

電子ジャーナルの長期保存—LOCKSS と Portico

時 実 象 一*

電子ジャーナルの長期保存は災害、出版社の業務停止、購読終了時のアクセス保障などの観点から最近関心を集めている。米国スタンフォード大学のプロジェクトとして始まった LOCKSS と、その応用である CLOCKSS、および米国の電子ジャーナルサイト JSTOR が中心となって開発した Portico についてその概要を紹介し比較する。

キーワード：電子ジャーナル、長期保存、LOCKSS、CLOCKSS、Portico

1. なぜ長期保存が必要か

学術出版の分野では電子ジャーナルは必須のものとなった。いまや電子ジャーナルとして提供されていない学術雑誌は存在しないと同等であるといえる。紙の雑誌は人が何の装置もなしに容易に読むことができる。しかし、電子データは人間が直接手にすることはできないし、ハードウェアやソフトウェアがなくては再現することができない。そしてそのハードウェアやソフトウェアは大変な速度で変化している。現在自分が持っている電子データが 10 年 20 年先にも同じように読めるという保証はない。

1980 年に PC-9801 を契機としてパソコンが立ち上がったとき、筆者は J-Word というワープロを購入した。しかし、その文書を現在の MS Word で読むことは不可能である。その後普及した「一太郎」や「Lotus 1-2-3」のファイルはかろうじて「MS Word」と「Excel」で読むことができることになっている。しかし、実際には罫線が崩れたり、表が紙面をはみ出したりして、もとのレイアウトが完全に再現できないことも多い。

HTML や PDF はインターネット上を自由に流通し、ほとんどのコンピュータ環境で閲覧ができる。しかし、これとて永遠のものではない。昔の HTML データは今のブラウザで閲覧しようとするとしばしば文字化けを起こすが、これは昔のデータが今のブラウザが要求する規格を満足していないためである。HTML に埋め込まれたプラグインなども将来にわたって動作が保証されるわけではない。PDF にもさまざまなバージョンがあり、さらには特定のメーカーの拡張バージョンもある。これらさまざまなバージョンで作成された PDF 文書がすべて将来の PDF リードにおいて正確に再現できるかどうかは約束できない。

また、電子データを保存するメディアも問題である。紙は燃えない限り何世紀も保存が可能である。マイクロフ

ィルムなどの写真メディアも 100 年以上の保存が可能であることが証明されている。一方磁気媒体や光ディスクなどの電子データのメディアの寿命はたかだか 10 年といわれており、長期に保存するためには頻繁なバックアップやメディア変換が必要である。

今大学等の図書館の主流は電子ジャーナルとなり、企業図書館などでは冊子の購読を中止するところが増えている。また、出版社もいずれは紙での出版を中止すると思われる。そのような背景から、電子ジャーナルの長期保存は重大関心となってきている。

2. 長期保存の諸問題

オランダ国立図書館の Hans Jansen によれば¹⁾、電子ジャーナルの保存に関して関係者はそれぞれ懸念を持っている。まず図書館は、(1)現在購読している雑誌が購読を中止したときには読めなくなるのではないかと、(2)出版社や提供サイトが何らかの理由で閉鎖されたら読めなくなるのではないかと、(3)現在の機器や技術が使えなくなった際にも今の電子ジャーナルの内容を読むことができるか、という懸念を持っている。論文の著者たちは、(1)自分の論文は永久に提供されるのだろうか、(2)自分の論文の内容が改変されないという保証はあるのだろうか、と心配している。さらに出版社は、現在出版している雑誌を将来に渡っても提供を保証できると断言するにはどうしたらいいだろうかと考えている。

また、スタンフォード大学図書館の Michael Keller によれば²⁾、長期保存を誰が担当するかが問題である。たとえば次のような担当者が考えられるが、それぞれに問題がある。

- a. ISP (Internet Service Provider) テープアーカイブ (テープの検索をどうするか、またすぐにはデータを取り出せない)
- b. 出版社 (監査をどうするか、継続性は保証されるか)
- c. 信頼できる第三者 (もとのイメージ (look & feel) を保存できるか、監査はどうするか、継続性は保証されるか、費用が高くないか)

*ときざね そういち 愛知大学文学部図書館情報学専攻

〒441-8522 愛知県豊橋市畑町 1-1

Tel. 0532-47-4467

tokizane@aichi-u.ac.jp

(原稿受領 2007.12.8)

- d. 各国の国立図書館（購読者のアクセスは保証されるか、国境の壁はないか）
- e. 大学図書館（コンテンツの安全は保証できるか、開発や維持のコストを負担できるか）
- f. コンソーシアム（開発や維持のコストを負担できるか）
- g. 政府機関（提供範囲が限定されないか、政治的に利用されないか、予算は保証されるかー米国環境保護庁で予算カットで図書館が閉鎖された）

このように見てくると、出版社との契約、情報の共有、参加、アーカイブとの契約、など複数の方法を併用することが好ましいと Keller は主張している。担当者を選択するには、

- a. 透明性（活動、契約、費用、特性）
- b. 監査と公開報告
- c. アプローチの多様性

などが要件として求められるとしている。

3. 長期保存に関する研究開発

長期保存についての初期の研究開発動向については松村氏がまとめている³⁾。米国 RLG と OCLC は 2000 年から電子情報の長期保存の調査をおこない、2002 年に報告書を発表している⁴⁾。これはその後の研究開発の基礎となった。米国 Andrew W. Mellon 財団では 2002 年に電子ジャーナルの保存方法を研究するための研究助成をおこなった⁵⁾。研究をおこなったのはコーネル、ハーバード、MIT、ペンシルバニア、スタンフォード、ユールなどの大学図書館とニューヨーク公共図書館である。スタンフォード大学図書館のチームは後述の LOCKSS の研究開発をおこなっている。また後述の Portico もこの研究助成から生まれている。

Kenney らが 2006 年に米国図書館情報資源評議会 (Council on Library and Information Resources: CLIR) へ提出した報告書「電子ジャーナル・アーカイブ：その境界：地平の調査 (E-Journal Archiving: Metes and Bounds: A Survey of the Landscape)」⁶⁾はこの問題についてのもっとも詳細な調査報告書である。ここでは電子ジャーナルの保存に関わる 12 件のプロジェクト (CISTI (カナダ), CLOCKSS, OCLC ECO, OhioLINK EJC, オランダ国立図書館の e-Depot, ドイツ国立図書館の kopal, LANL-RL, LOCKSS Alliance, Ontario Scholars Portal (カナダ), PANDORA (オーストラリア), Portico, PubMed Central) がさまざまな観点から比較されている。この中には PubMed Central のように特に長期保存のための技術を持たない通常のリポジトリも含まれている。長期保存のための技術開発をおこなっているプロジェクトとしては KB e-Depot^{7,8)}, kopal/DDB⁹⁾, PANDORA¹⁰⁾, LOCKSS/CLOCKSS¹¹⁾, Portico¹²⁾があげられる。

本記事では、上記で取り上げられたプロジェクトのうち LOCKSS/CLOCKSS と Portico について詳細に紹介する。なお後藤氏の記事^{13,14)}も参考とされたい。

4. LOCKSS と CLOCKSS

4.1 LOCKSS

LOCKSS (Lots of Copies Keep Stuff Safe) はスタンフォード大学図書館の Vicky Reich らが 1999 年頃始めたプロジェクトである。電子ジャーナルを購読していても、契約が切れるとアクセスが不可能になり、図書館としては「資料を保存する」という役割が果たせないことになるし、また出版社が倒産・解散した場合、アクセスが不可能になる。このような問題を解決するために、複数の図書館で電子ジャーナルのキャッシュを保存し、共有するという構想から開発をおこなった。2007 年 6 月にスタンフォード大学に Reich を訪問する機会があったので、その際得た知見を中心に紹介する。

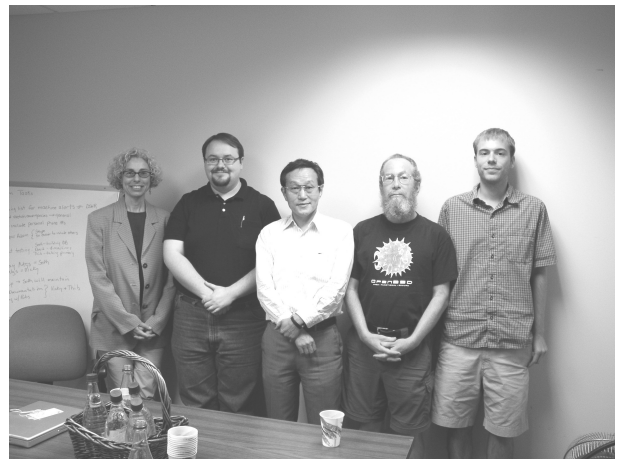


写真 1 LOCKSS のスタッフと筆者（一番左が Vicky Reich）

LOCKSS の仕組みは次のとおりである。各図書館は LOCKSS Box というものを用意するが、これは汎用の PC に汎用のディスクをつけたものである（図 1）。OS は OpenBSD, LOCKSS のソフトは Java で、どちらもオープン・ソフトウェアなので、誰でもダウンロードしてテストができる。UNIX の知識は必要ないので、小さい図書館

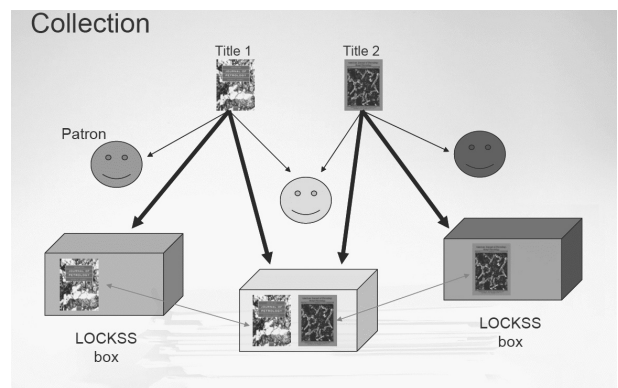


図 1 LOCKSS Boxes
(LOCKSS はデータを LOCKSS Box にキャッシュする (Keller²⁾)

でも運用可能であり、現在世界約 200 の図書館と 200 の出版者が参加しているとのことである。

LOCKSS のソフトは LOCKSS を承認した電子ジャーナル (LOCKSS-compliant) サイトをクロールし、電子ジャーナルページと PDF などアクセス可能なファイルを LOCKSS Box にキャッシュする。画像、音声、ビデオなどファイルの種類は問わないが、ストリーミング・ビデオや FLASH 中に埋め込まれたリンクなど、一部対応できないものもある。これらキャッシュされたページは次の場合に使用される。

- (1)電子ジャーナルサイトがメンテナンスなどで一時的にアクセスできない場合
- (2)ネットワーク障害などの場合
- (3)その電子ジャーナルが提供を中止したりした場合
- (4)図書館と出版者との契約が終了した場合は、購読していた期間中のデータは引き続きアクセスできる (出版社の許諾による)

この中で(4)が特に重要と思われる。

LOCKSS では「アクセス時形式変換 (migration on access)」方式を採用している¹⁵⁾。これは元のデータ (Web で提供されている HTML または XML) はそのまま保存し、アクセス要求があったときに、それを規格化された形式に変換する。その場合なるべく元のサイトの表現 (look and feel) を再現できるようにしている (図 2, 3)。

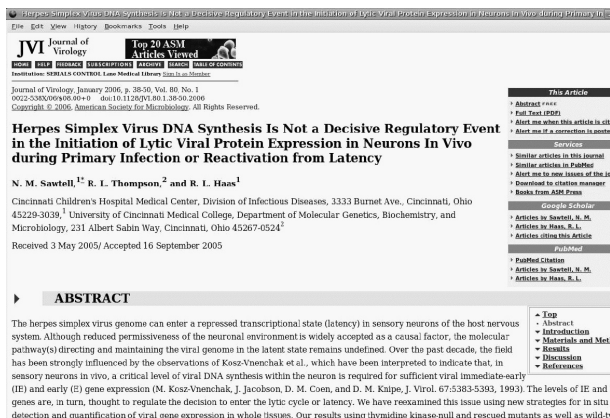


図 2 LOCKSS の元データ (Keller²⁾)

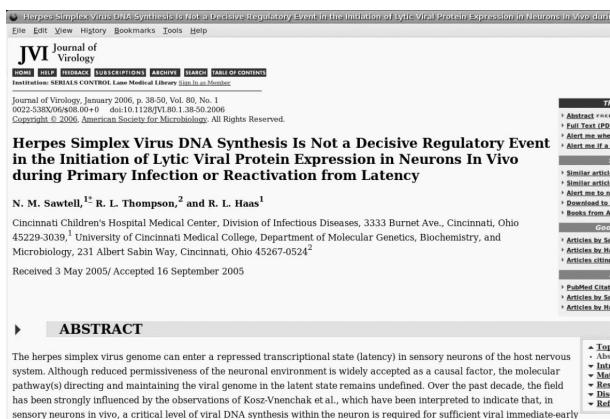


図 3 LOCKSS で再現したデータ (Keller²⁾)

LOCKSS は一種のプロキシ・サーバなので、利用者はアクセスが出版者サイトなのか、LOCKSS Box なのかを知る必要がない。またサイトの過去データに更新があったときも、時間がかかるが、最終的には LOCKSS のデータは更新される。また複数の LOCKSS Box 間でデータの相互チェックをおこなっているため、ディスク障害があったときもバックアップ可能である。実際出版社が LOCKSS のデータを復旧に用いたこともあるとのことであった。

LOCKSS に参加する出版社は電子ジャーナルページに次のメッセージを記載する。

“LOCKSS system has permission to collect, preserve, and serve this Archival Unit”

または Creative Commons のライセンス表示 (自由利用許諾) でもよい。これにより、LOCKSS がデータを自動的にキャッシュすることができ、図書館による上記の利用が可能になる。電子ジャーナルの利用契約や著作権そのものには変更がない。LOCKSS においては過去分の保存を図書館に許諾するので、現時点では商業出版社はほとんど参加していない、参加しているのは学会出版社と Oxford University Press などの大学出版者である。出版者のメリットとしては、図書館における認知度が高まることと、契約終了後のアクセス管理が不要なことである。

図書館が LOCKSS のメリットを享受するためには、会員 (Alliance) になる必要がある。会費は図書館の規模に応じて \$ 1,000-10,800 である。非会員でも若干のコンテンツはアクセス可能である。現在 200 図書館のうち会員は半数強である。なお著作権についての懸念から中国の図書館は会員として認めていない (現在は清華大学、中国科学院図書館が加盟しているようである)。また企業図書館の参加は認めていない。なお英国では現地の Digital Curation Center (DCC) が LOCKSS のサポートを代行している。

4.2 CLOCKSS

LOCKSS の技術を利用した大規模保存プロジェクトとして CLOCKSS (Controlled LOCKSS) がある。これは下記の大手商業出版社と学会出版社 11 社が参加している。

American Chemical Society
American Medical Association
American Physiological Society
Elsevier
Institute of Physics
Nature Publishing Group
Oxford University Press
SAGE Publications
Springer
Taylor and Francis
Wiley-Blackwell

このプロジェクトでは世界中の最大 15 館までの図書館と契約して、自分たちのデータを分散保存 (Dark Archive) しようというものである。現在参加が確定している図書館は下記の 7 館である。

Indiana University
New York Public Library
OCLC Online Computer Library Center
Rice University
Stanford University
University of Virginia
University of Edinburgh

安全のため、なるべく世界に分散して保存することを希望しており、日本やオーストラリアの図書館にも参加を求めていく考えである。

CLOCKSS のデータは通常は公開されないが、出版者の倒産、戦争などの提供不可能事態 (Trigger Events) が発生した場合は参加図書館を通して公開される。CLOCKSS は一種の私的な LOCKSS であり、その方針は参加出版社からなる理事会によって決定される。

5. Portico

JSTOR が始めた電子アーカイブ・サービス Portico について Fenton の発表を中心に紹介する¹⁶⁻²⁰⁾。ちなみに portico とは玄関のポーチのことである。

JSTOR では 1999 年に Andrew W. Mellon 財団の資金を得てアーカイブの研究を開始した (e-Archive Initiative)。ここでは大学図書館と 10 学術出版社に対してパイロットをおこなった。試行では、多くの技術課題や、協力の必要性、多方面からの資金提供などについて教訓を得たとしている。このプロジェクトがスピナウトして 2005 年に Portico となった。Portico は JSTOR と Ithaka によって設立され、Andrew W. Mellon 財団、米国議会図書館 NDIIPP などが資金提供している。

Portico はセンター方式のアーカイブである (図 4)。出版社は Portico に電子ジャーナルの元ファイル (source files, SGML, XML, PDF など、他にデータがない場合は HTML) を提供する。Portico はこれを NLM Archival DTD に基づき規格化された (normalized) データ (PorticoMETS) に変換して保存する。雑誌の知的内容を保存し、若干の内部リンクなどは保存するが、誌面イメージは保存されない。規格化データは時代の変遷に合わせ変

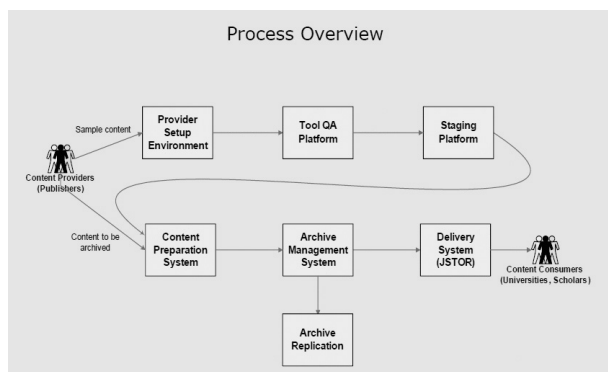


図 4 Portico の処理工程概要 (Owens²⁰⁾)

換していく。元ファイルはそのまま保存する。NLM DTD や JHOVE などのオープンソースをできるだけ使用する。

通常のコンテンツ提供は出版社がおこなうが、万一出版社がデータを提供できない事態が発生した場合 (trigger event) には、Portico が参加図書館に代理提供を開始する。提供不可能事態には次のものがある。

- (1) 出版社が機能を停止し、電子ジャーナルが提供されなくなる
- (2) 出版社がある電子ジャーナルの発行を停止し、それが他の出版社等からも提供されなくなる
- (3) 過去分が出版社のサイトから撤去され、他からも提供されていない
- (4) 災害等によって出版社からの提供システムに長期にわたって使えなくなる
- (5) 出版社が自分での提供をおこなわず、Portico に提供を依頼する (図書館の購読終了後などの無期限 (perpetual) アクセスなど)

また、契約図書館の司書は、提供不可能事態発生以前であっても内容確認をおこなうことができる。

財政的には出版社および図書館からの持続的支援が必要である。JSTOR, Andrew W. Mellon 財団, Ithaka, 米国議会図書館 (Library of Congress: LC) が投資している。

Portico に参加している出版社は Elsevier, Wiley, Nature, Springer, Taylor and Francis などの商業出版社, Oxford University Press, University of Chicago Press などの大学出版社, American Chemical Society, IEEE, Institute of Physics Publishing などの学会出版社など 45 社が参加しており、また図書館の会員は 386 館である (2007 年 11 月現在)。出資は出版社、図書館、政府などが規模に応じておこなう。たとえば出版社の場合 \$ 250-\$75,000/年、図書館の場合 \$1,500-\$24,000 である¹⁸⁾。図書館が支払いを停止すると脱会となり、Portico のサービスは受けられなくなる。

6. おわりに

LOCKSS と Portico を比較すると次のようになる。

	LOCKSS/CLOCKSS	Portico
開発者	Stanford University	Portico
保存場所	分散	集中
データの加工	おこなわない、表示の際に標準化	標準データに加工
ハードウェア	汎用 PC	センター・サーバ
利用形態	キャッシュ	ダークアーカイブ
購読終了後アクセス・オプション	あり	あり
コスト	低め	高め

このように、この 2 者はアプローチが非常に対照的で興味深い。なお主要商業出版社や大図書館は CLOCKSS と Portico の両方に参加しており、いわば保険をかけている

状態といえる。欧州ではオランダの国立図書館の e-Depot が IBM と協力して運営されており、Elsevier 社はこれを公式アーカイブの 1 つとしている⁹⁾。ドイツ国立図書館はこの e-Depot を利用する予定であったが、予算難のため現在あまり進んでいないと伝えられている。

わが国では電子ジャーナルの大出版社が存在しないためか、これまで長期保存についての議論は少なかった。しかし逆に、わが国は海外の商業・学会出版社に情報を頼る割合が極めて高いので、何かことが起きた場合には重大な問題となる。わが国でも今後積極的に電子ジャーナルの長期保存に関する議論を起こしていく必要があると考える。

参 考 文 献

- 1) Hans Jansen. Permanent access to electronic journals. Information Services & Use. 2006, 26, p.129-134, 129, <http://iospress.metapress.com/index/7DRBY91R8T4GF8A> P.pdf [accessed 2007-11-06].
- 2) Michael Keller. The Future of Digital Archiving...starting with the present. CrossRef Annual Meeting, 1 November 2006. <http://www.crossref.org/10meetings/2006/The%20Future%20of%20Digital%20Archiving.pdf> [accessed 2007-11-06].
- 3) 松村多美子. デジタル・アーカイビング電子ジャーナルを中心に. 情報管理. 2001, vol.44, no.9, p.622-628.
- 4) RLG. Trusted Digital Repositories: Attributes and Responsibilities, An RLG-OCLC Report. May 2002. <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf> [accessed 2007-11-17].
- 5) Andrew W. Mellon Foundation. Archiving Electronic Journals. <http://www.diglib.org/preserve/ejp.htm> [accessed 2007-11-13].
- 6) Anne R. Kenney, Richard Entlich, Peter B. Hirtle, Nancy Y. McGovern, Ellie L. Buckley. E-Journal Archiving: Metes and Bounds: A Survey of the Landscape. Council on Library and Information Resources. September 2006. <http://www.clir.org/pubs/reports/pub138/pub138.pdf> [accessed 2007-11-06].
- 7) e-Dopot. <http://www.kb.nl/e-depot> [accessed 2007-11-17].
- 8) Johan F. Steenbakkens. オランダ国立図書館における電子ジャーナルのアーカイビング (インタビュー). 情報管理. 2004, vol.47, no.6, 387-392.
- 9) kopal. <http://kopal.langzeitarchivierung.de/> [accessed 2007-11-17].
- 10) PANDORA, Australia's Web Archive. <http://pandora.nla.gov.au/> [accessed 2007-11-17].
- 11) LOCKSS. <http://www.lockss.org/> [accessed 2007-11-17].
- 12) Portico. <http://www.portico.org/> [accessed 2007-11-17].
- 13) 後藤敏行. 電子ジャーナルのアーカイブ - アクセスの観点からみた集中・分散の 2 方面戦略. 情報管理. 2005, vol.48, no.8, p.509-520.
- 14) 後藤敏行. 電子ジャーナルのアーカイビング: 論点, 動向, 将来展望. 図書館界. 2007. vol.58, no.6, p.320-331. <https://www.tulips.tsukuba.ac.jp/dspace/handle/2241/90928> [accessed 2007-11-12].
- 15) David S. H. Rosenthal, Thomas Lipkis, Thomas S. Robertson, Seth Morabito. Transparent Format Migration of Preserved Web Content. D-Lib Magazine. January 2005, vol.11, no.1. <http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html> [accessed 2007-11-11].
- 16) Eileen Fenton. History and Overview of Portico: A New Electronic Archiving Service. CrossRef Annual Meeting, 15 November 2005. <http://www.crossref.org/10meetings/2005/Portico.Overview.111505.FINAL.pdf> [accessed 2007-11-06].
- 17) Kirchhoff, Amy, and Eileen Fenton. Archiving Electronic Journals: An Overview of Portico's Approach. Papers from Portico. Portico, 2006. <http://www.portico.org/news/PapersFromPortico.1.Overview.pdf> [accessed 2007-11-11].
- 18) Eileen Fenton. Preserving Electronic Scholarly Journals: Portico. Ariadne. 2006, no.47. <http://www.ariadne.ac.uk/issue47/fenton/> [accessed 2007-11-13].
- 19) Eileen Gifford Fenton. An Overview of Portico: An Electronic Archiving Service. Serials Rev. 2006, vol.32, no.2, 81-86. <http://www.portico.org/news/Fenton.Serials%20Review%20article.031606.pdf> [accessed 2007-11-21].
- 20) Evan Owens, Suku Sukumar, John Meyer, David Copperman. A Format-Registry-Based Automated Workflow for the Ingest and Preservation of Electronic Journals. Digital Library Federation Fall Forum 2005 Charlottesville, Virginia. http://www.portico.org/news/Portico_DLF_Fall2005.pdf [accessed 2007-11-17].

Long-term preservation of electronic journals: LOCKSS and Portico. Soichi TOKIZANE (Library and Information Science, Faculty of Letters, Aichi University, 1-1 Machihata-cho, Toyohashi, Aichi Pref. 441-8522 JAPAN)

Abstract : Long-term preservation of electronic journals is a hot topic among libraries and publishers. It will preserve electronic journal contents so that they will continue to be available even in the case of disaster, termination of services from publishers, and termination of subscriptions (perpetual access). Two system, i.e., LOCKSS developed by the team of Stanford University Library, and its application product, CLOCKSS, and Portico, developed by JSTOR, etc., are discussed.

Keywords : electronic journals / long-term preservation / LOCKSS / CLOCKSS / Portico