèques numériques »)	38.578	30.565	28.902	
> titres Open Access du DOAJ	9.888	9.888	9.887	

Tableau 16 - Typologie des accès électroniques : le portefeuille de l'ULg (juillet 2013) – seuls sont comptés les titres possédant un ISSN ou un e-ISSN.

4.2.3.3.3 E-journaux payants²¹¹

Sur un total de 23.178 *holdings* ULg payants avec ISSN ou e-ISSN, 60% (soit 16.960 titres), correspondent à des titres présents dans au moins une solution tierce d'archivage : 32% sont présents dans LOCKSS, 36% dans CLOCKSS et 54% dans Portico. La couverture varie fort, logiquement, en fonction des catégories : les titres des plateformes primaires (contrats éditeurs ou agrégateur primaire) sont couverts par les solutions à hauteur de 83,7%, tandis que les titres des bases de données full-text ne sont couverts qu'à hauteur de 41,6%.

La couverture des bouquets et plateformes d'archives présents dans le portefeuille ULg est peu représentative : non seulement ils sont peu nombreux (Springer, JSTOR, pour les archives souscrites, PROLA, Annual Reviews, Institute of Physics Historical Archives pour les archives accessibles mais non souscrites), mais leur couverture dans les solutions n'est pas représentative puisque ce qui importe est précisément la couverture rétrospective, qui n'est pas reflétée ici.

La couverture des titres d'agrégateurs et de bases de données full-text n'est malheureusement pas non plus très significative étant donné que les solutions – et les agrégateurs – doivent obtenir les autorisations auprès des éditeurs, que l'ingestion des titres archivés se fait via les plateformes primaires de publication (LOCKSS, CLOCKSS) ou directement par transfert de fichiers auprès des éditeurs (Portico, CLOCKSS). Nous avons vu aussi (cf. supra 2.2.1.2) que les agrégateurs ne proposent pas, à l'heure actuelle, de garantie suffisante d'accès post-abonnement.

²¹⁰ Les collections JSTOR sont incluses dans les archives, et comprennent 51 titres courants, dont 30 sans embargo.

²¹¹ La source des e-journaux ULg utilisée a été l'export .csv de tous les titres activés dans sfx. En effet, c'est la seule source d'information comprenant tous les titres accessibles à l'ULg et présentant des champs normalisés et des records distincts (le cadastre Ebsco ne contient que les titres payants et, en outre, regroupés par bouquets en un record dans certains cas). Cependant, dans sfx, la seule manière de distinguer les titres payants des autres est de se baser sur le nom de la cible.

	Total Portefeuille	Dans LOCKSS	Dans CLOCKSS	Dans Portico	Dans Portico + PCA	Dans au moins une solution
<i>Accès payants</i>						
Accès	23.178	32,0%	36,0%	54,3%	51,0%	60,2%
Titres	16.960	26,1%	32,8%	50,0%	46,9%	55,8%
<i>Accès payants : Titres de plateformes primaires éditeur ou agrégateur²¹²</i>						
Accès	8.138	34,1%	58,4%	77,5%	74,6%	83,7%
Titres	7.921	33,9%	58,2%	78,2%	75,0%	84,7%
<i>Accès payants : Titres d'agrégateurs et base de données full-text</i>						
Accès	12.738	24,0%	20,6%	37,1%	33,5%	41,6%
Titres	10.241	22,4%	20,1%	36,3%	32,9%	40,8%
<i>Accès payants : Titres de bouquets d'archives</i>						
Accès	2.302	69,0%	42,1%	67,7%	64,7%	79,8%
Titres	1.912	65,0%	38,9%	65,1%	61,5%	78,1%

Tableau 17 - Synthèse de la couverture des solutions par catégorie de cible sfx

Les *holdings* éditeurs de plateformes primaires de publication (un peu plus de 8.000 accès) sont couverts à hauteur de 83,7% par au moins une solution d'archivage, ce qui est loin au-dessus des 30% de l'ensemble du portefeuille, et même des 60% des e-journaux payants. On remarque a priori une grande différence entre les solutions d'archivage : LOCKSS ne couvre que 34% des titres, alors que CLOCKSS couvre 58% et Portico 75% des titres – ce qui correspond à la comparaison de la couverture globale des solutions.

La prise en compte des éléments contractuels (licences éditeurs) fait apparaître une réalité plus relative.

D'abord, dans certains cas, comme nous l'avons vu, l'éditeur accorde un accès post-abonnement – et travaille avec des solutions tierces d'archivage, mais ceci ne concerne, pour l'institution, qu'une partie des titres auxquels elle souscrit auprès de cet éditeur. C'est en particulier le cas des contrats *big deals* Wiley et Elsevier, pour lesquels les titres hors *core collection* présents dans Portico avec accès post-abonnement ne peuvent bénéficier de cet accès (1.947 titres du bouquet pour Elsevier – dont 1.778 présents dans Portico, et 950 pour Wiley, dont 945 présents dans Portico). Les taux de couverture des solutions sans inclure ces titres sont plus proches : sur 5.398 accès, 50% sont couverts par LOCKSS, 49% par CLOCKSS et 71% par Portico²¹³ (cf. tableau 18).

Ensuite, si l'accès post-abonnement est autorisé sur la plateforme de l'éditeur (gratuitement ou moyennant un coût négligeable), les solutions tierces peuvent n'intervenir que pour l'archivage à long terme, pour pallier à une déficience éventuelle de l'éditeur (faillite économique, crash de serveur...). Si les contenus archivés ne sont pas disponibles localement (comme avec LOCKSS), la garantie d'accès post-abonnement contractuelle de Portico ne semble pas présenter davantage de sûreté que la garantie contractuelle de la licence éditeur, puisque Portico demande l'autorisation à l'éditeur avant d'ouvrir l'accès (cf. supra 4.2.2). Dans cette optique, CLOCKSS entre en ligne de

²¹² Ceci concerne les souscriptions à des titres via les éditeurs ou agents, que ces titres soient publiés sur une plateforme propriétaire de l'éditeur ou plateforme d'agrégation (comme Highwire par exemple). Le statut de bouquets spécifiques agrégateurs comme CAIRN ou JSTOR est difficile à catégoriser, mais nous avons fait le choix de le considérer avec les bases de données full-text (Ebsco, ProQuest...).

²¹³ D'autres éditeurs excluent l'accès post-abonnement, et la couverture réelle des solutions est donc un peu plus élevée qu'indiquée ici, mais le nombre de titres concernés est faible et les pourcentages de couverture changent donc peu, et l'information contractuelle est parfois moins claire (en particulier si elle vient d'Ebsconet). Ils n'ont donc pas été pris en compte ici.

compte au même titre que Portico, et la combinaison de LOCKSS et CLOCKSS, complémentaires, se révèle intéressante (cf. tableau 18) : LOCKSS et CLOCKSS couvrent ensemble 69% des titres, et Portico 71%. Sur les 1.043 titres présents dans CLOCKSS (sans accès post-abonnement) et non couverts par LOCKSS, 710 font l'objet d'une clause d'accès post-abonnement sur la plateforme éditeur, dans la plupart des cas gratuit ou presque (AMA, Elsevier, IOP, Nature, SAGE, Wiley).

Total holdings avec ISSN/e-ISSN	LOCKSS	LOCKSS + CLOCKSS	Portico	Portico + LOCKSS + CLOCKSS	Portico + LOCKSS
5.398	50%	69%	71%	79%	79%
Titres avec accès PA via la solution tierce	2.676	2.676	3.586	4.089	4.089
Titres sans accès PA via la solution tierce	-	1.043 (dont 710 avec accès PA sur la plateforme éditeur)	225	189	186
Total	2.676	3.719	3.811	4.278	4.275

Tableau 18 - Couverture du portefeuille ULg – titres payants éditeurs (hors Wiley et Elsevier non core collection) – avec indication de la présence d'une garantie d'accès post-abonnement

La fig. 14 donne un aperçu de la couverture des solutions par discipline²¹⁴, qui, de manière générale, reflète la couverture globale, même si LOCKSS s'approche davantage de Portico dans les sciences humaines et sociales, en mathématiques et en physique. Les disciplines les moins couvertes par les solutions sont, en pourcentage, les sciences économiques et des affaires, les sciences de l'information et la physique.

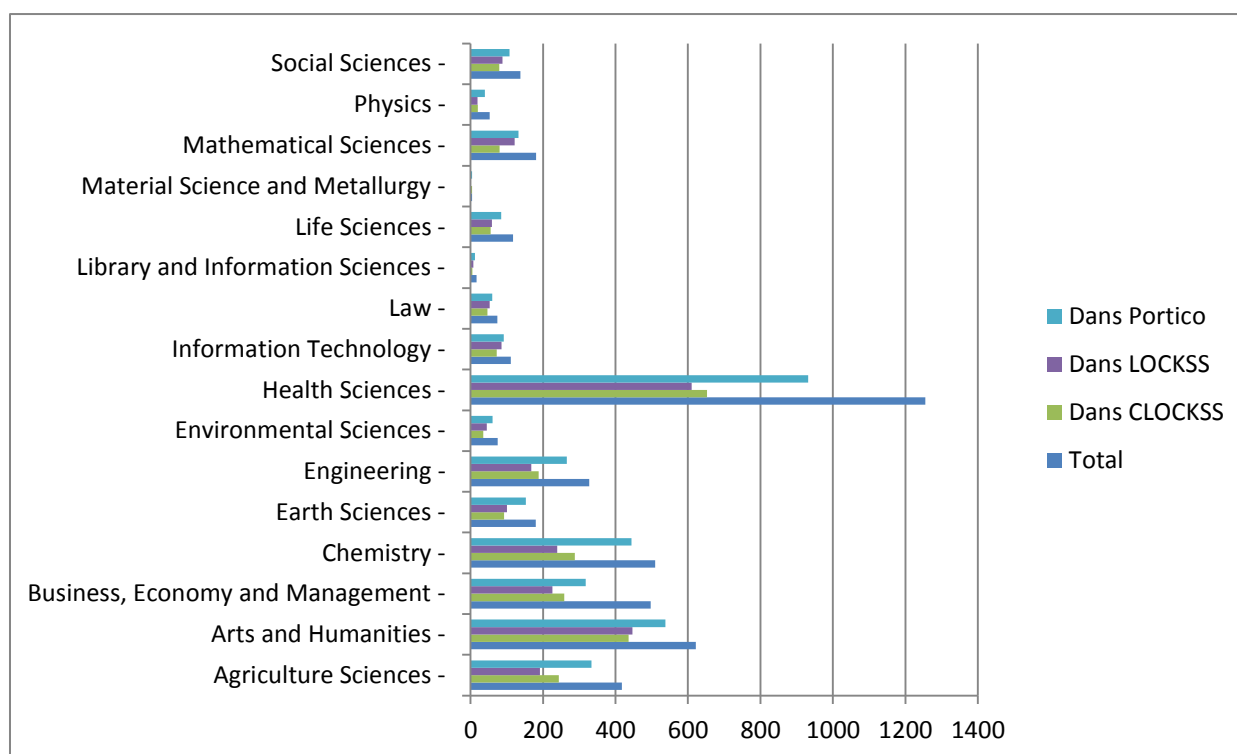


Figure 11 - Couverture du portefeuille ULg – titres payants éditeurs (hors Wiley et Elsevier non core collection) par discipline

Nous synthétisons ci-dessous la couverture des solutions en lien avec la problématique de l'accès pérenne et post-abonnement en situant les grandes cibles sfx sur une échelle : celles qui sont

²¹⁴ Les disciplines sont les disciplines de premier niveau de la classification de sfx.

absentes de toutes les solutions, et constituent des points d'action prioritaires si les institutions de la BICfB souhaitent participer à l'une ou l'autre solution mais étendre la couverture de leur portefeuille, celles qui sont absentes des solutions gérant l'accès post-abonnement (LOCKSS et Portico), mais proposent une alternative pour l'accès post-abonnement (CLOCKSS, e-Depot, accès via leur plateforme), celles qui sont présentes dans LOCKSS ou Portico et absentes de l'autre solution.

Cibles absentes ²¹⁵ (tableau détaillé en annexe)		
De toutes les solutions	OECD Periodicals ²¹⁶ , Metapress IOS Press, Jura e-journals, EDP Sciences, Peeters, Operation and Development, APS*, NCBI, International Press, John Libbey (mais aussi Kluwer, Masson)	
Des solutions avec PCA (LOCKSS et Portico avec PCA)	Annual Review* ²¹⁷ , ACS*, ASCE, AMA, CUP, Brill, APS Prola*, AIP, CSIRO, World Scientific Publications	Mais tous les grands absents de Portico avec accès post-abonnement sont présents dans Portico sans accès post-abonnement et parmi ceux-ci, Annual Reviews*, ACS*, AMA, CUP, AIP, CSIRO, proposent un accès post-abonnement sur plateforme.
De LOCKSS (mais dans Portico)	Elsevier* (y compris MD Consult, Cell Press*), Wiley*, MA Liebert, IOP, OVID Journals ²¹⁸ , titres Ingenta, Nature*	La majorité des grands absents de LOCKSS mais présents dans Portico sont archivés dans CLOCKSS : Elsevier*, Wiley*, ACS, ASCE, AMA, AIP, IOP, Nature*, Informa Healthcare, Annual Reviews* (<i>backfiles</i>), Brill. Sont absents de CLOCKSS : APS*, CUP, SCIRO, World Scientific Publications. Sont aussi présents dans e-Depot : Elsevier*, Blackwell*, Brill, DOAJ, Mary Ann Liebert
De Portico (mais dans)	Principalement des titres de la plateforme Highwire et RSC	

²¹⁵ Comme expliqué plus haut, nous avons dû travailler avec la notion de seuil pour contourner le fait que les listes de titres des solutions ne mentionnent pas la plateforme et que nous n'avons pu comparer les états de collection, afin de limiter le biais dû au transfert des titres et à la publication sur plusieurs plateformes. Nous avons fixé le seuil à 50% des titres présents dans la solution pour indiquer la cible comme couverte par cette solution. Nous reprenons ici seulement les cibles les plus importantes (au moins 5 titres).

²¹⁶ OECD iLibrary, la plateforme de publication des sources OECD (y compris les e-journaux), et négociée par la BICfB depuis 2012, garantit dans son offre un accès post-abonnement sur la plateforme pour les périodiques, ouvrages et articles, et sous la forme de fourniture de CD en fin de période d'abonnement pour les bases de données.

²¹⁷ V. Reich nous a communiqué par email en juillet 2013 qu'Annual Reviews donnait la permission d'ingestion dans la LOCKSS box bibliothèque par bibliothèque, et non via le GLN.

²¹⁸ En réalité, OVID est un agrégateur pour les e-journaux. La cible comprend en majorité des titres édités par Lippincott, Williams & Wilkins et par l'OCDE/OECD, ainsi que des titres de sociétés savantes (dont quelques-uns sont archivés dans l'une ou l'autre solution).

LOCKSS)

Tableau 19 - Grands éditeurs (cibles sfx) absents des solutions tierces et combinaison de solutions tierces et contractuelles. [* Contrats BICfB]

4.2.3.3.4 E-journaux gratuits et Open Access

La couverture des titres du DOAJ par les différentes solutions a déjà été examinée plus haut (cf. point 4.2.3.2.1). On peut relever ici les quelques grandes cibles supplémentaires (plus de 50 titres), présentes dans l'une ou l'autre solution :

Cibles Ex Libris	Total titres ULg	Dans Aucune	Dans CLOCKSS	Dans LOCKSS	Dans Portico	Dans Portico avec accès PA
BioMed Central Open Access Free	104	29,8%	17,3%	69,2%	52,9%	52,9%
Highwire Press Free	261	4,2%	36,0%	93,1%	46,7%	43,3%
Medknow Open Access Journals	75	33,3%	0,0%	66,7%	49,3%	49,3%
Astrophysics Data System	53	62,3%	18,9%	15,1%	35,8%	24,5%
DigiZeitschriften Open Access	141	85,8%	7,8%	11,3%	11,3%	11,3%
Hindawi Publishing Open Access Journals	93	61,3%	0,0%	36,6%	34,4%	34,4%
JSTOR Early Journal Content	383	72,8%	2,9%	14,9%	23,0%	20,1%

Tableau 20 - Couverture (en %) des grandes cibles Open Access (plus de 50 titres) par les solutions d'archivage CLOCKSS, LOCKSS et Portico

Il est aussi intéressant de rappeler que les titres Open Access archivés dans LOCKSS (GLN) sont accessibles à toute LOCKSS Box, que l'institution soit membre ou non de la LOCKSS Alliance. Rappelons également qu'en février 2013, 900 titres de 650 éditeurs du DOAJ étaient archivés dans e-Depot.

4.2.3.4 Conclusions

L'analyse des titres et éditeurs archivés par les différentes solutions, mise en regard avec le portefeuille ULg, a permis d'affiner les questions, problématiques et points d'action pour la BICfB en fonction de la typologie des contenus : titres courants acquis auprès d'éditeurs ou d'agents, bouquets d'archives, titres acquis via des plateformes d'agrégation (CAIRN, JSTOR) et des bases de données full-text (EBSCO, ProQuest), titres gratuits et Open Access.

En ce qui concerne les **titres acquis auprès des éditeurs ou des agents** (plateformes primaires), la couverture des solutions LOCKSS, CLOCKSS et Portico est mature et globalement **suffisante pour garantir une assurance sur l'investissement** : près de 80% de ces titres (hors titres des *big deals* non *core collection* pour lesquels l'accès post-abonnement n'est pas autorisé) sont couverts (Tableau 17) par l'une ou l'autre solution, Portico présentant la couverture la plus importante (71% des titres).

Cependant, les profils des solutions, et les garanties et contraintes qu'elles offrent aux éditeurs sont différentes et ne se recouvrent pas entièrement : CLOCKSS et Portico archivent davantage de collections de grands éditeurs, pour qui LOCKSS n'est pas une alternative envisageable (en particulier Wiley, Elsevier, APS). LOCKSS travaille avec davantage de petits éditeurs, en particulier ceux de la plateforme Highwire. En distinguant la problématique de l'accès post-abonnement, qui peut être garanti contractuellement sur la plateforme de l'éditeur, de la problématique de l'archivage pérenne (nécessairement via une solution tierce), on constate que **CLOCKSS et LOCKSS, en combinaison,**

garantissent l'archivage pérenne d'une portion du portefeuille examiné presque aussi importante (69%) que celle que couvre Portico (71%). Même s'il est peu probable que les couvertures des solutions franchissent les limites imposées par leurs choix de gouvernance, il est intéressant pour la BICfB, en fonction des choix qu'elle posera, d'explorer leur combinaison et d'agir en fonction, en particulier, pour les grands éditeurs : en négociant l'accès post-abonnement via la plateforme de publication et leur participation à CLOCKSS ou Portico ; pour les petits éditeurs, en négociant leur participation à l'une des solutions proposant un accès post-abonnement (LOCKSS, Portico avec accès PA).

Les **acquisitions via des agrégateurs et bases de données full-text** ont un statut beaucoup plus complexe et ne sont que fort peu couvertes par les trois solutions. Elles ne font visiblement pas partie du *core business* de ces dernières, ce qui s'explique par le statut d'intermédiaire des agrégateurs de données, déjà souligné lors de l'analyse des licences. Les données concernant la plateforme de publication ne sont d'ailleurs pas disponibles dans les listes des solutions, ce qui rend difficile l'évaluation de cette couverture (les titres inclus dans ces bases de données ou plateformes peuvent être couverts par une solution, mais via les canaux primaires de publication). Or, s'ils en avaient les droits, les agrégateurs pourraient jouer un rôle en assurant l'interface technique pour le moissonnage des données (LOCKSS, CLOCKSS) ou le transfert des fichiers (CLOCKSS, Portico) – comme dans le cas des plateformes primaires de publication (Atypon, Highwire, Ingenta...). La question nous semble ici liée au modèle d'acquisition et aux priorités définies en termes de gestion de collections : la souscription à des bases de données full-text doit-elle être considérée comme un achat pérenne ou plutôt comme une fourniture d'accès, évoluant dans le temps ?

Les **bouquets d'archives** des grands éditeurs sont présents dans les solutions en fonction de la politique globale de l'éditeur, comme les titres courants, et en termes de titres, ils sont donc bien couverts. Cependant, la couverture rétrospective n'est pas, actuellement, suffisamment assurée, ce qui plaide en défaveur de l'achat de bouquets d'archives. La situation n'est pas dramatique dans la mesure où il n'y a pas de barrière contractuelle et où cet état de choses n'est peut-être que ponctuel.

En ce qui concerne les **titres gratuits et Open Access**, les solutions ne se distinguent pas fondamentalement si ce n'est concernant les titres d'éditeurs publiant sur la plateforme Highwire, nettement plus présents dans LOCKSS que dans CLOCKSS ou Portico. A nouveau se pose ici la question de la définition de la collection par les bibliothèques, et de la responsabilité de l'archivage : quels titres Open Access sont destinés à faire partie de la *core collection* des bibliothèques – et doivent être archivés, et par qui ces contenus doivent-ils être archivés? Qui doit porter la responsabilité de l'archivage de contenus numérisés et accessibles sur des portails comme Gallica ? Ces questions sont difficiles et sortent en partie du champ de cette étude, mais ne pourront, à termes, être éludées.

Couverture	LOCKSS	CLOCKSS	Portico
Grands éditeurs	Un certain nombre, mais grands absents	La majorité, avec quelques exceptions	La majorité
	L'extension de la couverture a probablement atteint ses limites, car il s'agissait de contenus prioritaires dans les trois cas		
Petits éditeurs	Oui, nombreux	Oui, mais moins que LOCKSS	

	L'extension semble encore possible, mais limitée par des questions de temps, voire, dans le cas de CLOCKSS et Portico, par des questions de priorités et des questions financières	
Agrégateurs et bases de données full-text	<ul style="list-style-type: none"> • Peu, ne font pas partie de la <i>core collection</i> des solutions • L'extension est à la fois favorisée (au niveau technique) et limitée (au niveau juridique) par le statut d'intermédiaire des agrégateurs 	
Bouquets d'archives	<ul style="list-style-type: none"> • Oui, mais sans couverture rétrospective complète actuellement • L'extension n'est sans doute qu'une question de temps 	
Titres gratuits et Open Access	Oui, relativement, un peu plus que dans CLOCKSS et LOCKSS	Oui, relativement Oui, relativement
	L'extension est surtout une question de priorités et de définition des responsabilités	
Points d'action	Combiner les deux solutions, négocier avant tout l'accès post-abonnement sur la plateforme de l'éditeur pour les grands éditeurs.	
	Définir une politique d'archivage pour les agrégateurs et bases de données et les titres gratuits.	
	Négocier l'extension de la couverture rétrospective pour les bouquets d'archives (Springer, déjà acheté par la BICfB), et d'autres si souhait d'achat ; suivre la rétro-ingestion.	

Tableau 21 - Situation et points d'action en termes de couverture des solutions, par type de contenu

4.2.4 Contrôle

Le contrôle est l'un des aspects où transparaît de la manière la plus évidente la différence de philosophie des solutions. LOCKSS a pour objectif de rendre le contrôle des ressources archivées aux bibliothèques, CLOCKSS également, mais avec une gestion centralisée par quelques grandes bibliothèques académiques dans le monde, alors que Portico fonctionne plutôt comme une assurance-vie²¹⁹, assurant l'archivage en externe pour le compte des bibliothèques. Le degré de contrôle des bibliothèques est donc logiquement le plus important dans le cas de LOCKSS, et est extrêmement faible dans le cas de la solution d'*outsourcing* qu'est Portico.

Pour **LOCKSS**, le contrôle partagé est un élément de gestion des risques. LOCKSS assure le contrôle total des données physiques (*bits*) par les institutions, puisqu'elles sont hébergées en local. Le seul risque – très mineur – à ce niveau est l'utilisation d'un logiciel en réseau, mais le contrôle sur les aspects logiciels est également tout à fait possible puisque le logiciel est Open Source, et que les institutions ont le loisir de le tester, de décider du timing de mises à jour, et de contribuer à son développement. LOCKSS assure que c'est l'objectif, et travaille avec des responsables techniques en Angleterre (dans le cadre de la UK LOCKSS Alliance) et en Allemagne (dans le cadre du projet LUKII). Cependant, à l'heure actuelle, les contributions externes au logiciel sont encore très réduites (configuration de plug-ins²²⁰, projet LUKII). Malgré l'existence de la LOCKSS Alliance présentée comme un réseau où les membres ont des voix égales et peuvent suggérer des développements pour le logiciel et des priorités pour les négociations avec les éditeurs, et l'ouverture à la discussion de la

²¹⁹ La comparaison vient de Portico.

²²⁰ Conversation téléphonique avec V. Reich le 29 août 2013.

Directrice exécutive (V. Reich), il est difficile de se faire une idée exacte des processus de décision. La gestion est moins formalisée juridiquement que celle de Portico, et les processus également.

Dans le cas de **CLOCKSS**, malgré l'utilisation du même logiciel, les institutions membres n'ont aucun contrôle réel puisque les décisions (d'archivage et d'ouverture des accès en cas d'événement déclencheur) sont prises unilatéralement par le conseil d'administration de CLOCKSS. Toutefois, dans la mesure où il ne s'agit pas d'une société commerciale mais d'une collaboration où le conseil est composé des plus grands éditeurs et des bibliothèques académiques participantes – qui possèdent les données en local, les intérêts des bibliothèques sont en tous cas représentés par plusieurs acteurs importants, qui agissent au nom de leur institution.

La question du contrôle est le point faible de **Portico**. En effet, même si Ithaka est une société non commerciale et si son conseil d'administration, comme celui de CLOCKSS, est composé d'éditeurs et de bibliothécaires d'institutions importantes, en pratique, la souscription à Portico équivaut à un *outsourcing* total de la garantie d'archivage pérenne, même si sa collaboration avec e-Depot garanti la possession des données par au moins une bibliothèque majeure. En outre, malgré son statut et ses objectifs appropriés à la conservation pérenne, le *business model* de Portico est proche de celui des grands éditeurs commerciaux : les licences ont une durée de 3 ans, et les droits d'usage sont assez restrictifs pour les bibliothèques, de même que le processus d'autorisation d'accès aux ressources (cf. supra 4.1.2). Un accès à des fins de monitoring des données sur le serveur de Portico est autorisé pour les institutions, mais Portico spécifie explicitement en termes juridiques que toutes les technologies employées sont propriétaires.

En conclusion, seule la possession des données en local proposée par la solution LOCKSS offre un contrôle réel des institutions sur l'archivage pérenne. La gestion déléguée et l'assurance juridique proposées par Portico ne peuvent être jugée suffisante à notre avis dans le cadre des risques liés à l'archivage pérenne. Dans le cas de CLOCKSS, même si aucun contrôle effectif ne peut être effectué par les institutions, et malgré la délégation de la gestion par les bibliothèques qui n'hébergent pas les données, la possession des données par au moins trois bibliothèques majeures sur le plan international et la gestion collaborative du projet par les différents acteurs offre des garanties importantes en termes d'archivage pérenne, de même que l'utilisation du logiciel LOCKSS. La question cruciale est ici la viabilité de l'organisation : même si elle existe depuis 2006, elle a besoin de l'adhésion des bibliothèques pour conserver un poids dans le paysage de l'archivage pérenne et développer son périmètre d'archivage.

Degré de contrôle	LOCKSS	CLOCKSS	Portico
Contrôle sur les données	<ul style="list-style-type: none"> • Contrôle total des données, hébergées en local • Monitoring par les pairs 	<ul style="list-style-type: none"> • Outsourcing des données à un ensemble de grands éditeurs et bibliothèques, avec un contrôle mutuel • Pas de monitoring par les institutions non hébergeurs, mais système de monitoring par les pairs 	<ul style="list-style-type: none"> • Outsourcing complet, contrôle relatif du monde des bibliothèques via e-Depot (l'un des sites miroirs de Portico). • Monitoring très limité de l'archive via la plateforme
Contrôle sur le logiciel	Logiciel Open Source, mais la documentation technique détaillée semble encore assez limitée	documentation technique détaillée	Aucun contrôle sur le(s) logiciel(s), peu de visibilité, même si Portico envisage

			d'utiliser des modules Open Source existants pour certaines fonctionnalités
Contrôle sur la gestion	Gouvernance partagée en théorie du GLN, mais semble fort dépendante de Stanford (pour les choix de négociations et les développements de plug-ins)	<i>Outsourcing</i> à quelques grandes organisations, mais participation à la gouvernance, équilibrée entre les bibliothèques et les éditeurs	<i>Outsourcing</i> à une seule organisation, participation limitée à la gouvernance ; modèle de gouvernance plutôt orienté éditeurs
Points d'action	Vérifier l'effectivité de la participation et de la collaboration en termes de gouvernance	Tester la participation aux structures de gouvernance	Tester l'interface de monitoring de Portico

Tableau 22 - Synthèse de la comparaison du degré de contrôle dans LOCKSS, CLOCKSS et Portico

4.2.5 Coûts

On a d'abord identifié les coûts non couverts par la souscription à chaque solution et à charge des institutions, avec les investissements matériels et en ressources humaines correspondants, sur base d'un découpage en activités inspiré du modèle OAIS. Pour chacune des activités à charge des institutions, on a indiqué les actions spécifiques à effectuer pour chacune des solutions, en incluant les aspects techniques, de négociation et de gestion de l'information. Outre clarifier les coûts à charge des institutions, cette analyse peut éventuellement par la suite permettre de faire des parallèles avec d'autres types de solution, notamment des solutions *in house*, ou des solutions d'archivage papier²²¹. On a ensuite tenté d'évaluer les coûts globaux des différentes solutions (frais de participation, coûts matériels et humains), initiaux et sur une échelle de 10 à 20 années²²², en fonction des activités et de paramètres de coûts, en notant les limitations fortes inhérentes au modèle prévisionnel. On évalue ensuite le coût des différentes solutions en fonction du retour sur investissement en matière d'acquisitions électroniques.

4.2.5.1 Schéma d'activités non couvertes par les solutions (à charge des institutions)

Le tableau ci-dessous reprend, pour chacune des solutions, et en fonction des grandes catégories fonctionnelles de la préservation pérenne, les activités à charge des solutions. Chaque activité a été étiquetée en fonction de sa fréquence (I = coûts initiaux, R = coûts récurrents, P = coûts ponctuels (en fonction des titres ou plateformes)), et en fonction du grade des ressources humaines nécessaires (B1 : universitaire non informaticien, B2 : bibliothécaire gradué, I1 : universitaire informaticien, I2 : gradué informaticien). On a aussi identifié les ressources matérielles nécessaires pour LOCKSS.

Les activités de *preservation planning*, c'est-à-dire tout ce qui concerne le développement du système, la planification stratégique, la veille technologique, la création des *templates* et stratégies techniques pour les différentes opérations en fonction des types de contenus archivés, les négociations avec les fournisseurs de contenus (en l'occurrence les éditeurs et /ou agrégateurs), sont, dans tous les cas, assurées par les solutions tierces considérées ici. Ces activités sont d'ailleurs

²²¹ Ce modèle a pour objectif de schématiser les coûts de manière comparable, la répartition des activités au sein des différentes fonctions est quant à elle révisable, et est d'ailleurs effectuée différemment dans les différents modèles de coûts de la préservation pérenne examinés (KDRS, LIFE, CMDP).

²²² Selon la littérature, l'estimation des coûts devient plus hasardeuse au-delà, même si on peut aller jusque 20 ans.

les plus coûteuses de l'archivage pérenne²²³, et bénéficient donc majoritairement de la mutualisation des ressources. Nous n'avons pas considéré ici les activités initiales de sélection et d'acquisition des contenus (constitution de collection), ni leur catalogage et la gestion de l'accès courant (notamment via l'activation des cibles dans les résolveurs de liens). Ces opérations sont nécessaires à l'archivage pérenne mais en sont indépendantes.

Les opérations concernant la sélection des contenus à archiver et la vérification de leur présence dans la/les solution(s) tierces, par contre, sont bien à inclure dans les coûts d'accès pérenne aux ressources. Le coût de ces opérations dépend en partie des outils de gestion de l'information disponibles pour chacune des solutions (outils d'export et de comparaison de listes de contenus).

	LOCKSS	Portico	CLOCKSS
	ACTIVITES		
Preservation planning	Ressources humaines : [I+R] [B1] sélection des titres à archiver; vérification de leur présence dans la solution (f.titre); Pour les plateformes ou fournisseurs non inclus dans la solution : ([P] [B1] demande de création de préservation plan si nouvelle plateforme (f.plateforme)) ([P] [B1] négociation de l'autorisation de moissonnage avec les fournisseurs (f.fournisseur))		
Ingestion	Ressources humaines : [I+R] [B2] Activation des titres dans la LOCKSS box (f.titres) [I+R] [B2] vérification du statut d'import (f. titre * année)	Coûts matériels : 1 serveur basique ; 4 Tb stockage (recommandation LOCKSS 2013) (f. titre * année)	Ressources humaines : / / /
Administration, stockage physique et gestion des données	[I] [I1] Installation et configuration du logiciel [I+R][I2] Installation, configuration et remplacement media (f. durée de vie media et volume media) et serveur; [R] [I2] mât du logiciel, maintenance et monitoring du serveur ([R] [I1] installation et configuration / développement pour la réalisation de migrations périodiques (f. durée de vie format * nb fichier)		/ ([R] [B2] Vérification de la préservation sur serveur Portico) /
Accès	[I] [B2] Activation de la cible dans résolveur de liens ou configuration proxy avec [P] modification des liens dans le catalogue (f. plateforme / titre)	(en fonction des tests d'accès et de l'usage attendu, 1 serveur plus performant)	[P] [B1] Négociation de l'accès via Portico en cas de fin d'abonnement [I] [B2] Activation de la cible dans résolveur de liens ou [P] modification des liens dans le catalogue (f. plateforme / titre)

Tableau 23 - Activités de préservation prises en charge par les institutions dans LOCKSS, CLOCKSS et Portico
Légende: I = couts initiaux, R = couts récurrents, P = couts ponctuels (en fonction des titres ou plateformes)
B1 : universitaire non informaticien, B2 : bibliothécaire gradué, I1 : universitaire informaticien, I2 : gradué informaticien.

²²³ Le Cost Model for Digital Preservation (Slats, J., et R. Verdegem. « Cost Model for Digital Preservation ». In Proceedings of the IVth triennial conference, DLM Forum, Archive, Records and Information Management in Europe, 2005.
http://dlmforum.typepad.com/Paper_RemcoVerdegem_and_JS_CostModelfordigitalpreservation.pdf), identifiait le coût des activités initiales de développement et de mise en place du système à 1 à 2 année(s)/homme.

4.2.5.2 Coûts initiaux et annuels des différentes solutions

En fonction du schéma des activités établi ci-dessus et des informations sur le fonctionnement des solutions, on peut établir la synthèse des éléments de coûts à prendre en compte pour les différentes solutions. Outre les coûts liés à la gestion de l'information et à la vérification des titres disponibles dans les différentes solutions (*preservation planning*), Portico suppose des coûts de souscription et un minimum de coûts en ressources humaines, CLOCKSS suppose uniquement des coûts de souscription, LOCKSS par contre suppose des coûts de souscription, mais aussi des coûts en ressources humaines, y compris informatiques, et des coûts en matériel.

	LOCKSS	Portico	CLOCKSS
COÛTS			
Financiers (souscription)	Coûts récurrents (\$)		
	<i>Souscription annuelle obligatoire (sauf la première année) pour l'archivage et l'accès aux contenus</i>	<i>Souscription annuelle obligatoire pour l'accès aux contenus</i>	<i>Souscription annuelle facultative pour l'archivage et l'accès aux contenus</i>
	<i>Calcul : prix souscription (f (taille institution)) + à p. de l'année 2 : prix souscription (f (taille institution)) *1+ inflation annuelle de la souscription (à p. de année 2)</i>		
en ressources humaines (cf. schéma des activités)	Coût initiaux : (heures)		
	[B1] Preservation planning (sélection, vérification et négociations)	Preservation planning (sélection, vérification et négociations)	Preservation planning (sélection, vérification et négociations)
	[B2] <ul style="list-style-type: none"> Activation de la cible dans le résolveur de liens ou configuration proxy Activation initiale des titres dans la LOCKSS box (f.titre) et vérification de l'import [16h]	<ul style="list-style-type: none"> Activation de la cible dans le résolveur de liens ou configuration proxy (Vérification de la preservation sur le serveur Portico) vérifications et négociations d'accès post-abonnement via Portico en cas de fin de contrat [2h]	<ul style="list-style-type: none"> Activation de la cible dans le résolveur de liens ou configuration proxy [0h]
	[I1] <ul style="list-style-type: none"> Installation et configuration du logiciel [4h]	/	/
	[I2] <ul style="list-style-type: none"> Installation, configuration et media et serveur – f(nombre de titres, usage attendu) [4h]	/	/
	<i>Calcul : coût/heures (fonction) * nb heures (fonction)</i>		
	Coût récurrents ou ponctuels : (heures/mois)		
	[B1] <ul style="list-style-type: none"> Preservation planning (sélection, vérification et négociations pour les nouveaux abonnements) 	<ul style="list-style-type: none"> Preservation planning (sélection, vérification et négociations pour les nouveaux abonnements) vérifications et négociations d'accès post-abonnement via Portico en cas de fin de contrat [1h]	<ul style="list-style-type: none"> Preservation planning (sélection, vérification et négociations pour les nouveaux abonnements)

	[B2]	<ul style="list-style-type: none"> • Activation des nouveaux titres dans la LOCKSS box (f.titre) • Modification des liens dans le catalogue si configuration proxy et événement déclencheur (f. plateformes / titres) • Check statut d'import des nouveaux titres (f. titres - année) 	<ul style="list-style-type: none"> • Check preservation sur serveur Portico) (via l'interface de monitoring) 	/	
	[I1]	<ul style="list-style-type: none"> • maintenance logiciel (installation updates etc.) 		/	/
	[I2]	<ul style="list-style-type: none"> • maintenance et remplacement media (f. durée de vie media et volume media) et serveur 		/	/
Calcul : coût/heures (fonction) * nb heures (fonction) /année * avec augmentation annuelle en fonction de l'index (suite géométrique)					
en ressources matérielles (cf. schéma des activités)		<ul style="list-style-type: none"> • 1 serveur logiciel ; 4 Tb stockage (recommandation LOCKSS 2013) (f. année*titre) • (en fonction des tests d'accès et de l'usage attendu, 1 serveur plus performant) 		/	/
Calcul :					
<p>Serveur = coût serveur (f(capacités))/durée_vie_serveur * augmentation annuelle en fonction de l'index (suite géométrique)</p> <p>Stockage = coût stockage /Tb * Nb Tb_{année} /durée_vie_media * augmentation annuelle en fonction de l'index (suite géométrique)</p>					

Tableau 24 - Coûts des activités de préservation prises en charge par les institutions dans LOCKSS, CLOCKSS, Portico

Sur base de ces éléments, on a effectué une estimation des coûts de chaque catégorie (coûts de souscription, coûts en ressources humaines, coûts en matériel) pour les différentes solutions, avec une projection sur 20 ans, afin d'évaluer dans quelle mesure la gestion des données en local de LOCKSS pouvait être désavantageuse à terme en matière de coûts, et d'estimer le coût de l'archivage pérenne sur le cycle de vie des documents.

4.2.5.2.1 Coûts de souscription

Les **coûts de souscription** ont été évalués sur base des tarifs des différentes solutions en vigueur en 2013²²⁴ (cf. tableau 25). Les trois solutions ont une grille tarifaire basée sur la taille de l'institution, en

²²⁴ Les grilles tarifaires sont disponibles sur les sites web des solutions. Elles n'ont pas changé entre 2013 et février 2014. En ce qui concerne LOCKSS toutefois, seuls les prix pour des institutions US sont publiés, il faut contacter LOCKSS pour avoir les équivalents européens.

fonction du budget total en acquisitions documentaires²²⁵ pour Portico et CLOCKSS, en fonction de la classification JSTOR pour LOCKSS.

Souscription annuelle	Données institutions		LOCKSS ²²⁶		Portico	CLOCKSS
Institutions	Budget en acquisitions documentaires ²²⁷	Classification JSTOR ²²⁸	Tarif officiel US estimé	Prix proposé par LOCKSS	Tarif officiel	Tarif officiel
ULg	3.708.299,79 €	large	\$10.240,00	\$9.357,25	\$9.167,00	\$1.800,00
ULB	3.741.094,00 €	large	\$10.240,00	\$9.440,00	\$9.167,00	\$1.800,00
UCL	2.895.492,00 €	large	\$10.240,00	\$7.306,27	\$8.034,00	\$1.200,00
UNamur	977.552,43 €	medium	\$4.740	\$2.466,68	\$5.562,00	\$450,00
UMONS	512.081,00 €	medium	\$4.740	\$1.292,15	\$4.326,00	\$450,00
USL-B	298.779,75 €	small	\$3.930	\$753,92	\$3.193,00	\$450,00

Tableau 25 - Coûts de souscription annuelle aux différentes solutions pour les institutions de la BICfB (tarifs 2013)

Tant Portico que LOCKSS proposent immédiatement 5% de réduction (dans le cas de Portico, cette réduction s'applique si le consortium accepte de faire la promotion de Portico), CLOCKSS propose également des réductions consortiales.

La politique actuelle de LOCKSS est d'appliquer une augmentation annuelle de 3% pour les tarifs des nouveaux souscripteurs, de ne pas augmenter par contre le tarif des membres souscrivant déjà (ce qu'ils ont maintenu jusqu'à présent). La politique de Portico est de ne pas appliquer d'augmentation annuelle (il y a eu une augmentation des tarifs de 3% entre 2006 et 2013), par contre, le contrat a une durée de 3 ans, renouvelable ensuite. CLOCKSS vise une diminution des coûts à terme (cf. supra 4.2.1.3.2).

On peut donc dire que, théoriquement, les coûts de souscription de Portico et LOCKSS sont à peu près équivalents ; LOCKSS est très ouvert à la négociation, aucune négociation n'a par contre été entamée avec Portico. La souscription à CLOCKSS, facultative, est extrêmement peu coûteuse en comparaison, mais il faut garder à l'esprit que CLOCKSS est une *dark archive*, qui ne propose pas d'accès post-abonnement.

Il faut souligner finalement que Portico facture de manière indépendante la préservation des e-journaux et des autres contenus (e-books, D-Collections²²⁹), ce qui n'est pas le cas de LOCKSS, où le

²²⁵ Total Library Material Expenditure (LME). Le LME considéré par Portico est celui que définit l'ARL pour la récolte de statistiques des bibliothèques, soit le coûts total en acquisitions documentaires, imprimées et électroniques, incluant tous les types d'acquisitions (périodiques, monographies, autre type de ressource).

²²⁶ Les prix US estimés ont été estimés sur base des désignations des catégories JSTOR et des montants de souscription Portico, qui sont équivalents pour les grandes institutions à ceux proposés par LOCKSS. Les équivalents théoriques des coûts de souscription selon la classification JSTOR pour les institutions de la BICfB communiqués par mail par V. Reich en août 2013 étaient très désavantageux, sans doute suite à des erreurs de catégorisation ou à une mauvaise adaptation de la classification JSTOR à la situation des institutions de la BICfB. LOCKSS a donc proposé d'appliquer le rapport budget acquisitions documentaires / coût de la souscription de l'ULB, à laquelle elle a déjà fait une proposition financière en 2012, à l'ensemble des institutions de la BICfB. Ces équivalents calculés sont repris dans le tableau sous « prix proposés par LOCKSS », et sont extrêmement avantageux par rapport aux prix US estimés. Ils n'ont pas été validés par LOCKSS en tant que tels.

²²⁷ Source: statistiques CIUF, 2011.

²²⁸ Source: estimation théorique de la classification, la classification actuelle des institutions dans le cadre de leur souscription aux collections JSTOR n'étant vraisemblablement pas correcte.

paiement de la participation à la LOCKSS Alliance couvre l'utilisation du logiciel et l'aide de l'équipe de Stanford non seulement pour la préservation de tous les types de contenus (e-journaux et e-books) disponibles via la LOCKSS Alliance, mais également pour la création de réseaux privés LOCKSS (PNL).

4.2.5.2.2 Coûts en ressources humaines

Chacune des activités à charge des institutions a été évaluée en **coûts humains** en heures de travail pour les coûts initiaux et en heures/mois pour les coûts récurrents, pour les quatre fonctions identifiées plus haut. Certaines de ces activités sont fonction du nombre de titres ou de plateformes d'éditeurs, mais on a décidé de ne pas ajouter de paramètres supplémentaires, tablant sur un développement régulier de l'archive, et sur une augmentation faible du volume de données (cf. infra 4.2.5.2.3). L'évaluation en heures se base sur les informations fournies par les solutions (LOCKSS) et les retours d'expérience²³⁰ lorsque ces informations sont disponibles. Les salaires ont été évalués sur base des salaires en vigueur à l'Université de Liège²³¹ et de 1672 heures de travail par an.

Coûts en ressources humaines	LOCKSS	Portico	CLOCKSS
Coûts initiaux	<i>Heures</i>		
B1		0	2
B2		16	0
I1		4	0
I2		4	0
<i>Total coûts initiaux</i>	€ 712,92	€ 53,83	€ -
Coûts récurrents	<i>Heures/mois</i>		
B1		0,1	1
B2		2 ²³²	0,2
I1		2	0
I2		2	0
<i>Total coûts récurrents annuels</i>	€ 2.339,71	€ 588,52	€ -

Tableau 26 - Evaluation des heures d'activités de préservation dans LOCKSS, Portico et CLOCKSS

Dans tous les cas, les coûts en ressources humaines sont assez faibles, même dans le cas de LOCKSS qui suppose une maintenance logicielle et une activation des titres en local – qui ne prend pas plus de quelques heures par mois selon LOCKSS et les retours d'expérience. Nous n'avons pas inclus ici les coûts en ressources humaines liées à la sélection des titres à archiver et aux vérifications initiales des clauses contractuelles, qui peuvent être plus élevés, mais sont relativement indépendantes des solutions, dans la mesure où les outils de comparaison des titres archivés avec le portefeuille des institutions sont similaires dans les 3 solutions.

²²⁹ La préservation des e-books est facturée à la fois aux bibliothèques et aux éditeurs, la préservation des D-Collections uniquement aux éditeurs.

²³⁰ Final Report of the 2CUL LOCKSS Assessment Team, 2011.

<http://2cul.org/sites/default/files/2CULLOCKSSFinalReport.pdf>, p. 7, évalue le temps passé à la maintenance de la LOCKSS Box entre 2001 et 2011 à pas plus de 4 à 6 heures par an (mais très peu de titres préservés).

²³¹ 73.000 € brut pour un universitaire (bibliothécaire ou informaticien) avec 3 ans d'expérience, 45.000 € brut pour un bibliothécaire ou informaticien gradué.

²³² Estimation de la charge de travail d'activation et de gestion des collections par la UK LOCKSS Alliance (cf. LOCKSS Quickstart Guide, op. cit.)

4.2.5.2.3 Coûts en matériel

L'évaluation des **coûts matériels**, en particulier des coûts de stockage physique, nécessaire pour LOCKSS, est rendue difficile par la difficulté de disposer de données de volumes d'e-journaux en Mb représentatives et d'évaluer l'accroissement de ce volume, et par l'incertitude quant à l'évolution des coûts des supports physiques.

En ce qui concerne l'évaluation du volume et de l'accroissement, on a utilisé comme étalon de mesure l'**année-titre**, soit le volume de publication d'une année d'un titre de périodique²³³. L'évaluation du **volume de stockage** d'une année-titre a été estimée en fonction des données de stockage de la LOCKSS Box de Stanford : en juillet 2013 : 8,7 Tb pour 10.496 titres (7.501 avec indication d'années), soit 67.200 Archival Units²³⁴, ce qui correspond, d'après les indications des années archivées et en cours d'archivage dans la liste des titres du GLN, à environ 125.450 titres-année, soit environ **60 Mb pour 1 titre-année**. En imaginant que les publications sont natives électroniques à partir de 1996, et considérant que les publications natives électroniques ont un poids équivalent à environ 1/3 des publications numérisées pour un même nombre de pages²³⁵, on peut estimer un volume titre-année de **36 Mb pour les titres natifs numériques et de 106 Mb pour les titres numérisés**.

Le volume de stockage estimé sur cette base pour le portefeuille ULg courant archivable dans LOCKSS, hors titres Open Access, soit environ 2.600 titres (cf. tableau 18), pour lesquels on considère la couverture courante de 17 années (1996-2013), est de 2,65 Tb pour la première année. LOCKSS recommande cependant 4 Tb de stockage depuis 2013 (2 Tb auparavant), et l'estimation des coûts a été faite sur cette base. Le volume des données préservées dans LOCKSS est donc assez réduit, ce qui est rassurant pour la prévision des coûts.

Toutefois, cette estimation est difficile à maintenir sur un grand nombre d'années, puisque 1) la nature des fichiers d'articles et leur poids peut évoluer, notamment s'ils incluent de plus en plus d'images ou de contenus dynamiques et 2) les fichiers résultants s'ajoutent aux fichiers originaux après migration et la migration des fichiers peut en modifier le volume de stockage (*Life 3 Report*, 2010, cite Chapmann, 2004)²³⁶.

On a estimé le **coût initial au Tb** à 100 €²³⁷, en incluant l'utilisation de la technologie RAID. Aucun autre support de stockage n'a été étudié dans cette étude, la comparaison ayant été effectuée par D. Rosenthal en 2012²³⁸.

Le logiciel LOCKSS est conçu pour consommer des ressources limitées et peut être installé sur un serveur (ou une machine virtuelle) dont les capacités sont équivalentes à un PC de bureau de milieu de gamme²³⁹ ; le logiciel est fourni avec une installation Linux basée sur CentOS 5. Le coût du serveur

²³³ Cette unité présente l'avantage d'être utilisé par ailleurs pour les périodiques papier – ce qui peut faciliter la comparaison – et d'être une unité maniable.

²³⁴ Une Archival units est une unité d'archive (équivalent AIP du modèle OAIS), souvent équivalente à 1 année-titre.

²³⁵ Observations faites sur base d'un échantillon minimum de titres Springer et de titres du portail Persée.

²³⁶ La thèse de D. Rosenthal citée plus haut sur l'obsolescence des formats relativise cet argument dans le cas de LOCKSS.

²³⁷ Un disque LaCie 2big Quadra 4 Tb 2-Disk RAID Hard Drive 301432U coûte 500\$ sur Amazon en février 2014. Des disques durs sans RAID peuvent ne coûter que \$50 au Tb.

²³⁸ Rosenthal 2012 ont comparé le stockage sur disque au stockage via le cloud pour LOCKSS et ont conclu que le stockage sur cloud était beaucoup moins avantageux économiquement.

²³⁹ Les retours d'expérience (INIST, UK LOCKSS Alliance) ont confirmé que LOCKSS fonctionnait effectivement très bien sur des serveurs anciens aux capacités limitées.

a été estimé à 500 €. Ce coût pourrait être revu à la hausse dans le cas où les institutions souhaiteraient installer une LOCKSS Box fortement sollicitée pour l'accès.

Afin d'évaluer les coûts récurrents, on a estimé la **durée de vie des ressources matérielles**, à 4 ans pour le serveur et le disque de stockage.

On n'a pas inclus dans les estimations les coûts en électricité, locaux et bande passante, difficiles à isoler.

Coûts en ressources matérielles	LOCKSS	Portico	CLOCKSS
Coûts initiaux			
Serveur		500 €	-
Disque de stockage	100 €/Tb soit 400 € pour 4 Tb		-
Coûts récurrents			
Remplacement serveur	Tous les 4 ans		-
Remplacement disque de stockage	Tous les 4 ans		-

Tableau 27 - Paramètres des coûts matériels dans LOCKSS, CLOCKSS et Portico

Les coûts en matériel nécessaires pour LOCKSS, basés sur la configuration d'archivage recommandée par l'équipe de Stanford, sont donc extrêmement limités.

4.2.5.3 Projections et scénarios

Une **projection des coûts sur 20 ans** a été effectuée pour les différentes solutions, afin d'évaluer quel impact avait à terme la gestion des données en local de LOCKSS et d'estimer le coût de l'archivage pérenne sur le cycle de vie des documents. Une synthèse des données et résultats est disponible en Annexe 4.

4.2.5.3.1 Hypothèses de travail

De manière générale, on peut normalement estimer une **croissance faible des coûts** – ce qui est d'ailleurs un objectif de la préservation pérenne (cf. supra 1.4 et 4.1.5):

- les coûts **de souscription annuelle**, d'après les *business model* des différentes solutions, devraient rester assez stables dans le temps. Leur augmentation a été estimée à 0,75% pour LOCKSS et Portico (les coûts sont normalement stables, mais le contrat avec Portico, par exemple, a une durée de 3 ans), maintenue nulle pour CLOCKSS, qui vise explicitement une diminution des coûts à terme ;
- les **coûts en ressources humaines et les coûts matériels (pour LOCKSS) des différentes solutions** devraient évoluer en fonction de l'index. On a estimé ici une inflation basée sur l'index santé belge moyen depuis 2011 (2,24%) ;
- **les coûts de stockage physique** des données (pour LOCKSS) ont subi une diminution annuelle de 30% ces dernières décennies (loi de Kryder). Toutefois, nous adoptons une hypothèse plus pessimiste de stabilité des coûts, car la loi de Kryder est actuellement remise en cause (notamment D. Rosenthal, 2012) ; la durée de vie du matériel (serveur + stockage) a été estimée à 4 ans.

Un deuxième facteur important dans la projection des coûts pour LOCKSS, puisque les données sont préservées en local, est **l'accroissement prévu du volume de données**. Dans le cas des e-journaux, on a distingué deux scénarios possibles:

1. **l'archivage des titres courants natifs électroniques** (avec l'hypothèse d'une couverture de 17 années, soit 1996-2013²⁴⁰) selon la couverture actuelle des solutions, et sans augmentation du nombre de titres archivés ; c'est le scénario le plus probable en l'état actuel des choses puisque (1) les dernières statistiques (fig.1) montrent une stabilité ou une légère diminution du nombre de souscriptions annuelles et le contexte économique actuel et l'affaiblissement des *big deals* ne sont pas non plus propices à une extension du nombre de souscriptions et (2) nous avons vu plus haut que l'extension possible de la couverture de LOCKSS semblait limitée (en tous cas pour les grands éditeurs) ; dans ce scénario, la croissance annuelle d'années-titre équivaut au nombre de titres en n-1, soit, pour 2.600 titres à 36 Mb l'année-titre, à un peu moins de 0,1 Tb par an ;
2. **l'archivage de bouquets d'archives** (sachant qu'à l'heure actuelle, leur couverture rétrospective est loin d'être exhaustive), soit des collections numérisées d'avant 1996, dont le nombre d'années-titres peut aller de quelques années à plus de 100 ans ; on a ici, à titre illustratif, estimé les coûts pour la conservation des archives Springer (déjà achetées par les membres du consortium). En cas de préservation de bouquets d'archives, il n'y a normalement pas de croissance du volume de données, puisque l'ensemble de la couverture rétrospective a été numérisée. Le volume de données augmente éventuellement ponctuellement en fonction de l'achat de nouveaux bouquets d'archives.

Un paramètre utilisable ici également est la notion de **durée de vie** utile des publications électroniques, intervenant dans le processus régulier de sélection – élimination des collections. On pourrait estimer par exemple une élimination de 5% des volumes datant de n-15 tous les ans. Toutefois, dans le cadre de la présente étude, concernant des publications numériques peu anciennes (dans le cas des titres natifs) ou achetés pour leur valeur historique (dans le cas des bouquets d'archives), nous avons choisi de pas tenir compte de ce paramètre.

Dans le cas de la préservation d'e-books via LOCKSS, l'accroissement du volume de données préservées serait à évaluer séparément – mais nous n'avons pas envisagé ce scénario dans le cadre du présent rapport.

4.2.5.3.2 Risques

La difficulté dans l'établissement des coûts, soulignée également par plusieurs études, est que la précision et la validité du modèle dépendent d'une description correcte et suffisamment précise des activités, qui varient au cas par cas, selon le modèle de l'organisation, les workflows en place, le type de données à récolter. Le cas de la préservation des e-journaux via des solutions tierces n'est qu'un sous-ensemble des scénarios étudiés par les modèles de coûts, assez simple d'ailleurs : les activités à charge des institutions sont assez limitées, les formats sont assez standardisés (pas de vidéos, pas d'images, généralement des formats pdf ou html), les volumes sont nettement moins importants que pour d'autres types de projets (données de recherche, images...), le volume de données croît moins rapidement que dans d'autres contextes (dépôts de données et documents de recherche par exemple), et l'accès peut s'intégrer via les résolveurs de liens des institutions.

Toutefois, malgré cette relative simplicité et bien que le modèle d'accroissement des coûts utilisé ait l'avantage de permettre les comparaisons, plusieurs **restrictions s'imposent sur la validité des projections**, liées d'une part à l'incertitude des différents paramètres, d'autre part à l'utilisation,

²⁴⁰ En réalité, le nombre d'années sera moindre et variera – en fonction des disponibilités au sein des différentes solutions, en fonction du début de parution du titre en électronique, mais également pour les différentes institutions en fonction du moment du passage à l'électronique. Cette hypothèse permet toutefois d'estimer un coût par année-titre théorique

dans les limites de cette étude, de modèles simples qui devraient être affinés par l'utilisation de modèles financiers plus précis²⁴¹ :

- le taux d'inflation²⁴² : on a utilisé un taux d'index fixe, défini dans la littérature comme le DCF, mais qui présente plusieurs désavantages (Rosenthal et al., 2012) : propension à surestimer l'index, et impact linéaire non réaliste. Il serait utile de se baser sur des projections plus réalistes et non-linéaires ;
- le taux d'augmentation de coût de souscription (0,75% pour LOCKSS et Portico, 0% pour CLOCKSS) a été estimé en fonction des variables connues aujourd'hui et de la politique exprimée des solutions. Ceci est également difficile à prévoir sur le long terme, même si les modèles des solutions semblent relativement stables depuis plusieurs années ;
- l'évolution des coûts des stockages physiques est incertaine, puisqu'il est probable que l'application de la loi de Kryder devienne beaucoup moins évidente dans les années à venir (Rosenthal et al., 2012) et ces coûts sont par ailleurs plus difficiles à prévoir actuellement que sur les 30 dernières années. Or, des valeurs différentes pour ce paramètre peuvent grandement affecter les coûts réels dans le cadre de la préservation de gros volumes de données²⁴³. Dans le cadre de la préservation des e-journaux courants, où le volume et la croissance sont faibles, l'impact devrait rester limité, mais la question doit se poser en cas de préservation de bouquets d'archives.

La part importante des coûts de préservation pérenne assumée par les frais de souscription à la solution peut être un argument à considérer dans l'évaluation des risques financiers. En effet, même si les solutions tierces poursuivent un objectif de bonne gouvernance et de *business model* stable, les prévisions de coûts sont, pour elles aussi, soumises à des facteurs d'incertitude.

Plusieurs modèles de coûts de l'archivage électronique pérenne (Cost Model for Digital Preservation²⁴⁴, Keeping Research Data Safe²⁴⁵) ont établi que les activités les plus coûteuses du processus d'archivage pérenne étaient précisément les activités assumées par les solutions tierces : l'ingestion et la migration, et en particulier les aspects de développement et maintenance logiciels. Or, les stratégies des différentes solutions (LOCKSS et Portico) sont radicalement différentes sur ces points : Portico table sur une gestion centralisée, un import via transmission de fichiers, et une migration logique régulière, alors que LOCKSS automatise le processus via le développement logiciel en ingérant les contenus par *harvesting* et en évacuant la question de la migration (migration théorique *on access*). La stratégie la moins coûteuse semble être celle de LOCKSS : stockage plus réduit²⁴⁶, économie d'outils et activités de vérification et de normalisation de formats, pas d'intervention humaine (si ce n'est de monitoring) lors de l'ingestion et de l'accès. LOCKSS ne doit pas non plus, au contraire de Portico, développer d'interface d'accès public, ni assurer le stockage des données, mais par contre doit maintenir et développer un logiciel complexe. On a souligné plus

²⁴¹ C'est aussi l'un des éléments d'amélioration identifié par le projet 4C.

²⁴² Rosenthal et al., 2012 : « In practice, people applying DCF choose unrealistically high interest rates, making investment in long-term projects much more difficult to justify than it should be. » + problème de modèle en réalité non-linéaire.

²⁴³ L'étude de Rosenthal et al., 2012 met en évidence la différence de capital initial pour une assurance de préservation de stockage de 95%, avec un accroissement annuel du volume de données de 57%, en fonction de la valeur de la loi de Kryder : pour des valeurs de cette dernière entre 5 et 45%, les coûts de capital initial (endowment) varient entre 18 et 6 fois le coût de stockage initial, ce qui montre à quel point la projection à long terme est risquée.

²⁴⁴ 2009 phase 1 - 2012 phase 3 (<http://www.costmodelfordigitalpreservation.dk/>)

²⁴⁵ Beagrie, N., Lavoie, B. F., & Woollard, M. (2010). Keeping research data safe 2. HEFCE. Consulté à l'adresse <http://repository.essex.ac.uk/2147/1/keepingresearchdatasafe2.pdf>.

²⁴⁶ LOCKSS ne doit stocker qu'une copie originale du fichier sur une LOCKSS box ; Portico préserve les fichiers source (plus lourds que les fichiers de présentation) ; les fichiers normalisés, et – lorsqu'une migration aura lieu – les fichiers migrés.

haut que les revenus de Portico, sans financement spécifique externe, semblaient plus importants que les revenus de LOCKSS, ce qui pourrait effectivement confirmer que la préservation des contenus dans Portico est largement plus coûteuse que dans LOCKSS.

Le coût de l'archivage dans Portico étant soumis à la fois aux risques liés à l'évolution du coût de stockage des données et au choix d'un modèle de préservation plus « lourd », il n'est pas raisonnable d'exclure tout risque d'augmentation du coût de souscription. Dans le cas de LOCKSS, le risque financier, lié surtout au stockage des données, est plutôt porté par les institutions, mais reste limité dans le cas de la préservation des e-journaux courants.

Les avancées du projet 4C permettront peut-être d'avancer plus d'éléments, ainsi que des données chiffrées concernant les coûts d'ingestion et de migration.

4.2.5.3.3 Préservation des titres courants

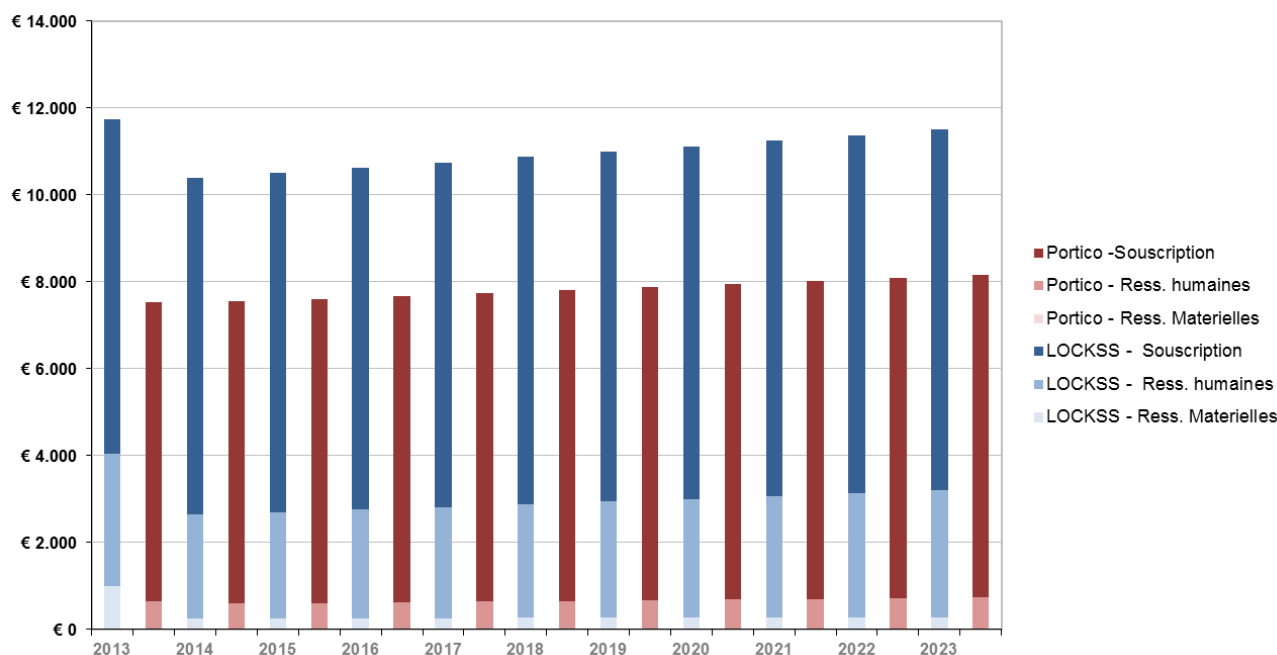


Figure 12 - Evolution des coûts de LOCKSS et Portico sur 20 ans par catégorie de coûts (exemple de l'ULg)

Si l'on n'archive que les titres natifs électroniques couverts par les solutions²⁴⁷, et avec une configuration matérielle de base telle que recommandée par LOCKSS (stockage unique sur disque dur RAID + serveur basique), la **part des coûts de stockage**, dans LOCKSS, à coûts de stockage constant et pour le volume de titres des institutions de la BICfB, est très faible (1.000 € la première année, soit environ 6.500 € sur 20 ans) par rapport au coût total de la mise en place de la solution (coût pour la première année : 11.500 € pour les grandes institutions, entre 7.000 € et 7.600 € pour les petites et moyennes institutions), en particulier de la souscription. Ceci est dû en partie à la politique économique de LOCKSS, dont l'une des forces techniques est justement d'assurer une sécurité maximale des bits sur un stockage à moindre coût, en partie au fait que les e-journaux, en tous cas actuellement, ne sont pas des données lourdes, puisqu'il s'agit majoritairement de textes. Ceci peut évidemment évoluer dans les années à venir, mais ce point (évolution des formats et volumes des publications électroniques) n'a pas été envisagé dans le cadre de cette étude.

²⁴⁷ La comparaison entre la couverture des autres institutions que l'ULg et celle des solutions n'ayant pas été effectuée, on a estimé un nombre de titres à préserver en multipliant le nombre de souscriptions par le rapport entre le nombre de souscriptions et le nombre de titres archivables dans chaque solution pour l'ULg. Le coût global de chaque solution n'est pas fort affecté par cette estimation puisque l'on part de toute façon sur un volume de données de 4 Tb, mais le coût par titre est une estimation plus aléatoire.

Dans l'hypothèse, peu réaliste, où les titres des *big deals* Wiley et Elsevier par exemple pourraient être archivés dans LOCKSS (environ 3.000 titres supplémentaires), et où les institutions privilégieraient cette solution, on pourrait imaginer un volume de stockage initial de 8 Tb, qui, au vu du coût estimé pour 4 Tb, ne modifierait pas radicalement l'ordre de grandeur global de l'investissement.

Le **coût en ressources humaines** est la principale cause de la différence de coûts entre LOCKSS et Portico, dont le montant de la souscription est presque équivalent. C'est évidemment un corollaire du degré de contrôle. Ceci peut représenter un frein pour les petites institutions, qui devraient consacrer à peu près les mêmes ressources pour l'installation et la maintenance de la LOCKSS box que les grandes institutions (un peu moins puisqu'elles doivent activer moins de titres).

La large **part de la souscription annuelle** à la solution dans le cas de LOCKSS et Portico est, malgré un taux d'inflation inférieur à l'index dans le cas de l'hypothèse conservatrice choisie, un facteur de risque (cf. point 4.2.1.3.). Il faut noter toutefois que LOCKSS accorde des réductions de la souscription en cas d'implication de l'institution dans le développement de plug-ins ou du logiciel²⁴⁸, et semble très ouvert à la négociation. Si l'on prend en compte le principe de prix proposé par LOCKSS à la BICfB (de manière non officielle), les coûts de souscription sur 20 ans, s'ils restent sensiblement les mêmes pour les grandes institutions, peuvent être divisés par 4 ou 5 pour les petites institutions.

4.2.5.3.4 Préservation des bouquets d'archives

Nous avons montré en analysant la couverture des solutions que la préservation des grands bouquets d'archives éditeurs ne peut actuellement être assurée via LOCKSS ou Portico. Nous estimons toutefois ici, à des fins décisionnelles, le surcoût hypothétique de la préservation du bouquet d'archives Springer dans LOCKSS – théoriquement, ce surcoût serait nul dans Portico et dans CLOCKSS, puisque le montant de la souscription annuelle n'est pas fonction de la quantité de contenus archivés.

Ne disposant pas du volume réel de données des bouquets d'archives éditeurs, nous l'avons estimé sur base du nombre d'années-titres dans le bouquet d'archives, multiplié par l'estimation calculée de 108 Mb par année-titre de collection numérisée.

Bouquets d'archives	Nb titres	Nb années-titre	Soit Nb moyen d'années par titre	Soit en Tb*
Springer OJA ²⁴⁹	1.015	25.575	25	2,7 Tb
SAGE Deep Backfile Package ²⁵⁰	463	9.225	19	1 Tb
OUP Archives	203	5.075 (estimation sur base de 25 années de collection par titre)	25 (hypothèse)	0,5 Tb

Tableau 28 - Estimation du volume de données de quelques grands bouquets d'archives. [*hypothèse : 1 année-titre numérisée = 108 Mb]

Si l'on conserve l'hypothèse de coûts de stockage constants et avec la configuration de base (100 €/Tb, durée de vie de 4 ans) identifiée plus haut, le surcoût annuel pour le stockage de 2,7 Tb, soit

²⁴⁸ Entretien avec V. Reich, août 2013.

²⁴⁹ L'analyse a été effectuée sur la base de la liste complète des archives du site de l'éditeur en octobre 2013, incluant les archives complètes 2005-2012. Les institutions de la BICfB ont acheté en deux temps l'ensemble des titres jusque 2011, ce qui donne 995 titres sur les 1.015 incluant 2012, soit 20 titres de moins.

²⁵⁰ L'analyse a été effectuée sur la base de la liste complète des archives du site de l'éditeur en octobre 2013.

l'équivalent des archives Springer, dans LOCKSS, à partir de la deuxième année, est de moins de 150 € par LOCKSS box.

En termes financiers, si le volume estimé des archives est correct, le coût de l'archivage dans LOCKSS des bouquets d'archives dont la couverture rétrospective serait assurée dans LOCKSS serait faible. Le paiement de fees spécifiques à l'éditeur pour un accès « post-abonnement » ou « post-achat » sur sa plateforme pourrait toutefois se justifier dans le cas d'un usage massif des archives, si les institutions ne souhaitent pas assumer les coûts d'un serveur d'accès plus performant que la LOCKSS Box de base.

4.2.5.4 *Evaluation de l'investissement sur les titres courants*

Le rapport Keeping Research Data Safe 2 (Beagrie, Lavoie, & Woollard, 2010) insiste avec raison sur l'importance, en parallèle à celle des coûts, de l'analyse des bénéfices, tout en soulignant ceux-ci, dans le cadre de projet de préservation, sont difficilement mesurables. Dans le cas de la préservation des e-journaux, la mesure du bénéfice s'exprime surtout en termes d'accès et en termes d'assurance de l'investissement effectué dans les acquisitions documentaires.

“A relatively small percentage of the overall budget could act as insurance against unforeseen circumstances and undue dependence on publishers to fulfil a role they have not traditionally taken”²⁵¹.

Le taux de cette « assurance préservation » devrait être estimé sur base du coût de souscription des e-journaux effectivement préservés dans les différentes solutions. En l'absence de données précises à ce sujet, nous avons tenté d'estimer ce coût, pour chaque institution, sur base du coût d'acquisition total des ressources électroniques (2011), duquel a été déduit le coût des bases de données acquises en consortium (2013). Le coût des *big deals* hors *core collection* devrait également être déduit pour arriver à un taux plus représentatif, puisque les titres non *core collection* ne sont pas archivés.

Sur base de cette simulation, et en termes de bénéfices, le pourcentage du budget annuel en acquisitions documentaires considéré pour les grandes institutions semble raisonnable (autour de 0,4 à 0,8% pour LOCKSS, autour de 0,4% pour Portico).

Pour les institutions non complètes, par contre, l'investissement en termes de pourcentage du budget est nettement plus considérable. Il faut toutefois rappeler que la directrice de LOCKSS s'est montrée très ouverte à la négociation et que les prix de souscriptions pourraient être divisés de moitié, voire plus, pour les petites institutions. L'approche de scénarios de collaboration entre les institutions pour l'archivage pérenne pourrait aussi être une solution de réduction des coûts (cf. chapitre 7).

²⁵¹ L'une des conclusions du rapport de Beagrie, Lavoie, & Woollard, 2010 est que « les coûts de l'archivage (stockage, *preservation planning and actions*) représentent relativement une très faible proportion des coûts généraux de ces ressources, et sont en particulier très faibles comparés aux coûts d'acquisition / *ingest* ».

Retour sur investissement	Institutions		LOCKSS ²⁵²			Portico			CLOCKSS	
	Budget acquisitions documentaires (hors databases acquises en consortium) ²⁵³	en Nb estimé de titres préservés	% annuel sur le budget (année 1 – inclut les coûts initiaux)	coût sur titre sur 20 ans	Coût au titre sur 20 ans	Nb estimé de titres préservés	% annuel sur le budget	coût sur titre sur 20 ans	Coût au titre sur 20 ans	% du coût annuel sur le budget en acquisitions documentaires
ULg	€ 1.884.865,20	2.600	0,6%	€ 88,79	3.500	0,4%	€ 46,52	0,1%		
ULB	€ 2.630.623,56	2.700	0,4%	€ 85,50	3.635	0,3%	€ 44,80	0,1%		
UCL	€ 1.504.914,41	3.342	0,8%	€ 69,07	4.499	0,4%	€ 32,12	0,1%		
UNamur	€ 635.034,33	1.857	1,2%	€ 76,45	2.500	0,8%	€ 41,83	0,1%		
UMONS	€ 143.659,54	1.238	5,3%	€ 114,67	1.667	2,7%	€ 50,77	0,2%		
USL-B	€ 85.059,53	1.238	8,2%	€ 104,10	1.667	3,6%	€ 39,78	0,4%		

Tableau 29 - Evaluation de l'investissement dans l'archivage pérenne tiers (LOCKSS, CLOCKSS, Portico) par institution

Un autre point de vue sur l'investissement est d'examiner, pour les titres de périodiques, les frais liés à sa préservation à long terme – qui peuvent être mis en relation avec leur prix d'achat²⁵⁴. Nous avons évalué le coût cumulé de préservation par titre sur 20 ans pour les différentes solutions, sur base des paramètres définis plus haut, pour 17 années de *holdings* la première année²⁵⁵. Ici aussi, des restrictions importantes s'imposent sur la validité de la projection, en particulier parce que le nombre de titres pris en compte a été estimé pour toutes les institutions sur base du portefeuille courant de l'ULg couvert par les solutions, et parce qu'il ne prend pas en compte une augmentation possible de la couverture des solutions, qui pourrait réduire sensiblement le coût au titre. Ceci étant dit, on arrive à une estimation d'un coût d'archivage pérenne de 40-50 € dans le cas de Portico, et du double dans LOCKSS.

4.2.5.5 Conclusion

L'estimation des coûts à charge des institutions pour l'archivage dans chacune des trois solutions (LOCKSS, CLOCKSS, Portico) a montré, malgré ses limites inhérentes suite à la difficulté d'évaluer certains paramètres, que les coûts de souscription aux solutions constituent l'investissement majeur pour l'archivage via des solutions tierces. L'archivage via LOCKSS nécessite un investissement

²⁵² Les prix US estimés ont été estimés sur base des désignations des catégories JSTOR et des montants de souscription à Portico, qui sont équivalents pour les grandes institutions à ceux proposés par LOCKSS. Les équivalents théoriques des coûts de souscription selon la classification JSTOR pour les institutions de la BICfB communiqués par mail par V. Reich en août 2013 étaient très désavantageux, sans doute suite à des erreurs de catégorisation ou à une mauvaise adaptation de la classification JSTOR à la situation des institutions de la BICfB. LOCKSS a donc proposé d'appliquer le rapport budget acquisitions documentaires / coût de la souscription de l'ULB, à laquelle elle a déjà fait une proposition financière en 2012, à l'ensemble des institutions de la BICfB. Ces équivalents calculés sont repris dans le tableau 25 sous « prix proposés par LOCKSS », et sont extrêmement avantageux par rapport aux prix US estimés. Ils n'ont pas été validés par LOCKSS en tant que tels.

²⁵³ Le budget global en acquisitions documentaires électroniques pour chaque institution est issu des statistiques CIUF, 2011. De ce montant global, faute de données plus précises, on a soustrait le coût 2013 des bases de données souscrites en consortium pour chaque institution.

²⁵⁴ Démarche initiée pour le monde des bibliothèques en 1994 par A. Stephen pour la British Library (Stephen, 1994), et reprise ensuite dans différents modèles d'analyse des coûts de la préservation pérenne (en particulier (« Cost Model for Digital Preservation »).

²⁵⁵ Le nombre de titres préservés a été estimé sur base de la comparaison du portefeuille ULg, de manière proportionnelle pour les différentes institutions. Cette estimation pourrait se révéler trop faible ou trop élevée lors de la comparaison réelle des portefeuilles des différentes institutions avec les titres archivés dans les solutions.

spécifique à la gestion des données en local, mais qui semble limité dans la configuration recommandée pour les e-journaux. La projection des coûts de stockage est soumise à plusieurs facteurs d'incertitude, mais étant donné le volume de données considérées (4 Tb pour les e-journaux natifs électroniques, quelques Tb supplémentaires pour les bouquets d'archives), le risque semble raisonnable.

La souscription à CLOCKSS est très faible par rapport à la souscription aux autres solutions, mais il faut rappeler qu'elle n'offre pas de garantie d'accès post-abonnement, et ne peut intervenir qu'à titre de complément. Elle est d'ailleurs facultative puisqu'il s'agit d'un soutien volontaire, ne modifiant aucunement les avantages en termes d'accès ou de choix de préservation.

A coûts de souscription et périmètre constants, Portico est la solution financièrement la plus intéressante²⁵⁶ pour la BICfB, puisque les coûts de souscription sont équivalents à ceux de LOCKSS. Toutefois, cet avantage est à estimer en fonction des autres projets d'archivage pérenne au sein des institutions. En effet, la souscription à la LOCKSS Alliance inclut également l'archivage des e-books disponibles et permet également de mettre en place un PLN et de répondre à d'autres besoins. Les coûts de souscription et en ressources humaines (en particulier ceux liés à l'initialisation et à l'apprentissage) peuvent être dans ce cas répartis sur différents projets.

Dans tous les cas, les coûts de souscription étant indépendants du volume de données archivées et/ou accessibles, les économies d'échelle sont importantes, et ne sont pas entièrement compensées par les tarifs dégressifs des solutions. En termes de coûts par titre archivé, Portico est donc beaucoup plus avantageux que LOCKSS, mais cet avantage est relatif si l'on considère l'articulation avec LOCKSS d'autres solutions d'accès et d'archivage, gratuites ou à très faibles coûts, que sont CLOCKSS et l'accès post-abonnement via la plateforme éditeur.

En termes d'investissement annuel, le coût de la préservation pérenne via une solution tierce représente entre 0,4 et 0,8% du budget en acquisitions documentaires électroniques hors bases de données souscrites en consortium pour les grandes institutions, mais jusqu'à 4,5% à un peu plus de 8% pour les petites institutions dans les cas les plus défavorables. Ceci plaide en faveur soit d'une négociation des coûts de souscription (LOCKSS y est visiblement sensible), soit d'une solution de collaboration entre grandes et plus petites institutions au sein de la BICfB.

Les économies d'échelle de l'archivage via les solutions tierces dédiées plaident également en faveur de l'achat de bouquets d'archives si la couverture rétrospective des solutions tierces dédiées était étendue.

La combinaison de l'archivage pérenne via LOCKSS et CLOCKSS et l'accès post-abonnement sur la plateforme de l'éditeur n'entraîne pas – ou peu si les institutions souscrivent à CLOCKSS – de coûts supplémentaires pour les institutions, et se justifie économiquement. Par contre, si la BICfB veut répartir la prise de risques en investissant dans deux solutions indépendantes (LOCKSS et Portico), le saut d'investissement nécessaire est majeur, et le bénéfice réel à bien examiner, puisque la multiplication des coûts atteint presque 2 pour un gain de couverture du portefeuille d'environ 10% (cf. point 4.2.3).

Il reste à souligner encore une fois que la projection des coûts effectuée est très incertaine pour les différentes solutions, puisque, outre de facteurs économiques et matériels difficiles à estimer sur le long terme, elle dépend de la stabilité du *business model* des solutions.

²⁵⁶ C'est certainement la raison pour laquelle le nombre de bibliothèques participantes atteint 820 en 2012 (alors que le nombre de bibliothèques participant à LOCKSS atteint entre 100 et 200).

4.2.6 Conclusion et recommandations

Les solutions tierces (LOCKSS, CLOCKSS, Portico) répondent à des exigences de la préservation pérenne des e-journaux : partage des responsabilités et des coûts et mutualisation des ressources, afin de minimiser les risques. L'analyse des garanties de préservation pérenne techniques et organisationnelles, des coûts, de l'accès et de la couverture de chacune de ces solutions soutient l'idée que ces solutions sont désormais suffisamment matures pour prendre le risque d'y investir : elles répondent aux différentes exigences reconnues de la préservation pérenne, prennent en compte les spécificités d'accès propres aux journaux électroniques, et permettent de couvrir le portefeuille des e-journaux payants courants des institutions à plus de 80%, pour des coûts relativement raisonnables par rapport aux dépenses investies chaque année dans la souscription aux ressources électroniques (entre 0,1 et 0,8% du portefeuille des grandes institutions selon la taille de l'institution et la solution choisie).

Il faut pourtant garder à l'esprit que ces solutions n'offrent pas de garantie totale ni de solution définitive²⁵⁷. Étant donné la nature des différents intervenants (éditeurs, bibliothèques, organisations tierces), ces solutions constituent toujours un équilibre précaire et à rediscuter (décisions ponctuelles d'accès, définition du *business model*...). Les incertitudes concernent aussi les coûts à long terme (en particulier pour l'ingestion, la migration et le stockage physique), même si l'évolution de la problématique au niveau international peut avoir un impact positif (développements partagés et collaborations, définitions de standards). La meilleure illustration de ces problématiques est la question de l'extension de la couverture des solutions. Dans quelle mesure est-elle une question de coûts financiers (ingestion des contenus), de droits (perte de contrôle par les éditeurs sur l'accès), d'investissement technique (compatibilité des plateformes) ou de communication (connaissance des enjeux et des contraintes) ? Quelle latitude existe encore pour la développer ? L'ingestion des contenus de bouquets d'archives pour les éditeurs participants n'est-elle qu'une question de temps ?

Nous avons donc identifié, pour chaque solution, et pour chaque aspect examiné (garanties d'archivage pérenne, accès, couverture, contrôle et coûts) une série de points d'action recommandés.

Le graphique ci-dessous (fig. 16) présente le résultat de l'évaluation globale des solutions pour l'accès pérenne aux e-journaux en fonction des différents aspects considérés. L'aspect « contrôle » montre que le choix de l'une ou l'autre solution n'est pas anodin puisqu'il suppose, même si les coûts et la couverture peuvent entrer en ligne de compte, un choix philosophique : les bibliothèques préfèrent-elles payer une sorte d'assurance externe (Portico) ou s'investir dans la préservation (LOCKSS) ?

²⁵⁷ B. Reilly, du CRL, écrivait dans Reilly, B. (2010). Portico security. *Library Journal*, 135(9), 12-12., que "LJ's annual Periodicals Price Survey ("Seeking the New Normal," by Kittie Henderson and Stephen Bosch, LJ 4/15/10, p. 36-40) infers from the certification of Portico by the Center for Research Libraries that with "this level of security, libraries can discard duplicate print journals with minimal risk" et insistait sur la prise en compte d'autres facteurs. D'autres facteurs tels le contrôle et les garanties et délais d'accès sont effectivement primordiaux et empêchent, à notre avis, que Portico ne présente des garanties suffisantes pour éliminer les collections papier.

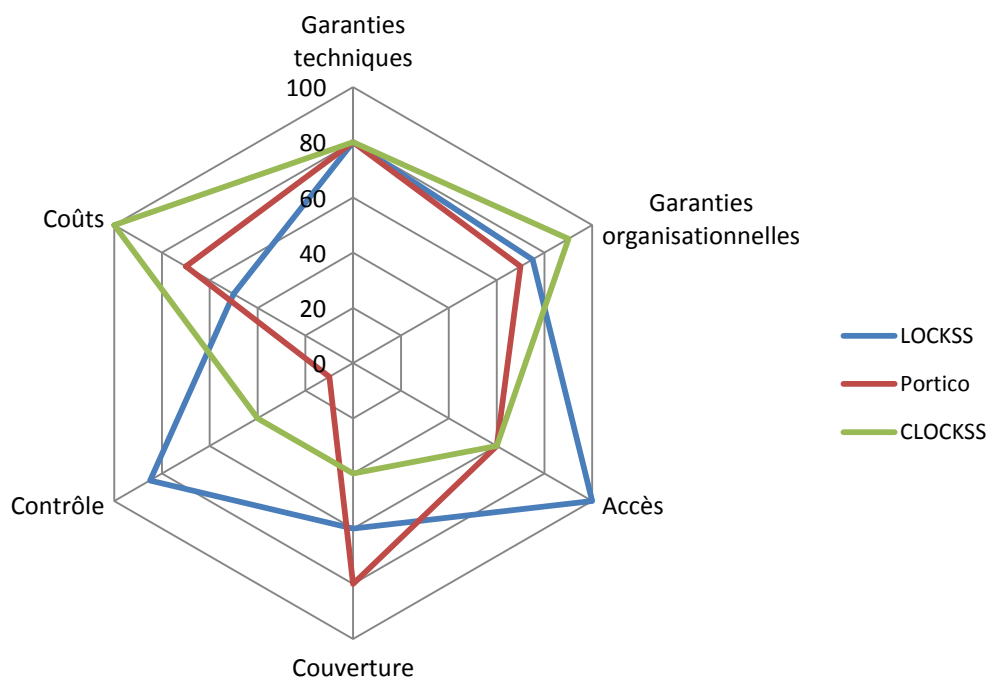


Figure 13 - radar de l'évaluation de LOCKSS, CLOCKSS et Portico

On voit aussi qu'aucune solution ne répond totalement aux besoins de la BICfB en termes de couverture (même si Portico présente une meilleure couverture globale du portefeuille), et qu'elles ont des philosophies fort différentes en termes d'accès.

La solution la plus économique et offrant des garanties d'archivage pérenne importante est CLOCKSS. Toutefois, elle présente des limites en termes de couverture et ne gère pas l'accès post-abonnement ni l'accès en cas de transfert de titres.

LOCKSS et Portico ont une couverture plus importante du portefeuille ; Portico présente la meilleure couverture en nombre absolu, mais le fait que Portico soit une solution d'*outsourcing* complet, présentant des limites fortes en termes d'accès (notamment le fait que l'arrêt de la souscription à Portico entraîne la perte d'accès aux titres archivés) en fait une solution trop risquée à notre avis pour pouvoir garantir un accès à long terme.

LOCKSS présente par rapport à Portico l'avantage majeur d'assurer aux institutions la possession des fichiers de données, ce qui, selon nous, justifie l'investissement plus élevé par rapport à la souscription à Portico. C'est la seule manière d'offrir une alternative suffisante par rapport à l'archivage papier, et de garantir un accès immédiat aux titres archivés des éditeurs ne pouvant pas gérer l'accès post-abonnement sur leur plateforme ou en cas de transfert de titres.

Ces différents éléments, complémentaires, plaident, selon nous, en faveur d'une solution mixte, déjà évoquée au cours de l'analyse, articulant les différentes solutions tierces d'accès pérenne en fonction de leurs points forts, et prenant en compte les éléments contractuels, que nous détaillons dans le chapitre 7.

5 Les initiatives nationales ou régionales

La présente étude s'est concentrée jusqu'ici sur l'évaluation des solutions d'accès pérenne dédiées spécifiquement aux e-journaux les plus importantes et avec une dimension collaborative internationale (CLOCKSS, LOCKSS, Portico). Nous avons montré que ces solutions, bien qu'encore perfectibles, sont suffisamment matures pour y investir et permettent de couvrir la plus large partie du portefeuille des institutions.

Avant d'aborder les scénarios possibles pour la BICfB, il nous semble intéressant de mentionner brièvement quelques initiatives spécifiques mises en place au niveau national ou régional en matière d'accès pérenne²⁵⁸. Il s'agit dans certains cas de systèmes centralisés, développés sous la responsabilité d'une ou plusieurs bibliothèques nationales, parfois en utilisant des outils commerciaux, soit de réseaux distribués. Il s'agit dans d'autres cas de volonté de mise en place d'une stratégie nationale, dont l'implémentation est encore à l'étude (Allemagne, Suisse). Ces initiatives peuvent constituer des sources de partage d'expérience, voire, éventuellement, des partenariats possibles.

5.1 Les licences nationales

Plusieurs pays (notamment l'Angleterre, l'Allemagne, la France, les Pays-Bas, l'Espagne, la Suisse, le Brésil²⁵⁹) ont mis en place des licences nationales, c'est-à-dire l'acquisition de contenus pour l'ensemble des institutions du pays, dans un cadre légal commun. Ces projets ont l'avantage de donner du poids aux institutions dans la négociation avec les éditeurs, et les licences nationales contiennent généralement des clauses d'accès pérenne (cf. par exemple NESLi2, mis au point par JISC Collections²⁶⁰). Ces licences nationales en tant que telles ne constituent pas des garanties d'accès pérenne, puisqu'il s'agit de mutualisation des efforts sur la question des droits – comme ce qui est réalisé lors des négociations de la BICfB.

Certains projets récents de licences nationales consacrés à l'achat d'archives incluent toutefois, outre ces aspects, la mise en place, en parallèle, d'un serveur de contenu avec garanties de préservation pérenne. On présente ici le cas de la France, qui est intéressant parce qu'il distingue les bouquets d'archives des autres contenus.

La France a lancé en 2010 à l'échelle nationale, via le Ministère de l'Enseignement supérieur et de la recherche, un projet destiné à acquérir des ressources via des licences nationales, financées directement par le gouvernement (<http://www.licencesnationales.fr>). En complément, le gouvernement a signé en 2012 le financement du projet ISTEEX (Initiative d'excellence de l'Information Scientifique et Technique) (<http://www.istex.fr>), dont l'objectif est « d'offrir, à l'ensemble de la communauté de l'enseignement supérieur et de la recherche, un accès en ligne aux collections rétrospectives de la littérature scientifique dans toutes les disciplines en engageant une

²⁵⁸ Il s'agit ici d'une sélection illustrative. Pour un panorama plus exhaustif, voir par exemple la publication de l'IFLA, Verheul, I. (2006). Networking for digital preservation : current practice in 15 national libraries. Munich: K.G. Saur. Consulté à l'adresse <http://www.ifla.org/files/hq/publications/ifla-publications-series-119.pdf>, ou Carbone, P. (2010). Coûts, bénéfices et contraintes de la mutualisation des ressources électroniques : éléments de comparaison internationale et propositions (No. 2010-012) (p. 91). Consulté à l'adresse http://media.enseignementsup-recherche.gouv.fr/file/2009/69/8/rapport_MRE_diffusion_159698.pdf, ou encore, plus récemment, mais dans une autre perspective, Aligning National Approaches to Digital Preservation, Skinner, & Educopia Institute, 2012.

²⁵⁹ Pour ces trois derniers pays, la source est Carbone, P. (2010), p. 37.

²⁶⁰ <https://www.jisc-collections.ac.uk/nesli2/>

politique nationale d'acquisition massive de documentation : archives de revues, bases de données, corpus de textes... »²⁶¹.

La première phase de sélection et d'acquisition des ressources a été effectuée en 2012-2013 ; la seconde phase, qui devrait être terminée en mai 2014, consiste en la création de la plateforme ISTEEX, qui offrira un accès centralisé à l'ensemble de ces ressources, hébergé sur le territoire français, assorti de la préservation pérenne de ces données. Un financement de 60 millions d'euros sur 3 ans a été accordé par le gouvernement, dont 55 millions destinés à couvrir l'achat des licences, et 5 millions destinés au développement du portail.

Actuellement, ont été acquis dans le cadre de ce projet les archives de revues et les collections d'e-books Springer, les Early English Books Online et les Eighteen Century Collections Online de ProQuest, les Dictionnaires et Grammaires classiques Garnier Numérique.

L'archivage pérenne des données est assuré par le CNRS avec l'appui du CINES (Centre informatique national de l'enseignement supérieur)²⁶². Le CINES a pour mission depuis 2004 « la mise en place d'un service pour l'archivage à long terme des données sous forme numérique destiné à la préservation à long terme du patrimoine scientifique ». Le CINES utilise depuis 2008 la plateforme PAC v.2, basée sur le logiciel Arcsys accompagné d'un module interne développé en Java et d'une base de données MySQL. Cette plateforme est destinée à stocker de larges volumes de données (plus de 40 Tb). Le CINES envisage la migration logique des données, et sélectionne rigoureusement les formats d'archivage, mais aucune migration n'a encore pu être réalisée sur les données réelles²⁶³.

5.2 Les initiatives des bibliothèques nationales liées au dépôt légal et les services de tiers archivage

La majorité des bibliothèques nationales européennes (on peut citer la France, l'Allemagne, l'Angleterre, les Pays-Bas, l'Italie, la Suède, le Danemark, la Suisse, la Finlande, la Tchéquie par exemple) travaillent activement sur la thématique de la préservation des contenus électroniques dans le cadre de leur mission de dépôt légal. Elles ont mis en place des systèmes d'archivage pérenne de contenus électroniques parfois depuis plusieurs années, et collaborent aux réseaux de partage internationaux dédiés à la thématique.

Plusieurs bibliothèques nationales tirent par ailleurs profit de leur expertise en matière de préservation numérique et de leur infrastructure capable d'héberger d'énormes quantités de données pour proposer un service de tiers archivage à des institutions extérieures.

Actuellement (2012-2014), la DNB est engagée dans la suite du projet « **Digital Preservation for Libraries – Engaged** »²⁶⁴, qui a pour objectif de proposer un panel de services de préservation à long terme des données numériques à des tiers. En juin 2013, le service a été présenté lors des nestor Tagungen à Munich par le project manager Karlheinz Schmitt ; les coûts sont de 1,49 €/Gb/an pour la location de serveurs, auquel il faut ajouter le coût du service en tant que tel, calculé au cas par cas sur base d'un modèle de coûts.

²⁶¹ Le projet est soutenu par le Centre National de la Recherche Scientifique (CNRS), l'Agence Bibliographique de l'Enseignement Supérieur (ABES), le Consortium Universitaire de Publications Numériques (Couperin1), et l'Université de Lorraine agissant pour le compte de la Conférence des Présidents d'Université (CPU).

²⁶² <http://www.cines.fr/>

²⁶³ d'après les informations disponibles sur le site web du CINES à la date de rédaction de ce rapport.

²⁶⁴

<http://www.dnb.de/EN/Wir/Projekte/Laufend/dp4lipEngaged.html;jsessionid=AD3AF6CF13571FB46FEAC1B345CBDEA5.prod-worker3>

En 2013, la BnF a décidé, comme la DNB, d'ouvrir son **système d'archivage numérique SPAR** (Système de Préservation et d'Archivage Réparti), lancé en 2010, à d'autres organisations, qui peuvent ainsi bénéficier de l'expertise et des infrastructures de la BnF, via le service de « **tiers archivage** »²⁶⁵. Le service propose un archivage à long terme sécurisé, avec vérification d'intégrité et d'authenticité sur des serveurs avec redondance (2 sites distincts en France), d'une capacité de 16 Po actuellement (susceptible d'augmenter). Ses aspects techniques, historiques et fonctionnels sont décrits en détail sur le site de la BnF²⁶⁶. Le service inclut la consultation possible des contenus via l'« espace Coopération » (extranet destiné aux partenaires de la BnF pour le partage de documents et le suivi des projets) de la BnF. La BnF a mis au point une offre tarifaire basée sur le volume de données, le nombre de copies préservées, le niveau de service attendu et la durée du contrat. Pour une durée de 5 ans, le tarif annuel au Tb, pour une ingestion sur mesure, est de 2.680 € pour 2 à 5 Tb²⁶⁷, soit 6.700 € pour 2,5 Tb.

Au Danemark, le *Bitmagasin*²⁶⁸, développé conjointement par la Bibliothèque Royale, les Archives de l'Etat et l'Université d'Etat d'Aarhus et qui devrait être disponible très bientôt²⁶⁹, proposera un service de préservation d'octets partagé aux autres institutions culturelles de conservation au Danemark.

Si ces initiatives sont fondamentales en termes de développement des connaissances et des technologies en matière de préservation numérique pérenne, elles n'ont pas vocation à archiver spécifiquement les collections d'e-journaux des bibliothèques universitaires, et présentent, dans cette perspective, certaines limites.

Le dépôt légal électronique, s'il a pour but de garantir effectivement la préservation à très long terme des publications électroniques, et prend en compte l'obtention des données auprès des éditeurs, souffre de limitations en termes de couverture et d'accès :

- la couverture du dépôt légal ne correspond pas à celle des collections utiles à la recherche universitaire : elle est normalement limitée aux publications nationales (même si la frontière est de plus en plus difficile à définir) ; par ailleurs, les types de publications concernées par le dépôt légal ne sont pas les mêmes dans les différents pays, créant des « gaps » dans l'archivage de l'héritage numérique mondial²⁷⁰ ;
- les bibliothèques nationales n'ont pas les mêmes priorités en termes d'accès que les bibliothèques universitaires²⁷¹ : elles ne prennent pas en compte l'accès post-abonnement, et proposent généralement un accès limité à la consultation sur place.

²⁶⁵ Voir, pour une présentation de la philosophie et des aspects fonctionnels du projet : Fauduet, L., & Peyrard, S. (2013). Destination: Shared Repository. The National Library of France's Journey to Third-Party Archiving. In Proceedings of the 10th International Conference on Preservation of Digital Objects (iPRES). Lisbonne. Consulté à l'adresse

http://purl.pt/24107/1/iPres2013_PDF/Destination%20Shared%20Repository%20The%20National%20Library%20of%20France%27s%20Journey%20to%20Third-Party%20Archiving.pdf.

²⁶⁶ http://www.bnf.fr/fr/professionnels/spar_systeme_preservation_numerique.html

Une bibliographie est disponible sur http://www.bnf.fr/documents/preservation_numerique_bibliographie.pdf (consulté le 4 février 2014).

²⁶⁷ Voir les tarifs sur http://www.bnf.fr/documents/archivage_num_tarifs.pdf (septembre 2013).

²⁶⁸ <http://digitalbevaring.dk/det-nationale-bitmagasin/>

²⁶⁹ La page web annonçait, le 5 février 2014, que le Bitmagasin devrait être disponible en 2013.

²⁷⁰ Voir Aligning National Approaches to Digital Preservation, Skinner, & Educopia Institute, 2012, p. 47 et ss. pour davantage de précisions et des propositions concernant cette problématique.

²⁷¹ Voir aussi le point 2.1. et Aligning National Approaches to Digital Preservation, Skinner & Educopia Institute, 2012, p. 43.

Le projet e-Depot (cf. supra 3.3.4), de la Bibliothèque nationale des Pays-Bas, est un cas particulier, puisque, conçu à l'origine pour préserver les données du dépôt légal numérique, il a été très rapidement étendu aux éditeurs internationaux et fonctionne globalement comme une *dark archive* avec des *triggers events* semblables à ceux de CLOCKSS.

Les services de tiers archivage présentent l'avantage de mettre au service d'autres institutions, au niveau national, des compétences, des technologies et des infrastructures, pour des coûts finalement assez comparables aux solutions LOCKSS et Portico. Malheureusement, ces services ne sont pas spécifiquement dédiés aux e-journaux et ne proposent pas de mutualisation des ressources aussi avancées que celles des solutions dédiées : ils ne prennent pas en compte les aspects de gestion des droits sur les contenus (négociations avec les éditeurs) ni, généralement, l'obtention des données et leur traitement avant ingestion dans le système, ni et la gestion des accès à ces contenus dans le cadre des systèmes de gestion de l'information des bibliothèques universitaires, qui restent à charge des institutions. Ces services ne sont donc pas une alternative aux solutions dédiées, mais pourraient être envisagés pour les contenus spécifiques non couverts par les solutions dédiées.

En Belgique, nous n'avons pas connaissance à l'heure d'initiative développée proche de ce modèle. La KBR et les archives du Royaume, si elles travaillent sur la problématique, n'ont pas encore développé à notre connaissance un système de préservation numérique global offrant ce service. La Bibliothèque royale utilise actuellement DigiTools d'ExLibris pour l'archivage pérenne des journaux belges numérisés et des contenus du dépôt légal électronique, mais s'orientera vraisemblablement dans le futur plutôt vers une solution développée en interne²⁷².

5.3 Les réseaux de préservation nationaux et les initiatives nationales liées à LOCKSS

L'Allemagne et l'Angleterre ont investigué fort tôt la piste de la préservation partagée et mis en place deux réseaux concrets LOCKSS : LUKII et la UK LOCKSS Alliance.

En 2008, suite à un projet pilote mis en place par le JISC était lancée la **UK LOCKSS Alliance**²⁷³, un réseau LOCKSS de bibliothèques universitaires anglaises participant au GLN pour préserver les e-journaux mais assurant une gouvernance spécifique au niveau national via EDINA, fournissant une aide à l'installation et à la gestion des LOCKSS Boxes, ainsi qu'une communauté de négociation et de sélection de titres. La UK LOCKSS Alliance compte aujourd'hui 15 institutions. Elle publie des études de cas et des rapports réguliers et dispose d'une importante compétence technique. La UK LOCKSS Alliance est un exemple atypique d'implémentation de gouvernance LOCKSS, dont la BICfB pourrait s'inspirer.

Le projet allemand **LUKII (LOCKSS und KOPAL Infrastruktur und Interoperabilität, 2010-2012)**²⁷⁴ visait à combiner les forces de LOCKSS (archivage des bits via un réseau peer-to-peer) et de KoLiBRI (archivage logique, migration préventive), logiciel Open Source développé entre 2004 et 2007 dans le cadre du projet KOPAL, qui fonctionnait à l'origine avec la solution propriétaire DIAS (Digital Information Archiving System) d'IBM (sous licence) pour le stockage et la gestion des données. La

²⁷² Nous remercions Frédéric Lemmers (KBR) de nous avoir communiqué cette information.

²⁷³ <http://www.lockssalliance.ac.uk/>

²⁷⁴ <http://www.lukii.hu-berlin.de/>

Le projet, financé par le ministère de l'éducation et de la recherche allemand et porté par la DNB, la Niedersaechsische Staats- und Universitaetsbibliothek Goettingen (SUB), la Gesellschaft fuer wissenschaftliche Datenverarbeitung mbH Goettingen (GWDG), en collaboration avec IBM Deutschland, s'est déroulé entre le 01-11-2009 et le 31-08-2012. Voir Ostrowski, F., & Sieber, J. (2012). LOCKSS im Rahmen des LuKII-Projekts. Technische Erweiterungen für Datenaustausch, Metadatenverwaltung und die prophylaktische Formatmigration. *Bibliothek Forschung und Praxis*, 36(1), 135-142. doi:10.1515/bfp-2012-0017. doi:10.1515/bfp-2012-0017 pour la présentation générale et le fonctionnement technique de la solution.

phase de tests intensifs du prototype, basée sur le *harvesting* des données des dépôts institutionnels allemands, a commencé en 2012.

Dans le cadre du projet LUKII, un PLN de 9 LOCKSS Boxes a été installé en Allemagne (HU Berlin, ULB Münster, HBZ, DNB, SLUB Dresden, KIT, UB Stuttgart, Uni Konstanz, BSB²⁷⁵). L'Allemagne a dans ce cadre développé des plug-ins spécifiques et développé un centre de compétences LOCKSS²⁷⁶ à la HU Berlin. Dans la même ligne, l'Université Humbolt (HU) de Berlin participe depuis fin 2011 à CLOCKSS en tant que « nœud » d'archivage et membre du comité directeur.

Plusieurs universités italiennes sont très impliquées dans la LOCKSS Alliance ou dans CLOCKSS, et il y a un intérêt exprimé de l'Università del Sacre Cuore (Milan) à constituer un Réseau Privé LOCKSS afin de pouvoir préserver les publications scientifiques italiennes ou présentant un intérêt particulier pour la recherche en Italie²⁷⁷.

5.4 Les stratégies nationales ou consortiales à l'étude

Nous mentionnons ici deux initiatives en cours dont nous avons connaissance, sans aucun souci d'exhaustivité. Nous n'avons pas identifié via les informations disponibles dans la littérature d'études de ce type actuellement en cours au niveau consorcial en Italie en France ou aux Pays-Bas, mais ces pays sont très impliqués dans la préservation pérenne, soit via des projets nationaux déjà mentionnés, soit via la participation à CLOCKSS, LOCKSS et Portico²⁷⁸. Le JISC a par ailleurs déjà réalisé plusieurs projets et études liées à la préservation des e-journaux, mentionnées régulièrement dans ce rapport, et qui constituent des fondements incontournables en la matière.

5.4.1 Suisse

Le portail e-lib (<http://www.e-lib.ch>) a pour objectif de créer un portail national d'accès à l'information scientifique.

Le consortium des bibliothèques universitaires (CBU) suisses a publié en 2005 une étude conséquente sur la question de l'e-archiving²⁷⁹. Entre 2005 et 2007, il a mené le projet « E-Archiving », qui avait pour objectif de déterminer la faisabilité d'une archive locale pour des articles de journaux électroniques ; suivait en 2008-2011 le projet « E-Depot » (sous-projet d'e-lib) qui a testé la préservation de tous les articles de journaux Elsevier publiés entre 1995 et 2004 avec l'outil DigiTools d'Ex-Libris. L'outil n'ayant pas donné satisfaction en matière de traitement d'une telle masse de données, le comité décidait en 2011 de clôturer E-Depot et d'identifier des solutions alternatives²⁸⁰ : LOCKSS, Portico, CLOCKSS. En juin 2012, le consortium prévoyait de mettre à disposition de ses membres le rapport final E-Depot et l'évaluation de LOCKSS et CLOCKSS, et de faire un sondage sur les souhaits de participation de ses membres. En fonction de ce sondage, le

²⁷⁵ Voir la présentation de Fromm, N. (2013, juin 18). Archivierung elektronischer Ressourcen mit LOCKSS. Consulté à l'adresse <http://files.dnb.de/nesstor/veranstaltungen/Praktikertag2013/2013-06-lockss-fromm.pdf>

²⁷⁶ Cf. <http://www.lockss.de/> ; contact Niels Fromm (Competence Center LOCKSS).

²⁷⁷ Voir D'Amato, G., & Mazzocchi, R. (2008). Conservando più, copie rendiamo sicure le nostre collezioni digitali. (Italian). *Biblioteche Oggi*, 26(7), 9-13. Cet intérêt a été confirmé par G. D'Amato (Università del Sacre Cuore) lors d'une conversation téléphonique en été 2012. Les universités italiennes suivantes soutiennent CLOCKSS : Politecnico di Milano, Università Cattolica del Sacro Cuore* (en tant que « nœud »), Università degli Studi di Bergamo, Università degli Studi di Brescia, Università degli Studi di Milano, Università degli Studi di Napoli Federico II, Università degli Studi di Verona, European University Institute (Florence).

²⁷⁸ La participation à Portico est parfois soutenue au niveau consorcial, comme dans le cas de l'Irish Universities Associations, qui a rejoint Portico en 2008.

²⁷⁹ Voir le rapport Etude préliminaire « e-archiving » Consortium des bibliothèques universitaires suisses. (2005). Consulté à l'adresse http://lib.consortium.ch/external_files/Konzeptstudie_F_V1_2_1.pdf.

²⁸⁰ Manz, S. (2012). Ensuring perpetual access to licensed content: introduction to the main issues. Consulté à l'adresse <http://e-collection.library.ethz.ch/view/eth:5613>.

consortium prévoyait de négocier avec LOCKSS et Portico et avec les éditeurs pour les convaincre de participer à ces solutions tierces. Actuellement, il existe une LOCKSS Box en Suisse²⁸¹ et 6 bibliothèques suisses participent à Portico²⁸².

5.4.2 Allemagne

En 2008, l'« Alliance of German Science organisations » définissait l'initiative prioritaire « Digital Information »²⁸³ (2008-2012), avec un volet dédié à la question de la « **National Hosting Strategy** »²⁸⁴. Le rapport «**Ensuring perpetual access**» (Beagrie)²⁸⁵, publié dans sa version finale en février 2010, constituait la base de réflexion pour les actions à mettre en place sur cette question. Ce rapport faisait les recommandations suivantes concernant l'accès post-abonnement : privilégier à court terme l'accès sur les plateformes éditeurs, et explorer deux options pour la préservation à long terme des journaux électroniques : le développement d'une solution nationale spécifique en collaboration avec des partenaires internationaux (qui pourrait se baser sur les solutions existantes d'archivage à long terme), et la participation à Portico. Au vu du succès de cette première phase 2008-2012, une seconde phase (2013-2017) a été définie²⁸⁶. Un **appel à propositions** a été lancé par la Deutsche Forschungsgemeinschaft (DFG) pour la traduction en recommandations concrètes, planning, et répartition des coûts des recommandations théoriques de l'étude préparatoire (Beagrie, 2010), afin de pouvoir implémenter rapidement une stratégie de préservation nationale²⁸⁷. En particulier, le projet devait prendre en compte les différents niveaux de licence (nationale, régionale, locale), et identifier des solutions complémentaires à LOCKSS et/ou Portico pour les contenus non couverts²⁸⁸. Le financement a été attribué à l'automne 2013 à un projet d'une durée de deux ans, impliquant plusieurs Bibliothèques allemandes de recherche. Si un réseau privé LOCKSS est déjà installé en Allemagne dans le cadre du projet LUKII, en janvier 2014, il n'y avait toujours, d'après les informations disponibles sur le site de Portico, aucune institution allemande participante.

5.5 Conclusion

Si les différentes initiatives évoquées ci-dessus ne peuvent répondre directement à la problématique de la préservation pérenne des journaux électronique pour la BICfB, elles sont fondamentales dans le cadre du développement d'outils, de bonnes pratiques et de recommandations internationales, et

²⁸¹ Information communiquée par V. Reich, en août 2013, cf. aussi F. Lecocq, 2013.

²⁸² Federal Institute of Technology, Zurich, University of Basel, University of Fribourg, University of Lausanne, University of St. Gallen, University of Zurich (source: <http://www.portico.org/digital-preservation/who-participates-in-portico/participating-libraries/country/switzerland>).

²⁸³ <http://www.allianzinitiative.de/en>

²⁸⁴ Les autres volets sont les suivants : Open Access, Research data, Virtual Research environment, National Licensing, Legal Frameworks. La priorité pour le volet «*National Hosting strategy*» est définie comme suit : « *Working in a coordinated process at national level, the National Hosting Strategy working group identifies ways in which research organisations can satisfy the demand for the permanent accessibility of digital publications by creating a back-up hosting solution ("safety net").* »

²⁸⁵ Beagrie, Neil. Ensuring perpetual access establishing a federated strategy on perpetual access and hosting of electronic resources for Germany. Edited Final Report, 2010. http://www.allianzinitiative.de/fileadmin/hosting_studie_e.pdf.

²⁸⁶ http://www.allianzinitiative.de/en/core_activities/national_licensing/

²⁸⁷ Deutsche Forschungsgemeinschaft (DFG). « Förderprogramm «Elektronische Publikationen». Ausschreibung «Nationales Hosting elektronischer Ressourcen» ». Consulté le 3 février 2014. http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_elektronische_publikationen_12_0430.pdf.

²⁸⁸ Nous remercions le Dr. Johannes Fournier (Deutsche Forschungsgemeinschaft) et le Dr. Hildegard Schäßler (Bayerische Staatsbibliothek), qui ont aimablement répondu à nos questions sur l'échéance et le périmètre de ce projet.

peuvent permettre d'identifier des collaborations possibles ou des modèles de mise en place de solutions complémentaires à LOCKSS, CLOCKSS ou Portico.

6 L'archivage papier partagé

Le groupe de travail « Conservation partagée des périodiques » (CPP), créé par le CIUF, travaille depuis 2003 à rationaliser et limiter les coûts de l'archivage papier via l'archivage partagé des collections de périodiques imprimés au sein des institutions de la BICfB, en distribuant le rôle de préservation entre les différentes institutions. Ceci permet des gains d'espace notables, réels ou potentiels²⁸⁹ selon que les collections effectivement éliminées pour cause de coûts de stockage ou de manque d'espace, ou qu'elles sont maintenues temporairement au sein des différentes institutions.

Il reste important de rappeler, comme déjà souligné dans l'introduction (cf. 1.2.), que les collections imprimées offrent des garanties de préservation à long terme plus importantes, par nature, que les collections électroniques, puisque d'une part, le support a montré qu'il peut survivre des centaines d'années sans intervention de conservation a priori – au contraire de l'information numérique – et que, d'autre part, le modèle d'acquisition papier entraîne de facto la possession des documents. L'archivage papier partagé a donc fait ses preuves, et reste la solution qui offre le plus grand degré de contrôle puisqu'il n'implique pas nécessairement l'intervention de tiers.

Pour les titres souscrits en électronique du portefeuille BICfB, l'archivage papier partagé pourrait donc être envisagé 1) comme une alternative globale à l'archivage électronique via des solutions tierces s'il présentait une couverture suffisante du portefeuille électronique et se révélait plus avantageux financièrement ou 2) comme alternative spécifique pour certains cas où il constituerait la seule solution pour offrir des garanties suffisantes d'archivage et d'accès.

6.1 Aspects économiques

Nous analysons ici quelques éléments de coûts de la préservation papier en regard des coûts de la préservation via LOCKSS, la plus coûteuse des solutions tierces dédiées, afin d'évaluer si la préservation papier pourrait constituer une alternative viable économiquement à la préservation numérique.

6.1.1 Éléments de coût de la préservation papier

L'estimation des coûts de la préservation papier, comme celle des coûts de la préservation numérique, est complexe car elle dépend de plusieurs facteurs contextuels : taille et organisation de l'institution, mode de stockage (sur campus ou dans un magasin distant, dans des compactus ou étagères ouvertes), importance des usages, et est susceptible, comme la préservation numérique, d'économies d'échelle plus ou moins importantes. Nous n'avons pu, dans le cadre de cette étude, ni procéder à la récolte systématique de l'ensemble des données, ni identifier dans la littérature des données chiffrées récentes utilisables dans le contexte de la BICfB²⁹⁰. Nous posons donc quelques

²⁸⁹ « Entre 2003 et 2010, environ 600 titres de revues d'abstracts ont ainsi fait l'objet de mesures de protection. Le travail d'élimination des doubles sur l'ensemble des bibliothèques participantes a permis de libérer 850 mètres de rayonnage » (BICfB - Proposition de Projet 2012 : Renforcement du projet « Conservation partagée des Périodiques », présentée et approuvée par l'Assemblée générale 2012 de la BICfB). Le budget supplémentaire alloué au projet en 2012 devait permettre le traitement supplémentaire de 2 km de rayonnages.

²⁹⁰ Voir par exemple, pour des calculs de coûts sur la totalité du cycle de vie des documents, soit imprimés, soit imprimés et électroniques : Stephen, 1994 et Schonfeld, R. C., King, D. W., Okerson, A., & Gifford Fenton, E. (2004). The Nonsubscription Side of Periodicals : Changes in Library Operations and Costs between Print and

jalons indicatifs utiles à la comparaison et à l'analyse d'après les données en notre possession, sans viser à une analyse exhaustive ni à des projections précises.

6.1.1.1 Coûts de stockage

Les coûts de stockage de documents papier incluent des coûts d'amortissement de la construction des bâtiments, et des coûts de maintenance (entretien, chauffage, électricité...). En ce qui concerne l'amortissement de la construction, nous considérons les données concernant la construction du nouveau Magasin à livres envisagée en 2012-2013 à l'ULg : 7 millions d'euros pour 45 km linéaires de collections (y compris bâtiment, rayonnages etc.), sans compter le coût du terrain ni le transport des collections. Avec un amortissement sur 25 ans du bâtiment²⁹¹, on arrive à un amortissement annuel du bâtiment de 280.000 €, soit 6,22 € au mètre de rayonnage²⁹².

Coûts de stockage	
Amortissement annuel de la construction sur 25 ans	6,22 €/mètre de rayonnage
Maintenance (électricité, chauffage...)	Non estimée

Tableau 30 - Estimation du coût de stockage annuel au mètre de rayonnage sur base des données concernant la construction d'un Magasin à livres de 45 km linéaires de rayonnages (2012-2013)

6.1.1.2 Coûts d'accès

La conservation de documents papier en réserve, sur campus ou en magasins distants, voire dans une autre institution en cas d'archivage papier partagé, implique des coûts de traitement pour la fourniture de documents (transport du document ou copie ou scan et envoi postal papier ou électronique), tant pour les demandes internes que pour les demandes externes (PIB), en particulier dans le cas de l'archivage papier partagé, où la collection de référence assure la fourniture de toutes les demandes des institutions collaborant.

On peut compter un coût minimal de traitement de 2 € à 3 €, ce qui correspond à 5 minutes de travail au tarif horaire d'un documentaliste (27 € brut) avec éventuellement 15 pages de copie à 0,06 € la page. Ce montant correspond aux tarifs commande de document en réserve ou sur campus distant pratiqués dans les bibliothèques de la BICfB (entre 3 € et 5 € environ) (sauf lorsque le prix facturé correspond uniquement aux frais de copie²⁹³) ; les tarifs de prêt inter-bibliothèque (entre 6 et 8 €/article dans les institutions de la BICfB) sont plus élevés.

Electronic Formats (research report). Washington, D.C.: Council on Library and Information Resources. Consulté à l'adresse <http://www.clir.org/pubs/reports/pub127/pub127.pdf>; Cooper, M. D. (2006). The Costs of Providing Electronic Journal Access And Printed Copies of Journals to University Users. *Library Quarterly*, 76(3), 323-351. Outre le fait que les données sont anciennes, l'importance du contexte géographique et organisationnel les rend difficilement utilisables dans le contexte de la BICfB. Nous n'avons pas pu consulter Genoni, P. (2013). An international review of the development and implementation of shared print storage. *Australian Academic and Research Libraries*, 44(1), 50-66. doi:10.1080/00048623.2013.773867.

²⁹¹ Durée souvent utilisée, mais on pourrait choisir une durée plus longue. Sur 30 ans par exemple, on arrive à un coût de stockage par titre sur 20 ans de 108,89 €.

²⁹² Si le montant devait être emprunté, le coût annuel au mètre serait plus important puisqu'il devrait tenir compte du taux d'intérêt.

²⁹³ 0,06 €/page soit 0,90 € pour un article de 15 pages au tarif de commande articles@BST de l'ULB, soit le prix de la copie uniquement.

6.1.2 Eléments de comparaison des coûts de la préservation papier avec les coûts de la préservation électronique via LOCKSS

La préservation via LOCKSS étant la plus coûteuse des trois solutions tierces envisagée, mais également la seule à fournir des garanties de préservation similaires à l'archivage papier, c'est la seule à avoir été considérée ici. On a tenté de mettre en parallèle les différents éléments de coûts estimés, afin d'évaluer la pertinence économique de la préservation papier comme alternative à la préservation électronique. Nous avons envisagé plusieurs scénarios mettant en évidence les économies d'échelle, l'impact de l'accroissement des collections, et l'impact du nombre d'exemplaires de préservation, mais il reste un certain nombre d'inconnues, en particulier l'effet des économies d'échelle du stockage papier. En outre, la fiabilité des estimations de volumes, qui détermine la validité de la comparaison en matière de stockage, n'est pas assurée. En ce qui concerne la préservation numérique via LOCKSS, l'impact est faible vu la faible part du coût de stockage dans le modèle (nous avons utilisé les étalons calculés sur base du contenu de la LOCKSS Box de Stanford (cf. point 4.2.5.2.3.) ; il est nettement plus important dans le cas de la préservation papier. Nous nous sommes basés sur une estimation de 10 volumes de périodiques (1 volume annuel) au mètre, ce qui correspond à l'étalon utilisé par le CIUF²⁹⁴.

Comparaison des coûts LOCKSS et archivage papier	LOCKSS – coût annuel pour une grande institution sur base de la moyenne du coût sur 20 ans avec remplacement des media tous les 4 ans	Magasin à livres – coût annuel sur base de l'amortissement du bâtiment sur 25 ans (pas de taux d'intérêt)
INGESTION Transfert	/ Inclus	Non inclus (frais de transfert ou de conditionnement des documents) + frais d'acquisition du papier en cas d'archivage papier complémentaire à l'archivage électronique
STOCKAGE	Hors maintenance électricité, réseau... (inclut souscription LOCKSS, stockage, frais en personnel)	Hors maintenance (chauffage, électricité...), hors frais de personnel pour la gestion du bâtiment
Coût annuel/volume, hypothèse de l'occupation maximale (hypothèse théorique)	Coût fixe, quasiment indépendant du volume Coût/an pour 2 à 8 Tb : entre 12.000 € et 12.500 €/an, soit un coût directement dégressif par rapport au volume stocké : ex : <ul style="list-style-type: none"> • Si 50.000 volumes courants (36 Mb) sur 2 Tb : 0,25 €/vol • Si 70.000 volumes d'archives (108 Mb) sur 8 Tb : 0,18 €/vol 	Coût directement proportionnel au volume <ul style="list-style-type: none"> • Si 1 vol = 10 cm : 0,62 €/vol • Si 1 vol = 20 cm : 1,24 €/vol

²⁹⁴ Etalon utilisé dans la récolte des statistiques CIUF, où 1 volume de périodique relié = 1 année de parution, et où le métrage est fixé à 10 volumes reliés par mètre.

	<ul style="list-style-type: none"> • Si 40.000 volumes (60 Mb) sur 4 Tb : 0,31 €/vol. 	
Coût annuel/titre si prévision d'accroissement annuel des collections sur 20 ans (hypothèse : 3.000 titres)	<ul style="list-style-type: none"> • 4 Tb suffisent pour 60.000 vol. (60 Mb), soit environ 4,05 €/titre 	Si 1 vol = 10 cm : prévoir 6 km de stockage pour 20 ans soit 37,32 €/titre avec un amortissement de 6,22 €/m.
ACCES	Inclus	+2 à 3 € par demande au minimum

Tableau 31 - Comparaison des coûts de l'archivage électronique (LOCKSS) et de l'archivage papier (estimation basée sur une nouvelle construction)

En ce qui concerne le stockage, il est difficile de comparer les coûts avec précision. LOCKSS semble dans tous les cas plus avantageux pour le stockage au niveau institutionnel (1 LOCKSS Box, une souscription), même avec l'hypothèse d'occupation maximale sans accroissement de volume. Si l'on envisage l'installation d'une LOCKSS Box dans chaque institution de la BICfB (soit une multiplication des coûts par 4,5 à 5,5 selon que la souscription à LOCKSS est négociée ou non), le coût au titre (avec 6 exemplaires préservés au sein de la BICfB) pour 3.000 titres avec accroissement prévu sur 20 ans est à peu près équivalent au coût de stockage au titre papier (1 exemplaire) avec espace d'accroissement prévu pour 10 ans (hors maintenance des bâtiments, prix du terrain, frais de traitement initiaux du document qui peuvent être importants²⁹⁵). Indépendamment de ces estimations conjecturales, la dégressivité maximale du rapport coût/stockage, l'investissement linéaire et la flexibilité en termes d'accroissement du modèle LOCKSS (ajouter un ou quelques Tb de stockage est sans difficulté – ajouter des m² est plus coûteux...) plaident dans tous les cas en faveur de la préservation numérique.

Le coût de l'usage est un argument supplémentaire en faveur de la préservation numérique. Dans le cas de la préservation électronique via LOCKSS, les coûts d'accès aux documents sont inclus dans les coûts de gestion globaux, que le document soit préservé ou non, puisque l'accès passe des infrastructures existantes dans le cadre d'autres fonctions (résolveurs de liens, réseau...²⁹⁶). Si l'on se base sur un coût de 2 à 3 € la demande d'accès à un titre papier rangé en réserve, ce qui est sous-évalué dans nombre de configurations²⁹⁷, il suffit de deux demandes de fourniture d'article par an pour couvrir le coût total d'archivage électronique d'un titre de périodique via LOCKSS. Or, en cas d'événement déclencheur (fin d'abonnement, faillite de l'éditeur, transfert de titre...) concernant un titre archivé en format papier, si l'accès électronique n'est pas disponible par ailleurs (via l'accès post-abonnement sur la plateforme de l'éditeur ou une solution tierce), la fourniture de document (copie, scan ou transfert) est la seule solution possible pour assurer l'accès au document aux usagers. Pour ces titres, dont la préservation papier serait la seule garantie d'accès pérenne, il serait d'ailleurs raisonnable, selon la politique définie pour le groupe de travail CIUF de Conservation partagée des périodiques (Groupe CPP), de conserver 2 exemplaires – l'un de consultation, l'autre de préservation – avec la multiplication des coûts correspondante.

²⁹⁵ Cooper, 2006 et Schonfeld, King, Okerson, & Gifford Fenton, 2004, indiquent que dans tous les cas, les coûts de traitement des documents sont plus élevés pour le papier que pour l'électronique.

²⁹⁶ Comme déjà signalé, les coûts en bande passante ne sont pas comptabilisés, puisqu'il serait impossible de distinguer l'accès à des titres préservés d'autres usages.

²⁹⁷ Par exemple dans le cas de réserves distantes ne disposant pas de service de scan sur place, dans le cas d'une demande faible imposant des transferts nombreux pour un nombre relativement faible de documents, etc.

On peut aussi estimer que le coût de préservation annuel via LOCKSS pour une grande institution est équivalent au coût de traitement d'environ 4.000 à 6.000 demandes de fourniture de documents par an.

Finalement, dans le cas où l'on considère l'archivage papier comme une alternative de secours pour les publications accessibles en électronique, il faut ajouter aux coûts de stockage et d'usage un coût supplémentaire d'acquisition du document, puisque celui-ci doit être à la fois souscrit en électronique et acheté en papier (généralement sous la forme « print + electronic » ou « print add on », avec un surcoût limité par rapport à l'achat de la version électronique seule). Actuellement, l'effet peut, dans certains cas, être bénéfique puisque l'achat papier peut permettre de bénéficier d'une TVA de 6% au lieu d'une TVA de 21%. Toutefois, cet effet bénéfique n'est pas généralisable ni sans doute permanent, et ne peut entrer en ligne de compte comme un argument financier en faveur de l'archivage papier que dans des cas très spécifiques.

6.2 Autres aspects

Au vu des éléments présentés ci-dessus, l'archivage papier, même partagé, ne semble pas une alternative globale intéressante à l'archivage électronique via LOCKSS lorsque celui-ci est possible. LOCKSS offre en effet des garanties d'archivage pérenne qui se rapprochent de celles du papier (contrôle local des données notamment) pour un coût inférieur.

Au-delà des arguments économiques, l'archivage papier, s'il offre un degré de contrôle maximal, présente également plusieurs limitations en termes de garanties à long terme et de service :

- l'archivage papier n'est pas possible pour tous les titres du portefeuille électronique des institutions : il ne concerne que les titres pour lesquels il est (encore) possible d'acheter une version imprimée, et en ce sens, ne peut être une solution globale à long terme ;
- l'archivage papier ne présente pas le même degré de confort d'accès pour les usagers : l'accès n'est pas immédiat, et les fonctionnalités électroniques sont absentes, même en cas de scan du document.

6.3 Conclusion et recommandations

Au vu des éléments de coûts, de garanties à long terme et d'accès, l'archivage papier pour des titres disponibles par ailleurs en électronique nous semble devoir être considéré :

- seulement comme une solution de secours à l'accès électronique pérenne, si celui-ci n'est pas assuré de manière satisfaisante via l'archivage dans une solution tierce et/ou l'accès post-abonnement sur la plateforme de l'éditeur – au niveau du titre ou d'états de collection spécifiques ;
- uniquement pour les titres que l'institution estime appartenir au noyau de sa collection ;
- via l'archivage partagé puisque, pour des publications électroniques, un accès papier au sein de chaque institution ne nous semble pas se justifier financièrement.

L'argument est valable à la fois pour les titres courants, natifs numériques, et pour les bouquets d'archives. Toutefois, les bouquets d'archives nécessitent une attention particulière puisqu'il est important qu'une vérification qualité ait été effectuée sur la numérisation.

Pour les cas où l'archivage papier serait considéré comme la solution d'archivage adéquate, il est important de revoir ce choix sur base régulière, en fonction de l'évolution de la couverture des solutions tierces d'archivage, de l'évolution du format de parution du titre, et de l'évolution des négociations des clauses contractuelles et/ou de la politique de l'éditeur.

7 Scénarios envisageables pour la BICfB

Nous proposons d'abord une organisation hiérarchique des solutions de conservation pérenne des ressources électroniques, sous la forme de deux scénarios, en fonction des garanties qu'ils proposent, de leur coût et de la portion du portefeuille de titres BICfB qu'ils permettent de couvrir. Nous traduisons ensuite l'implémentation de ces scénarios sous la forme de recommandations concrètes pour la BICfB.

7.1 Une échelle de solutions d'accès pérenne électronique

En fonction des résultats de l'analyse coûts-bénéfices des solutions tierces d'archivage et des clauses contractuelles des licences BICfB, nous proposons à la BICfB une échelle de solutions permettant d'assurer l'accès pérenne à ses ressources électroniques. Nous distinguons un scénario prioritaire, qui couvre une part importante du portefeuille de titre des institutions avec un niveau de contrôle important et à coûts raisonnables, d'autres pistes complémentaires, plus conjecturales, permettant de couvrir les parties de collections non archivées via le premier scénario (la *long tail*). Tous ces scénarios supposent à la fois des garanties contractuelles et l'intervention de solutions d'archivage mutualisées.

7.1.1 Scénario prioritaire

La combinaison de deux solutions d'archivage articulées avec des garanties contractuelles auprès des éditeurs garantissent l'accès pérenne pour une large part des collections des institutions membres de la BICfB :

- l'installation de LOCKSS Boxes au sein des institutions et l'adhésion à la LOCKSS Alliance ;
- la combinaison de clauses d'accès post-abonnement sur la plateforme éditeur, assorties de garanties contractuelles d'accès en cas de transfert de titres, avec les garanties d'archivage de CLOCKSS.

La seconde solution est la moins coûteuse, la participation à CLOCKSS n'étant pas obligatoire pour bénéficier des garanties du projet, et est déjà effective, sans action de la part des institutions, pour une part du portefeuille de grands éditeurs (notamment les titres Elsevier, Wiley, Springer). C'est une solution particulièrement satisfaisante en termes d'analyse coûts-bénéfices pour les titres des grands éditeurs, qui ont les moyens techniques d'assurer l'accès sur leur plateforme, et qui sont les moins susceptibles de cesser leur activité. Notons qu'en fonction de l'évolution d'e-Depot (vers une gouvernance partagée, avec plus de communication) la solution pourrait être considérée comme une alternative à CLOCKSS pour l'archivage pérenne de certains titres.

L'installation de LOCKSS boxes constitue un investissement et nécessite d'être testée, mais elle présente l'avantage majeur d'assurer aux institutions la possession des fichiers de données. La couverture de LOCKSS est très complémentaire de celle de CLOCKSS et c'est la solution qui nous semble actuellement la plus adaptée aux petits éditeurs et aux titres Open Access. L'approche « en amont » de LOCKSS oblige à des bonnes pratiques de sélection et de gestion. L'installation de LOCKSS boxes est aussi l'occasion de développer des compétences et de tester une solution utilisable également pour l'archivage d'autres données numériques (seule ou en combinaison).

Pour les titres publiés par les grands éditeurs, la négociation de l'accès post-abonnement sur la plateforme éditeur est également utile même en cas de l'archivage de leurs titres au sein des LOCKSS boxes institutionnelles, car elle délègue la gestion des accès à l'éditeur, évitant une trop forte sollicitation du serveur LOCKSS.

La souscription à Portico doit être considérée à notre avis comme une assurance-risque complémentaire facultative, sauf éventuellement en cas d'événement déclencheur pour des titres jugés prioritaires (fin d'abonnement pour un éditeur majeur n'offrant pas de garanties d'accès post-abonnement sur sa plateforme par exemple, ou faillite d'un éditeur non présent dans LOCKSS ou CLOCKSS). Cette remarque vaut également, à notre avis, pour l'accès via des plateformes d'archives avec garanties d'archivage pérenne via souscription (JSTOR), puisque l'accès est conditionné par la souscription. Les plateformes d'archives ouvertes dépendant d'un organisme national, telles Pubmed Central, constituent, de la même manière, une assurance complémentaire, mais non suffisante à notre avis.

7.1.2 Scénarios complémentaires

Pour couvrir la *long tail*, c'est-à-dire les titres du portefeuille des institutions qui ne peuvent bénéficier de l'une des deux solutions du scénario prioritaire, soit que les garanties contractuelles ne soient pas suffisantes, soit que les titres ne soient pas archivés par LOCKSS ou CLOCKSS, on peut envisager une série de pistes complémentaires, plus conjecturelles en fonction des collaborations et projets existants ou possibles.

Les trois premières pistes supposent l'obtention des fichiers physiques de données de la part de l'éditeur ou du fournisseur (à négocier idéalement lors de l'acquisition, en alternative au scénario prioritaire), ainsi qu'un travail d'ingestion plus ou moins important selon les contenus et les systèmes. Etant donné les coûts rédhibitoires de la mise en place d'un système complet d'archivage pérenne, en particulier pour la quantité de contenus finalement peu importante que constitue la *long tail*, nous n'envisageons ici que les pistes qui permettent une certaine mutualisation des coûts, soit entre institutions, soit entre projets d'archivage.

1. La mise en place d'un ou plusieurs réseaux LOCKSS privés (PLN), en collaboration avec les pays voisins ou plus éloignés, semble une possibilité intéressante par la réutilisation de la technologie, non seulement sans coûts de souscription ou d'achat logiciel supplémentaire mais qui contribuerait au développement des compétences et au renforcement mutuel des projets. Plusieurs modalités sont envisageables :

- la participation à un réseau existant impliquant des institutions de la BICfB ou des institutions belges (KBR, ...) mais pour d'autres types de contenus (par exemple le projet développé à l'ULB pour la préservation des données du répertoire institutionnel et de la bibliothèque numérique) ;
- la création d'un réseau avec les pays voisins par exemple pour des contenus européens ou régionaux n'intéressant pas suffisamment d'institutions de la GLN ;
- la mise en place, avec des partenaires internationaux²⁹⁸, d'un réseau dédié à la préservation de bouquets d'archives, dans le cas où la couverture rétrospective de CLOCKSS ne se révélerait pas suffisante.

Ceci nécessite cependant un investissement supplémentaire en termes de matériel et surtout de ressources humaines, afin d'identifier, d'établir et de gérer les partenariats, et de mettre en place la solution technique²⁹⁹.

2. L'utilisation de services de tiers archivage proposés par la KBR ou les Archives de l'Etat. Nous n'avons pas, dans le cadre de cette étude, exploré en détail cette possibilité avec les organismes concernés, qui ne proposent pas actuellement à notre connaissance un service équivalent à celui que

²⁹⁸ Pour rappel en effet, un réseau de nœuds tous situés en Belgique ne satisfait pas aux conditions de préservation pérenne (risque physique).

²⁹⁹ S'il s'agit du même logiciel dans le cas de la LOCKSS Alliance et des PLN, le paramétrage est quant à lui spécifique, de même que les plug-ins pour l'ingestion de contenu.

l'on trouve en France ou en Allemagne. La BICfB pourrait explorer cette piste plus concrètement sur base de la définition de contenus concernés en fonction de ses choix en termes de solution(s) tierce(s) d'archivage. Les coûts au Tb proposés par le service de tiers archivage de la BnF et du CINES, par exemple, sont relativement abordables (cf. point 5.1), mais ces coûts ne comprennent ni le traitement des données pour l'ingestion, ni la gestion de l'accès pour les membres des institutions. L'utilisation de services de tiers archivage de la BnF ou de la DNB pourrait aussi être une possibilité si la BICfB considérait que la gestion de l'archivage d'une partie des e-journaux acquis en Communauté française pouvait être confiée entièrement à des institutions étrangères.

3. L'utilisation de solutions d'archivage sur base de développements locaux ou sur base de logiciels commerciaux mises en place au sein des institutions pour d'autres types de contenus (Dépôt institutionnel, bibliothèque numérique, archives...). Nous n'avons pas connaissance de solutions de ce type abouties au sein des institutions de la BICfB, mais la piste pourrait être investiguée en fonction de la définition de contenus concernés sur base des choix éventuels de la BICfB en matière de solutions tierces et de l'évolution des développements au sein des institutions.

4. Au cas par cas et en fonction de la situation, pour les contenus dont il ne serait pas possible d'obtenir les fichiers sous un format et pour des coûts acceptables, ou si aucune solution d'archivage électronique complémentaire n'était mise en place, **l'archivage papier partagé** pourrait être une solution, éventuellement temporaire en fonction de l'évolution de la situation. Les éléments contractuels, en parallèle avec l'importance des titres pour l'institution (*core collection*), pourraient instruire le choix du nombre d'exemplaires à préserver (1 exemplaire de préservation accompagné ou non d'un exemplaire de consultation) : les garanties d'accès sur la plateforme de l'éditeur (en cas de fin d'abonnement, de transfert de titres et de cessation de publication) pourraient plaider en faveur de la préservation d'un exemplaire de préservation uniquement.

7.1.3 Synthèse

Le tableau suivant synthétise l'ensemble des scénarios et solutions d'accès pérenne selon l'échelle que nous avons présentée ci-dessus comme la plus adaptée aux besoins de la BICfB.

	Solutions d'accès pérenne	Note
Scénario prioritaire	<ul style="list-style-type: none"> • LOCKSS Alliance (en particulier pour les plus petits éditeurs et les titres Open Access) • Accès post-abonnement* sur la plateforme éditeur + TRANSFER + CLOCKSS (en particulier pour les grands éditeurs) 	<p>Tous les titres disponibles peuvent être archivés dans la LOCKSS box, ou seulement ceux qui ne présentent pas toutes les garanties via CLOCKSS + clauses contractuelles.</p> <p>e-Depot pourrait être une alternative à CLOCKSS pour certains éditeurs ou titres Open Access.</p>
Scénarios complémentaires	<ul style="list-style-type: none"> • Transfert de fichiers éditeurs + LOCKSS PNL • Transfert de fichiers éditeur + tiers archivage ou archivage local • Archivage papier partagé 	<p>Solution à choisir en fonction de la conjoncture : possibilités de collaboration, choix effectués pour d'autres projets d'archivage numérique, possibilité d'obtention des fichiers de données, analyse coûts-bénéfices.</p>

Solutions n'offrant pas toutes les garanties (risque économique et/ou absence de contrôle)	<ul style="list-style-type: none">• Portico• Accès post-abonnement sur plateforme éditeur + accès via plateformes d'archives avec garanties d'archivage pérenne (Pubmed Central...)• Souscription à des plateformes d'archives avec garanties d'archivage pérenne (JSTOR)	Ces solutions peuvent intervenir en tant qu'assurances complémentaires ou à défaut d'autre solution en cas d'événement déclencheur.
---	---	---

Tableau 32 - Echelle de solutions d'accès pérenne à considérer pour la BICfB. [*Uniquement pour les titres souscrits, pas pertinent pour les titres Open Access]

7.2 Recommandations

En fonction de l'échelle de scénarios garantissant l'archivage pérenne présentée ci-dessus, nous recommandons différents points d'action à la BICfB.

- tester LOCKSS, soutenir CLOCKSS ;
- négocier les clauses contractuelles des éditeurs selon un ordre de priorités adapté au scénario préférentiel recommandé ;
- réfléchir à la notion de collection pérenne et gérer l'information ;
- investiguer les possibilités de partenariat (en interne, sur d'autres projets, et en externe, au niveau national ou européen).

7.2.1.1 Tester LOCKSS, soutenir CLOCKSS

Si LOCKSS est la solution d'archivage numérique qui permet le plus de contrôle par les institutions, c'est aussi celle qui demande le plus à être testée rapidement, puisque (1) le moissonnage des données est lent et qu'il doit être effectué pendant l'autorisation d'accès aux contenus et (2) pour les titres non encore disponibles dans LOCKSS, il est nécessaire de tester l'efficacité du workflow décisionnel de la LOCKSS Alliance et ses limites. Par ailleurs, LOCKSS est Open Source et utilisable en combinaison avec d'autres logiciels pour le développement de certaines fonctionnalités (audit, gestion de collections, gestion logique et migration de fichiers), qui pourraient également faire l'objet de tests ou de développements complémentaires. L'analyse des coûts a mis en évidence la possibilité d'économies d'échelle importantes, dans tous les cas, puisque le coût de souscription à la solution représente la part la plus importante des coûts et est indépendante du nombre de titres et volumes archivés. Il est donc recommandé d'archiver un grand nombre de titres dès le début, après la période de test, ce qui diminue l'impact du processus de sélection dans le workflow d'archivage.

La question du niveau de participation à la LOCKSS Alliance et de l'organisation de la gestion des LOCKSS Boxes pour les institutions de la BICfB, en particulier pour les petites institutions, n'est pas résolue a priori, et serait à définir lors de la première phase d'implémentation³⁰⁰. Il est possible que les coûts de l'installation d'une LOCKSS Box en local, en particulier en ressources humaines, dépassent les capacités financières des petites institutions. Une piste de solution pourrait être la délégation de tout ou d'une partie des activités de gestion de ces LOCKSS Boxes à une grande institution partenaire : soit uniquement la délégation des activités de maintenance informatique (installation du logiciel, mises à jour, monitoring), soit, en cas de partenariat existant au niveau du cadastre des ressources électroniques, la délégation des activités de maintenance et des activités d'activation des titres. Ce principe de délégation va à l'encontre du principe de distribution des risques au cœur de LOCKSS, mais pourrait être acceptable dans la mesure où le stockage physique

³⁰⁰ Les différences essentielles entre les différentes modalités d'organisation des réseaux LOCKSS est synthétisée très clairement par Rosenthal, D. S. (2013). Architectural Choices in LOCKSS Networks (preprint). Consulté à l'adresse <http://www.lockss.org/locksswp/wp-content/uploads/2013/12/ArchitecturalChoices.pdf>

resterait distribué. Une participation à la LOCKSS Alliance avec une seule LOCKSS Box pour la BICfB est théoriquement possible mais ne serait peut-être pas acceptée par la LOCKSS Alliance, qui fonctionne normalement sur des bases institutionnelles. En outre, les collections courantes des institutions étant gérées – et différentes – au niveau institutionnel, la mise en place d'un serveur centralisé de préservation des collections courantes pour l'ensemble de la BICfB nécessiterait l'ajout d'une gestion spécifique des accès, dont les possibilités techniques et l'impact financier devraient être évalués.

Nous recommandons à la BICfB, si elle en a les moyens (au niveau institutionnel ou consortial), de soutenir financièrement l'initiative CLOCKSS, non seulement, et en priorité, parce que c'est une initiative extrêmement économique et portée de concert par les bibliothèques et les éditeurs, qui permet d'assurer la préservation pérenne de collections de titres représentant un investissement majeur pour les institutions de la BICfB, mais aussi parce qu'elle utilise le logiciel LOCKSS et contribue dès lors à sa pérennité.

7.2.1.2 Négocier les clauses des licences avec les éditeurs selon un ordre de priorité

L'échelle de stratégies d'accès pérenne aux e-journaux que nous recommandons ci-dessus définit aussi les **objectifs de négociation avec les éditeurs** ou fournisseurs lors de la souscription aux journaux électroniques :

- 1) l'accès post-abonnement sur la plateforme éditeur et la participation au moins à LOCKSS ou à CLOCKSS, voire à une autre solution d'archivage pérenne qui peut être une *dark archive* (par exemple e-Depot) ; insister aussi sur les garanties en cas de transfert de titres (par exemple via la signature du TRANSFER Code of practice) ;
- 2) si l'accès post-abonnement sur la plateforme éditeur n'est pas possible, négocier au moins la participation de l'éditeur à LOCKSS ;
- 3) si ni la solution 1) ni la solution 2) ne sont acceptées, négocier au moins le transfert de fichiers dès le début de l'abonnement, afin de pouvoir éventuellement assurer en local la préservation de ces contenus ;
- 4) si aucune des solutions précédente n'est acceptée ou possible à des coûts raisonnables, négocier une offre avantageuse pour l'achat d'un exemplaire papier de préservation.

Parmi les grands éditeurs auprès desquels les institutions souscrivent des abonnements, les éditeurs prioritaires pour les négociations sont les suivants :

- en premier lieu, les éditeurs ne fournissant pas d'accès post-abonnement via leur plateforme et ne participant ni à LOCKSS ni à CLOCKSS : par exemple IOS Press (Metapress), OVID Journals, Jura, Future Medicine, Peeters ;
- ensuite, et dans une moindre mesure lorsque ces éditeurs participent à Portico ou à d'autres solutions d'archivage pérenne, les éditeurs ne participant pas à LOCKSS ni à CLOCKSS mais proposant un accès post-abonnement via leur plateforme : par exemple ACS, Cambridge University Press.

Rappelons aussi l'importance, soulignée point 2.3.1, pour les clauses déjà existantes comme pour les clauses nouvellement négociées, de définir clairement la couverture, les conditions et le coût de chaque clause d'accès et d'archivage.

7.2.1.3 Définir la notion de collection pérenne et améliorer la gestion de l'information

Afin de pouvoir gérer l'accès pérenne à son portefeuille, via l'identification et la mise en place pour chaque titre ou collection du scénario d'accès et d'archivage adapté en fonction du contrat et de la

politique de l'éditeur, il est fondamental que chaque institution définisse des priorités au sein de son portefeuille et se donne les moyens de disposer de l'information nécessaire.

La définition des priorités passe par la notion de collection pérenne – c'est-à-dire dont l'acquisition ou la sélection a été envisagée afin de permettre un accès sur le long terme : cette notion s'applique-t-elle à l'ensemble des titres souscrits, y compris les titres des bases de données full-text et les titres Open Access répertoriés ? Ou bien faut-il distinguer au sein des acquisitions, des acquisitions à long terme et des acquisitions destinées uniquement à l'accès courant (par exemple via les bases de données avec texte intégral), et appliquer une politique conséquente en matière de négociation de prix et de solutions d'archivage ? Qui doit assumer la responsabilité de l'archivage pérenne des titres Open Access ? Les bibliothèques doivent-elles assurer et contrôler l'archivage pérenne de l'ensemble des titres Open Accès qu'elles sélectionnent ? Ou participer à l'archivage de l'ensemble des titres Open Access publiés ? Les réponses de chaque institution à l'ensemble de ces questions lui permettront, en combinaison avec l'analyse des solutions disponibles au cas par cas, d'établir les priorités d'archivage au sein de son portefeuille.

Pour les titres Open Access publiés sur des plateformes d'archives ouvertes, par exemple, les garanties d'archivage assumées par l'éditeur en conformité avec les principes de l'archivage pérenne (via un service d'archivage tiers par exemple, comme dans le cas de Persée) peuvent être jugées satisfaisantes.

Nous avons vu aussi que la gestion de l'information est cruciale en matière d'archivage pérenne, 1) puisqu'il faut disposer des clauses contractuelles pour chaque titre et état de collection et 2) puisqu'il faut savoir, pour chaque titre, dans quelle(s) solution(s) d'archivage (LOCKSS, CLOCKSS, Portico, solution interne, archivage papier) il est disponible, et pour quels états de collection.

Il est donc important d'améliorer, dans la mesure du possible, la prise en compte de cette information, qui inclut notamment une gestion de l'historique des contrats, par les systèmes de gestion des ressources électroniques et les solutions d'archivage : via l'ajout de champs spécifiques à l'archivage pérenne s'ils ne sont pas présents dans les systèmes de gestion des ressources électroniques en place, via le remplissage systématique (et idéalement automatique) de ces champs, via l'amélioration du détail d'information disponible dans les solutions d'archivage (notion de plateforme), et via l'amélioration des fonctionnalités de comparaison et de transfert d'information entre les systèmes. Tous ces éléments supposent notamment l'utilisation de standards communs (par exemple les déclinaisons du vocabulaire ONIX). L'articulation de solutions d'archivage pérenne différentes complexifie cette gestion mais la rend nécessaire.

Le projet JISC « Post-cancellation Entitlement Registry Scoping Project » a mis au point une liste des champs à considérer pour l'inclusion dans un système de gestion de l'information³⁰¹. Il s'agit d'un premier pilier auquel il faudrait préciser les informations opérationnelles (sources de données, workflows et harvesting, vocabulaires...), et éventuellement compléter. Ces informations interviennent en amont dans la gestion des e-ressources et se distinguent, tout en partageant de nombreux points communs, des métadonnées de préservation proprement dites, en particulier PREMIS³⁰², qui constituent les métadonnées associées à tout objet préservé dans un entrepôt « comme fondement du processus de préservation ». Lors de l'archivage proprement dit, une partie

³⁰¹ JISC, Post-cancellation Entitlement Registry Scoping Project, final report mai 2012, Appendix B.

³⁰² Le vocabulaire PREMIS ("PREservation Metadata: Implementation Strategies"), mis au point par le PREMIS Working Group, une collaboration internationale d'experts impliqués dans des activités de préservation créée en 2003, sponsorisé par l'OCLC et le RLG, représente le standard actuel de métadonnées de préservation (version 1 du vocabulaire en 2005, dernière version en 2008). Voir aussi Comprendre PREMIS, 2009. http://www.loc.gov/standards/premis/Understanding-PREMIS_french.pdf

de ces données (en particulier les droits d'accès) peut être transférée dans le système d'archivage, en même temps que l'objet.

L'amélioration de la gestion de l'information passe par un dialogue constant avec les éditeurs, les développeurs de systèmes de gestion de l'information et les solutions tierces.

7.2.1.4 *Investiguer les possibilités de partenariat*

La question de l'archivage numérique à long terme est une problématique fondamentale dans la société de l'information, et elle se pose concrètement pour un grand nombre d'organisations. L'archivage des e-journaux scientifiques non assuré – en tous cas pour le moment – via le scénario prioritaire recommandé pourrait bénéficier de solutions mises en place pour d'autres types de contenu au sein des institutions ou au niveau national ou européen. En fonction des priorités définies en matière de collection pérenne, il serait donc intéressant que la BICfB investigue, éventuellement dans un deuxième temps, les possibilités de partenariat avec :

- les solutions d'archivage collaboratives mises en place dans les institutions, par exemple pour le dépôt institutionnel ou les collections numérisées (par exemple le projet pilote de LOCKSS PNL à l'ULB) ;
- les solutions d'archivage mises en place au niveau national ou international, en particulier pour des collaborations éventuelles pour des contenus « e-journaux » spécifiques (titres locaux ou européens, archives).

Par ailleurs, la réalité de l'archivage pérenne évoluant rapidement au niveau international, il est souhaitable de maintenir une veille d'information plus générale sur l'évolution des projets (LOCKSS, CLOCKSS, Portico, mais aussi d'autres initiatives), sur les choix effectués par d'autres pays ou consortia européens (en particulier par l'Allemagne, qui étudie actuellement la mise en place d'une solution nationale d'archivage des e-journaux, et la Suisse, cf. point 5.4), sur les outils et logiciels disponibles (par exemple les outils Open Source), et les standards.

7.2.1.5 *Synthèse et planning*

Nous proposons ci-dessous un planning des actions proposées à la BICfB sur deux ans, en fonction des scénarios présentés dans les points précédents.

Année 1 :

Prioritaire :

- **tester la LOCKSS box** dans une institution au moins (installation, configuration, temps d'ingestion des titres OA, réactivité de l'équipe de Stanford, complexité du logiciel, présence de documentation et transparence de l'organisation et de la prise de décisions, intégration avec les résolveurs de liens et systèmes de gestion des ressources électroniques existants, performance en termes d'accès) ; archiver en priorité les titres sans clause post-abonnement sur la plateforme éditeur ; si possible, dans un second temps, tester l'installation dans l'ensemble des institutions de la BICfB ; attention à bien définir rôles et responsabilités du staff et le workflow de préservation (sélection, activation...) ;
- **négoier auprès des éditeurs** pour les titres « courants » et les bouquets d'archives la couverture des licences pour l'accès post-abonnement ainsi que les solutions d'archivage pérenne selon le degré de priorité défini en 8.2.1.2 ;
- vérifier et enregistrer les informations concernant l'archivage pérenne dans les systèmes de gestion des ressources électroniques existants au sein des institutions ; définir les améliorations nécessaires.

Facultatif mais recommandé :

- **soutenir CLOCKSS**, au moins au nom de la BICfB si ce ne peut être au niveau institutionnel ;
- s'impliquer dans le développement ou la documentation de LOCKSS, tester éventuellement les articulations de LOCKSS avec d'autres couches logicielles, pour la gestion des collections ou la préservation logique ;
- intensifier le dialogue avec les éditeurs pour le respect des standards et l'amélioration des données (KBART, ONIX, TRANSFER) ;
- s'impliquer dans la gouvernance des solutions LOCKSS et CLOCKSS, et notamment participer à la réflexion sur l'évolution des *business models* des solutions tierces ;
- effectuer une veille sur les projets internationaux et nationaux, les projets consortiaux (Solution nationale allemande, suisse, JISC Entitlement Registry...).

Année 2 :

En fonction des résultats de l'année 1 :

- si résultats positifs (test de la LOCKSS Box, et, éventuellement, extension de la couverture des licences éditeurs, développement de la communauté LOCKSS) : souscrire à la LOCKSS Alliance, installer la LOCKSS Box dans toutes les institutions BICfB et effectuer l'ingestion des titres disponibles dans LOCKSS pour chaque institution (temps d'ingestion) – voir s'il est possible d'accélérer le processus via ingestion directe via backup de la première LOCKSS box.

En fonction de la conjoncture (collaborations, moyens financiers) :

- examiner les possibilités de partenariats, en Belgique ou en Europe, notamment pour un PNL, pour la partie non couverte du portefeuille ;
- éventuellement souscrire à Portico comme assurance-risque complémentaire.

Année 3 et suivantes :

- maintenir la veille et l'implication dans les solutions tierces d'archivage, et **réévaluer les choix périodiquement !**

8 Conclusion

La présente étude avait pour objectif d'éclairer les choix de la BICfB en matière d'archivage pérenne des ressources électroniques, en examinant en particulier les solutions tierces d'archivage existantes (LOCKSS, CLOCKSS, Portico), en évaluant le coût des solutions envisagées, en articulant ces solutions avec l'archivage papier et en proposant des scénarios d'application.

Elle a rappelé et illustré (chapitre 1) l'urgence de la question et l'importance de la distinction des concepts en matière d'archivage pérenne électronique : accès post-abonnement, archivage pérenne, accès pérenne ne sont pas des concepts purement théoriques mais distinguent des réalités très différentes en matière de solution d'archivage, d'accès et de couverture. Seul l'accès pérenne impliquant d'une part la garantie d'accès post-abonnement et d'autre part la garantie d'archivage pérenne peut être une solution acceptable pour la BICfB. Par ailleurs, étant donné la nature de la problématique et l'investissement nécessaire, seules des solutions mutualisant les ressources devraient être considérées.

L'examen des licences contractuelles (chapitre 2) de produits souscrits en consortium par la BICfB – qui sont prioritaires sur le cadre légal, de toutes façons insuffisant – ainsi que de quelques autres bouquets de périodiques d'éditeurs importants a mis en évidence l'importance d'actions rapides en

termes de négociation et d'information. En l'état actuel, l'accès pérenne aux titres électroniques souscrits par la BICfB n'est garanti que dans un nombre de cas limités, malgré la présence de clauses d'archivage pérenne, négociées ou non, dans la majorité des contrats signés. En effet, à l'exception de CLOCKSS, les solutions tierces d'archivage pérenne des e-journaux électroniques nécessitent un investissement financier et humain de la part des institutions, pour majeure partie sous la forme d'une souscription. En outre, les clauses contractuelles sont souvent imprécises en ce qui concerne les modalités d'accès pérenne : coût d'accès post-abonnement sur la plateforme de l'éditeur, coûts et modalités de transfert des fichiers le cas échéant, garanties d'archivage pérenne via des solutions tierces, processus décisionnel en cas d'alternatives. Finalement, dans nombre de contrats, seule une partie des contenus souscrits, en particulier dans le cadre des *big deals*, est couverte par les garanties d'accès post-abonnement ; l'accès courant ne peut donc être équivalent à l'accès aux contenus archivés. De la même manière, les titres accessibles en électronique dans le cadre du contrat courant mais non souscrits en électronique depuis l'origine (globalement, les années 1997-2010) posent problème car ils sont généralement absents des clauses post-abonnement, mais ne sont pas non plus proposés à l'achat dans le cadre des bouquets d'archives.

La prise en compte des spécificités de la préservation des e-journaux par rapport à l'archivage numérique pérenne en général permet de cadrer les caractéristiques minimales des solutions de préservation envisageables (point 3.1) : un mandat spécifique, le résultat d'une collaboration entre bibliothèques et éditeurs, une mutualisation des ressources et une répartition des responsabilités, des conditions d'accès suffisantes pour les usagers des bibliothèques, une couverture large, multidisciplinaire. Des solutions d'archivage numérique intéressantes par ailleurs, tels les dépôts institutionnels, les logiciels de préservation numérique et les plateformes d'archives, ne peuvent être considérées comme des solutions satisfaisantes à la problématique (point 3.2).

Les solutions dédiées à l'archivage pérenne des e-journaux et présentant un certain degré de mutualisation des ressources (LOCKSS, CLOCKSS, Portico) ont été comparées (chapitre 4) sur base de six critères : garanties techniques et organisationnelles de pérennité, couverture, accès, contrôle et coût. Ceci a permis de mettre en évidence les philosophies fondamentalement différentes – et complémentaires – de ces projets, et a mené à plusieurs conclusions : les trois solutions existent depuis le milieu des années 2000 et peuvent être considérées comme matures, mais seule LOCKSS GLN offre selon nous des garanties suffisantes d'accès pérenne pour les institutions membres de la BICfB, même si Portico, qui travaille avec la majorité des grands éditeurs, est plus riche en matière de couverture. La combinaison d'un accès post-abonnement sur la plateforme des éditeurs et de la solution CLOCKSS peut présenter également des garanties suffisantes, sauf dans le cas de transfert de titre d'un éditeur à un autre. La littérature et le bon sens recommandent de limiter les risques en participant à différents projets de préservation. Malheureusement, le rapport coûts-bénéfices de la participation à LOCKSS et à Portico est prohibitif, puisqu'il représenterait, pour un gain assez relatif en termes de couverture du portefeuille, une augmentation de coûts de l'ordre de 80%.

Les coûts de mise en place des solutions dépendent en majeure partie des coûts de souscription annuelle aux solutions, qui varient entre environ \$3.000 et \$10.000 (prix officiels) selon la taille des institutions. Ces coûts, si on les considère comme un retour sur investissement, semblent raisonnables pour les grandes institutions de la BICfB, puisqu'ils représentent moins de 0,5 ou de 1% des investissements annuels en ressources électroniques. Pour les plus petites institutions, le rapport peut atteindre 8%, ce qui rend l'investissement plus difficile. Ceci en fait le facteur majeur d'incertitude même si les différentes solutions tierces affirment viser la limitation des coûts au maximum. Le coût de stockage physique et la gestion humaine ne sont dans aucune solution directement proportionnels au nombre de titres archivés, ce qui permet, quelle que soit la solution, des économies d'échelle. Les simulations de coûts effectuées sur 20 ans pour LOCKSS et Portico présentent toutefois trop de facteurs d'incertitude (coûts d'ingestion et de migration pour les

solutions, coûts de stockage physique dans le cas de LOCKSS) actuellement pour permettre une planification réelle des coûts.

Les initiatives nationales en matières de préservation pérenne liées aux e-journaux ou plus larges sont nombreuses en Europe (chapitre 5) : systèmes mis en place par les bibliothèques nationales en lien avec le dépôt légal, licences nationales et systèmes de préservation associés (en particulier en France), réseaux de préservation distribuée (en Allemagne et en Angleterre). Plusieurs pays (en tous cas l'Allemagne et la Suisse) sont également actuellement en train d'étudier la mise en place d'une solution globale de préservation pérenne des e-journaux au niveau national impliquant LOCKSS et/ou CLOCKSS et/ou Portico et/ou d'autres solutions complémentaires. Ces projets sont intéressants à suivre, que ce soit en vue de s'inspirer de bonnes pratiques (par exemple dans le cas de UK LOCKSS) et/ou en vue de collaborations éventuelles.

La comparaison des coûts de l'archivage électronique dans LOCKSS (chapitre 6) – avec les restrictions déjà signalées sur la validité des projections – avec les coûts de l'archivage papier, sans pouvoir prendre en compte l'ensemble des données de l'archivage papier mais en se basant sur quelques éléments chiffrés concernant le stockage et l'accès, a montré que l'archivage papier, même partagé, n'est vraisemblablement pas une alternative financièrement intéressante à l'archivage électronique mutualisé. L'archivage papier présente également des limites en termes d'accès et de disponibilité sur le long terme, et ne doit être considéré, à notre avis, que comme une solution de dernier recours pour garantir l'accessibilité à long terme des publications scientifiques pour lesquelles aucune solution de préservation numérique n'a été mise en place.

Sur base de ces éléments, nous avons proposé (chapitre 7) une échelle de solutions garantissant l'accès pérenne aux périodiques scientifiques électroniques, en distinguant une stratégie prioritaire, basée sur la participation à la LOCKSS Alliance, la présence de clauses garantissant l'accès aux titres sur la plateforme éditeur en cas de transfert, de cessation de publication ou de fin d'abonnement, et la prise en compte de l'archivage via CLOCKSS, et des pistes complémentaires, plus diverses et plus ou moins coûteuses, à sélectionner au cas par cas en fonction d'éléments externes et d'analyse coûts-bénéfices, pour garantir la préservation pérenne du portefeuille non couvert par le premier scénario. Il nous semble important, au vu de la maturité des solutions et de l'importance de la problématique, que la BICfB suive une démarche pragmatique et s'investisse dès maintenant dans la mise en place du scénario recommandé, en testant LOCKSS, en soutenant CLOCKSS, en s'impliquant dans les négociations avec les éditeurs et dans la communauté internationale, et en améliorant la gestion de l'information d'accès pérenne dans les systèmes de gestion des institutions (cf. point 7.2.1.5). En parallèle, il est important que les institutions définissent une politique d'archivage pérenne de leurs collections électroniques, notamment en fonction de réponses aux questions de responsabilité d'archivage (en particulier pour les titres Open Access) et de modèles d'acquisition (en particulier pour les bases de données avec texte intégral).

9 Bibliographie

- Aligning National Approaches to Digital Preservation, M., Nancy Y, Skinner, K., & Educopia Institute. (2012). *Aligning National Approaches to Digital Preservation*. Atlanta, Ga.: Educopia Institute Publications. Consulté à l'adresse http://www.educopia.org/sites/default/files/ANADP_Educopia_2012.pdf
- Altenhöner, R., & Steinke, T. (2010). Kopal: cooperation, innovation and services: Digital preservation activities at the German National Library. *Library Hi Tech*, 28(2), 235-244. doi:10.1108/07378831011047640
- Altman, M., & Crabtree, J. (2011). TRAC-Based Auditing of LOCKSS: Using the SAFE-Archive System - archiving2011_altman_crabtree.pdf. Présenté à Archiving 2011. Consulté à l'adresse http://thedata.org/files/thedata_new2/files/archiving2011_altman_crabtree.pdf
- Ayris, P. (2009). LIBER's involvement in supporting digital preservation in member libraries. *LIBER Quarterly*, 19(1), 22-43.
- Beagrie, N. (2010). *Ensuring perpetual access establishing a federated strategy on perpetual access and hosting of electronic resources for Germany* (Edited Final Report). Consulté à l'adresse http://www.allianzinitiative.de/fileadmin/hosting_studie_e.pdf
- Beagrie, N., Lavoie, B. F., & Woollard, M. (2010). *Keeping research data safe 2*. HEFCE. Consulté à l'adresse <http://repository.essex.ac.uk/2147/1/keepingresearchdatasafe2.pdf>
- Becker, C., Kulovits, H., Guttentbrunner, M., Strodl, S., Rauber, A., & Hofman, H. (2009). Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries*, 10(4), 133-157. doi:10.1007/s00799-009-0057-1
- Beh, E., & Smith, J. (2012). Preserving the Scholarly Collection: An Examination of the Perpetual Access Clauses in the Texas A&M University Libraries' Major E-Journal Licenses. *Serials Review*, 38(4), 235-242. doi:10.1016/j.serrev.2012.10.005
- Bote, J., Fernandez-Feijoo, B., & Ruiz, S. (2012). The Cost of Digital Preservation: A Methodological Analysis. *Procedia Technology*, 5, 103-111. doi:10.1016/j.protcy.2012.09.012
- Brindley, G., Muir, A., & Proberts, S. (2004). Provision of digital preservation metadata: A role for ONIX? *Program*, 38(4), 240-250.
- Burnhill, P. (2013). Tales from The Keepers Registry: Serial Issues About Archiving & the Web. *Serials Review*, 39(1), 3-20. doi:10.1016/j.serrev.2013.02.003
- Burnhill, P., & Guy, F. (2010). Piloting an E-journals Preservation Registry Service (PEPRS). *Serials Librarian*, 58(1-4), 117-126. doi:10.1080/03615261003622742
- Carbone, P. (2010). *Coûts, bénéfices et contraintes de la mutualisation des ressources électroniques : éléments de comparaison internationale et propositions* (No. 2010-012) (p. 91). Consulté à l'adresse http://media.enseignementsup-recherche.gouv.fr/file/2009/69/8/rapport_MRE_diffusion_159698.pdf
- Carr, P. L. (2011). The Commitment to Securing Perpetual Journal Access: A Survey of Academic Research Libraries. *Library Resources & Technical Services*, 55(1), 4-16.
- Cooper, M. D. (2006). The Costs of Providing Electronic Journal Access And Printed Copies of Journals to University Users. *Library Quarterly*, 76(3), 323-351.
- Crabtree, J. (2011). *LOCKSS Auditing using the SAFE Archive System*. Consulté à l'adresse http://www.metaarchive.org/public/resources/presentations/2010PLN_presentations/Crabtree_DataPASS_2010_SAFE_Audit_Overview_Final.pdf

- Crabtree, J., Altman, M., (2011), *TRAC-Based Auditing of LOCKSS: Using the SAFE-Archive System*, Archiving 2011 (16-19 mai 2011, Salt Lake City, Utah), Proceedings. Consulté à l'adresse http://thedata.org/files/thedata_new2/files/archiving2011_altman_crabtree.pdf
- CRL Report on Portico Audit 2010. (2010). Consulté à l'adresse <http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf>
- D'Amato, G., & Mazzocchi, R. (2008). Conservando più, copie rendiamo sicure le nostre collezioni digitali. (Italian). *Biblioteche Oggi*, 26(7), 9-13.
- Dale, R. L., & Ambacher, B. (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Consulté à l'adresse http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
- Dalton, P., & Conyers, A. (2008). Evaluation of the JISC UK LOCKSS Pilot. Consulté à l'adresse <http://www.era.lib.ed.ac.uk/handle/1842/3336>
- Deutsche Forschungsgemeinschaft (DFG). (2012). Förderprogramm « Elektronische Publikationen ». Ausschreibung « Nationales Hosting elektronischer Ressourcen ». Consulté à l'adresse http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_elektronische_publicationen_120430.pdf
- Earney, L. (2012). Post-Cancellation Entitlement Registry Scoping Study-Final Report. Consulté à l'adresse <https://www.jisc-collections.ac.uk/Reports/entitlementregistryreport/>
- E-only scholarly journals: overcoming the barriers | Research Information Network*. (2010). Consulté à l'adresse <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/e-only-scholarly-journals-overcoming-barriers>
- Etude préliminaire « e-archiving » Consortium des bibliothèques universitaires suisses*. (2005). Consulté à l'adresse http://lib.consortium.ch/external_files/Konzeptstudie_F_V1_2_1.pdf
- Fauduet, L., & Peyrard, S. (2013). Destination: Shared Repository. The National Library of France's Journey to Third-Party Archiving. In *Proceedings of the 10th International Conference on Preservation of Digital Objects (iPRES)*. Lisbonne. Consulté à l'adresse http://purl.pt/24107/1/iPres2013_PDF/Destination%20Shared%20Repository%20The%20National%20Library%20of%20France%27s%20Journey%20to%20Third-Party%20Archiving.pdf
- Fenton, E. G., & Kirchhoff, A. (2006). An Overview of Portico: An Electronic Archiving Service. *Serials Review*, 32(2), 81-86. doi:10.1016/j.serrev.2006.03.004
- Final Report of the 2CUL LOCKSS Assessment Team*. (2011). Consulté à l'adresse <http://2cul.org/sites/default/files/2CULLOCKSSFinalReport.pdf>
- Fromm, N. (2013, juin 18). *Archivierung elektronischer Ressourcen mit LOCKSS*. Consulté à l'adresse <http://files.dnb.de/nestor/veranstaltungen/Praktikertag2013/2013-06-lockss-fromm.pdf>
- Gust, P. (2012). Using a digital preservation system to ensure access to online digital content. *Library Hi Tech News*, 29(9), 5-9. doi:10.1108/07419051211294437
- Hole, B., Wheatley, P., Li Lin, McCann, P., & Aitken, B. (2010). The Life3 Predictive Costing Tool for Digital Collections. *New Review of Information Networking*, 15(2), 81-93.
- JISC Collections. (2010, février). Ensuring that « e » doesn't mean ephemeral. A practical guide to e-journal archiving solutions. Consulté à l'adresse http://www.jisc-collections.ac.uk/Documents/practical_guide_to_ejournal_archiving.pdf
- Jones, M. (s. d.). e-Journals: Archiving and Preservation Briefing paper. Consulté 29 juillet 2013, à l'adresse http://sitecore.jisc.ac.uk/publications/briefingpapers/2007/pub_ejournalspreservationbp.aspx

- Kastellec, M. (2012). Practical Limits to the Scope of Digital Preservation. *Information Technology & Libraries*, 31(2), 63-71.
- Kejsler, U. B., Nielsen, A. B., & Thirifays, A. (2011). Cost model for digital preservation: Cost of digital migration. *International Journal of Digital Curation*, 6(1), 255–267.
- Kenney, A. R., Entlich, R., & Hirtle, P. B. (2006). *E-Journal Archiving Metes and Bounds: A Survey of the Landscape*. Consulté à l'adresse <http://www.clir.org/pubs/reports/pub138/pub138.pdf>
- King, D. W., & Tenopir, C. (2011). Some economic aspects of the scholarly journal system. *Annual Review of Information Science and Technology*, 45(1), 295–366.
- Kirchhoff, A. J. (2009). Expanding the Preservation Network: Lessons from Portico. *Library Trends*, 57(3), 476-489.
- Kirchhoff, A., & Fenton, E. (2008). Preserving E-Journals: An Overview of Portico. *Serials Librarian*, 53(4), 113-123. doi:10.1300/J123v53n04_08
- LOCKSS Audit Report -2007. (2007). Consulté à l'adresse http://www.crl.edu/sites/default/files/attachments/pages/LOCKSS_Audit_Report_11-07.pdf
- Manz, S. (2012). Ensuring perpetual access to licensed content: introduction to the main issues. Consulté à l'adresse <http://e-collection.library.ethz.ch/view/eth:5613>
- Maple, A., Wright, C., & Seeds, R. (2003). Analysis of format duplication in an academic library collection. *Library Collections, Acquisitions, and Technical Services*, 27(4), 425-442. doi:10.1016/j.lcats.2003.09.003
- McLeod, R.; Wheatley, P.; Ayris, P. (2006). *LIFE 1 : Final Report*. Consulté à l'adresse <http://discovery.ucl.ac.uk/1854/1/LifeProjMaster.pdf>
- Micah Altman. (2012, décembre 25). *Auditing Distributed Preservation Networks*. Consulté à l'adresse <http://fr.slideshare.net/drmaltman/auditing-distributed-preservation-networks>
- Minimum Criteria for an Archival Repository of Digital Scholarly Journals: Criteria. V.1.2. (2000). Consulté à l'adresse <http://old.diglib.org/preserve/criteria.htm>
- Montgomery, C. H., & King, D. W. (2002). Comparing library and user related costs of print and electronic journal collections: A first step towards a comprehensive analysis. *D-Lib Magazine*, 8(10).
- Morrissey, S. M., Stoeffler, M., Cheruku, V., Howard, W. J., Meyer, J., & Kadirvel, S. (2012). Migration at Scale: A Case Study. Consulté à l'adresse <http://www.portico.org/digital-preservation/wp-content/uploads/2012/11/MigrationAtScale.pdf>
- Morrow, T., Beagrie, N., & Jones, M. (2008). *A Comparative Study of e-Journal Archiving Solutions : a JISC funded investigation* (Final report) (p. 41). Consulté à l'adresse http://www.jisc-collections.ac.uk/Documents/e_journal_archiving_%20solutions_comparative_study.pdf
- Neil, A. (2009). Portico Holdings Comparison Service: Are We Really Insured? *Serials*, 22(2), 174-177.
- nestor–Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung. (2006). *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version 1 (Entwurf zur öffentlichen Kommentierung)*. Frankfurt-am-Main: nestor. Consulté à l'adresse <http://edoc.hu-berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf>
- nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. (2010). Consulté 24 mai 2012, à l'adresse http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf
- Nielsen, A. B., Thirifays, A., & Kejsler, U. B. (2012). Costs of archival storage. In *Archiving 2012 - Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organizations, Final Program and Proceedings* (p. 205-210).

- NISO SERU Standing Committee (mai 2012). SERU: A Shared Electronic Resource Understanding. NISO RP-7-2012. Consulté à l'adresse http://www.niso.org/publications/rp/RP-7-2012_SERU.pdf
- Oltmans, E. (2004). Cost Models in Digital Archiving: An overview of Life Cycle Management at the National Library of the Netherlands. *LIBER Quarterly*, 14(3/4). Consulté à l'adresse <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-113393>
- Oltmans, E., & Kol, Nanda. (2005). A Comparison Between Migration and Emulation in Terms of Costs. Consulté à l'adresse <http://worldcat.org/arcviewer/2/OCC/2009/08/11/H1250012115408/viewer/file2.html>
- Ostrowski, F., & Sieber, J. (2012). LOCKSS im Rahmen des LuKII-Projekts. Technische Erweiterungen für Datenaustausch, Metadatenverwaltung und die prophylaktische Formatmigration. (German). *Bibliothek Forschung und Praxis*, 36(1), 135-142. doi:10.1515/bfp-2012-0017
- Palaiologk, A. S., Economides, A. A., Tjalsma, H. D., & Sesink, L. B. (2012). An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS. *International Journal on Digital Libraries*, 12(4), 195-214. doi:10.1007/s00799-012-0092-1
- Ras, M. (2009). The KB e-Depot: Building and Managing a Safe Place for e-Journals. *Liber Quarterly: The Journal of European Research Libraries*, 19(1), 44-53.
- Reference model for an Open Archival Information System (OAIS). Recommendation for Space Data System Practices.* (2012). Consultative Committee for Space Data System. Consulté à l'adresse <http://wiki.digitalrepositoryauditandcertification.org/pub/Main/WebHome/652x0r1candidate-update-typocorrected.doc>
- Reich, V. (2008). CLOCKSS--It Takes a Community. In *Serials Librarian* (Vol. 54, p. 135-139). Taylor & Francis Ltd. Consulté à l'adresse <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=33247887&site=ehost-live>
- Reich, V., & Rosenthal, D. (2009). Distributed digital preservation: Lots of copies keep stuff safe. In *Proceedings Indo-US Workshop on International trends in Digital Preservation March* (p. 24–25). Consulté à l'adresse <http://humgroup.lockss.org/locksswiki/files/ReichIndiaFinal.pdf>
- Reich, V., & Rosenthal, D. (2009). Distributed Digital Preservation: Private LOCKSS Networks as Business, Social, and Technical Frameworks. *Library Trends*, 57(3), 461-475. doi:10.1353/lib.0.0047
- Reilly Jr., B. F. (2008). Summary of the Test Audits Portico and LOCKSS. *Charleston Advisor*, 9(3), 61-62.
- Research Information Network, & JISC. (2010). *E-only scholarly journals: overcoming the barriers*. Consulté à l'adresse http://www.publishingresearch.net/documents/E-only_reportRIN2010.pdf
- Research Libraries Group. (2002). *Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report*. Mountain View, CA. Consulté à l'adresse <http://oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
- RLG-OCLC. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA. Consulté à l'adresse <http://oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
- Rogers, S. (2009). Survey and Analysis of Electronic Journal Licenses for Long-Term Access Provisions in Tertiary New Zealand Academic Libraries. *Serials Review*, 35(1), 3-15. doi:10.1016/j.serrev.2008.11.002
- Rosenthal, D. S. (2014). Architectural Choices in LOCKSS Networks (preprint, à paraître). Consulté à l'adresse <http://www.lockss.org/locksswp/wp-content/uploads/2013/12/ArchitecturalChoices.pdf>
- Rosenthal, D. S. H., Lipkis, T., Robertson, T. S., & Morabito, S. (2005). Transparent Format Migration of Preserved Web Content. *D-Lib Magazine*, 11(01). doi:10.1045/january2005-rosenthal

- Rosenthal, D. S. H., Robertson, T., Lipkis, T., Reich, V., & Morabito, S. (2005). Requirements for Digital Preservation Systems. *D-Lib Magazine*, 11(11). doi:10.1045/november2005-rosenthal
- Rosenthal, D. S., & Vargas, D. L. (2013). Distributed digital preservation in the cloud. Présenté à 8th International Digital Curation Conference, Amsterdam. Consulté à l'adresse <http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf>
- Rosenthal, D. S., Rosenthal, D. C., Miller, E. L., Adams, I. F., Storer, M. W., & Zadok, E. (2012). The economics of long-term digital storage. In *Memory of the World in the Digital Age Conference, Vancouver, BC*. Retrieved from <http://www.lockss.org/locksswp/wp-content/uploads/2012/09/unesco2012.pdf>. Consulté à l'adresse http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Rosenthal_et_al_27_B_1330.pdf
- Rosenthal, D., Robertson, T., Lipkis, T., Morabito, S., & Reich, V. (2005). Requirements for Digital Preservation Systems. A bottom-up approach. *D-Lib Magazine*, 11(11). Consulté à l'adresse <http://lockss.org/locksswiki/files/DlibRequirements.pdf>
- Rusbridge, A. (2012, janvier 31). *UK LOCKSS Alliance: Today's scholarly content, secured for tomorrow*. Présenté à Trust and e-Journals, London. Consulté à l'adresse http://dpconline.org/component/docman/doc_download/731-trust-and-e-journals-2012-rusbridge
- Rusbridge, A., & Ross, S. (2008). Establishing a community-based approach to electronic journal archiving: the UK LOCKSS Pilot Programme (p. 8). Présenté à iPRES 2008 Fifth International Conference on Preservation of Digital Objects. Joining up and working: Tools and Methods for Digital Preservation, 29-30th September 2008, British Library, London. Consulté à l'adresse <http://eprints.gla.ac.uk/4635/1/4635.pdf>
- Schmidt, K. (2013, juin 18). *Der Langzeitarchivierungsservice der Deutschen Nationalbibliothek*. nestor Praktiker-Tag 2013, *Dienstleistungen und Kooperationen für die digitale LZA*. Consulté à l'adresse <http://files.dnb.de/nestor/veranstaltungen/Praktikertag2013/2013-06-dnb-schmitt.pdf>
- Schonfeld, R. C., King, D. W., Okerson, A., & Fenton, E. G. (2004). Library Periodical Expenses. *D-Lib Magazine*, 10(1). doi:10.1045/january2004-schonfeld
- Schonfeld, R. C., King, D. W., Okerson, A., & Gifford Fenton, E. (2004). *The Nonsubscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats* (research report). Washington, D.C.: Council on Library and Information Resources. Consulté à l'adresse <http://www.clir.org/pubs/reports/pub127/pub127.pdf>
- Schroger, A. (2013, juin 18). *Die Einführung von Rosetta an der Bayerischen Staatsbibliothek als Basis für Langzeitarchivierungs-services*. nestor Praktiker-Tag 2013, *Dienstleistungen und Kooperationen für die digitale LZA*. Consulté à l'adresse <http://files.dnb.de/nestor/veranstaltungen/Praktikertag2013/2013-06-bsb-schroger.pdf>
- Seadle, M. (2010a). Archiving in the Networked World: Interoperability. *Library Hi Tech*, 28(2), 189-194.
- Seadle, M. (2010b). Archiving in the Networked World: LOCKSS and National Hosting. *Library Hi Tech*, 28(4), 710-717.
- Seadle, M. (2011). Archiving in the networked world: by the numbers. *Library Hi Tech*, 29(1), 189-197. doi:10.1108/07378831111117001
- Seadle, M. (2011a). Archiving in the networked world: metrics for testing. *Library Hi Tech*, 29, 557-564.
- Seadle, M. (2011b). Archiving in the networked world: Open access journals. *Library Hi Tech*, 29(2), 394-404.
- Seadle, M. (2012). Archiving in the networked world: Authenticity and integrity. *Library Hi Tech*, 30(3), 545-552. doi:10.1108/07378831211266654

- Skinner, K., Schultz, M., & MetaArchive Cooperative (U.S.). (2010). *A guide to distributed digital preservation*. Atlanta, Ga.: Educopia Institute.
- Slats, J., & Verdegem, R. (2005). Cost Model for Digital Preservation. In *Proceedings of the IVth triennial conference, DLM Forum, Archive, Records and Information Management in Europe*. Consulté à l'adresse http://dilmforum.typepad.com/Paper_RemcoVerdegem_and_JS_CostModelfordigitalpreservation.pdf
- Sparks, S., Look, H., Bide, M., & Muir, A. (2010). A registry of archived electronic journals. *Journal of Librarianship and Information Science*, 42(2), 111 -121. doi:10.1177/0961000610361552
- Stemper, J., & Barribeau, S. (2006). Perpetual Access to Electronic Journals: A Survey of One Academic Research Library's Licenses. *Library Resources & Technical Services*, 50(2), 91-109.
- Stephen A. (1994), *The Application of Life Cycle Costing in Libraries: A Case Study Based on Acquisition and Retention of Library Materials in the British Library*, IFLA Journal, 20(2), 1994, 130-140.
- UKSG. (2008). TRANSFER Code of Practice : version 2.0. Consulté à l'adresse http://www.uksg.org/sites/uksg.org/files/TRANSFER_Code_of_%20Practice_2_0.pdf
- Verheul, I. (2006). *Networking for digital preservation : current practice in 15 national libraries*. Munich: K.G. Saur. Consulté à l'adresse <http://www.ifla.org/files/hq/publications/ifla-publications-series-119.pdf>
- Waller, A., & Bird, G. (2006). « We Own It »: Dealing with "Perpetual Access in Big Deals. *Serials Librarian*, 50(1/2), 179-196.
- Wang, J., & Schroeder Jr., A. T. (2005). The Subscription Agent as E-journal Intermediary. *Serials Review*, 31(1), 20-27. doi:10.1016/j.serrev.2004.11.007