February 2009

Springer's eBook Preservation Strategy

Heather Ruland Staines

SpringerLink, heather.staines@springer.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Staines, Heather Ruland (2009) "Springer's eBook Preservation Strategy," *Against the Grain*: Vol. 21: Iss. 1, Article 12.

DOI: <https://doi.org/10.7771/2380-176X.2495>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Springer's eBook Preservation Strategy

by **Heather Ruland Staines** (Global eProduct Manager, SpringerLink, Springer Science + Business Media, 233 Spring Street, New York, NY 10013; Phone: 646-201-8474) <heather.staines@springer.com> www.springer.com www.springerlink.com

Just two weeks into my job as Global eProduct Manager of **SpringerLink** for **Springer Science + Business Media**, I discovered that my new responsibilities would include our digital preservation efforts. With a publications program that oversees production of some 2,000 journals and more than 3,500 eBooks annually, and an ever-increasing number of researchers, scientists, and students accessing scientific content online, the importance of such projects was immediately clear to me. I soon learned, however, that the universe of digital preservation initiatives is more complex than I would have believed. Neither our licensing managers nor our library partners have time to research this vital issue in-depth. I quickly found myself learning and educating others about digital preservation of e-journal content. It was not long before I found myself on the receiving end of another preservation question: what is **Springer's** policy in regards to eBooks?

As a digital preservation neophyte, I understood the concern that, as journals moved increasingly — and perhaps entirely — into an e-only production mode, print versions might not be available as a fallback option should anything go wrong with an e-distribution model. The term “eBook” though at least implied the existence of a “p-book” somewhere, and this no doubt explains why digital preservation initiatives have only recently begun to examine the rationale and ramifications for ingesting eBook content. “eBooks should absolutely be preserved,” says **Sue Polanka**, Head of Reference at **Wright State University** and author of the blog on eBook issues *No Shelf Required*. “The philosophy of the librarian or archivist is to retain copies of anything necessary for research purposes. A change in format doesn't make that need go away.”

Springer is the largest STM book publisher, and the first to move its complete book portfolio online. **Springer** launched its eBook program in 2006 and has, since that introduction, seen increasing adoption and usage rates of eBook content. **Cynthia Cleto**, **Springer's** Global eProduct Manager for eBooks, notes a trend in online usage: “eBooks, which constitute only 15% of content on **SpringerLink**, already generate 25% of its total usage, after a few years in existence.” The digital generation of students expects online access to content — making no discernable distinction between articles and book chapters — regardless of their library's hours or their geographical location. As library renovation projects are setting aside smaller areas for physical collections, eBooks often solve library space-allocation issues. eBooks are easily found via Internet search engines, electronic resources

management tools (ERMs), and library OPACs (online public access catalogs). All of these factors render eBooks an ever-more important part of a university's collection development investment. Thus, for the library community digital preservation of eBook content is essential.

Libraries around the world share a similar concern. **Anthony Ferguson**, University Librarian at the **University of Hong Kong**, sums up many concerns:

“[Our library] has purchased upwards of 1.6 million eBooks. It is critical that this content be preserved. For about 400,000 Chinese language eBooks and 5,000 e-journals we serve as mirror site partners with the publishers but realize that if the publishers were to disappear in time we would be left with a huge responsibility to maintain and develop the software needed to access these materials. Other publishers promise that they would provide digital copies of the content but we would have to confront the software problem for these as well. The problems of preserving e-content, as well as giving long life to the stop-gap solutions we are employing, are too big for a single library to tackle. We need to partner with other libraries and the publishers themselves to provide long-term, sustainable solutions.”

Denise Koufogiannakis, Collections and Acquisitions Coordinator at **University of Alberta**, also stresses the importance of preservation initiatives when she works with vendors. “If they can show that they are working towards this type of community effort to preserve eBooks, it gives me a great deal more trust in the longevity of our purchase, as opposed to those vendors/publishers that are only focused on internal company solutions.”

This article is meant to provide an overview of **Springer's** participation in major digital preservation initiatives and to explore some of the technical and legal issues that surround preservation of eBook content. Due to limited space, I will focus on four initiatives: **CLOCKSS**, **Portico**, the **German National Library**, and **Koninklijke Bibliotheek** (National Library of the Netherlands).

Major Preservation Initiatives

My introduction to digital preservation began with a board meeting of **CLOCKSS** (Controlled Lots of Copies Keep Stuff Safe). As a Governing Publisher, **Springer** had been involved in the successful **CLOCKSS Pilot Project** which was concluding in spring of 2008. During the next six months, as **CLOCKSS** prepared to incorporate as a non-

profit, I learned a new vocabulary of “dark archives,” “trigger events,” and “archive nodes.” Conceived of as a collective of concerned libraries and publishers, **CLOCKSS** had during its pilot focused exclusively on the ingest of journal content, but during a subsequent board meeting, the question was raised: would **CLOCKSS** also accept eBook content?

My preservation education would continue once I made contact with **Portico**, the electronic archiving service of **Ithaka Harbors, Inc.**, which launched in 2005. **Springer** had signed an agreement with **Portico** in the fall of 2007 and began depositing **Springer** journal content shortly thereafter. At the 2008 **American Library Association (ALA)** Annual meeting in Anaheim, **Portico** announced that it would be accepting eBook content. **Eileen Fenton**, Executive Director of **Portico**, explains: “Libraries increasingly expect publishers will have established reliable preservation arrangements for eBooks in much the same way that they have for e-journals. As libraries respond to the growing demand from students and faculty for digital content, the expectation is that preservation and adoption of the eBook genre will unfold in a much more synchronized way than was the case for e-journals.” I was eager to find out more. Certainly, we would soon be fielding customer questions about our own participation. Discussions and preparations began immediately to determine whether **Springer** eBook content could be successfully ingested by **Portico** and to address any issues beyond our existing cooperation on preserving journal content.

Again and again, I would find myself explaining to my colleagues why **Springer** was participating in both **CLOCKSS** and **Portico**. The two preservation solutions have different models, both in terms of funding and in terms of content availability if a “trigger event” should occur. Listening to our library partners, **Springer** decided that it was important for us to participate in both of these programs with our eBook content. As we would learn, however, eBooks are different animals than e-journals, and our approach would require great care and attention to our contractual obligations.

In addition to **CLOCKSS** and **Portico**, **Springer** is mandated to deposit versions of record with designated repositories. Since 2005, we have contributed journal content to the **National Library of the Netherlands's e-Depot** project at the **Koninklijke Bibliotheek (KB)**. In contrast to the other programs mentioned above, the **e-Depot** at the **KB** is not a dark archive. As a national library, it provides onsite access to pass holders. The **KB** does not allow Internet access, nor can a researcher print out an article. Thus, this initiative serves as a living library of electronic content with each access by a scholar acting as a test to ensure that content remains renderable. **Marcel Ras**,

continued on page 30

the Manager of **e-Depot**, informed me that, in accordance with its overall mission, the **KB** would begin preserving eBook content in 2009.

The fourth major initiative with which **Springer** is associated is the **German National Library (Deutsche Nationalbibliothek, or DNB)** program which began collecting our content in 2003 and began to ingest eBooks in 2007. Along with other institutions, they are tackling the issue of preserving digital publications and ensuring their accessibility through the **Kopal** project. Like the **KB**, the **DNB** allows users to access the content only through designated terminals in the library itself. Internet access is only possible with right holder permission.

Technical and Contractual Issues

Technically, ingesting eBook content should not be any more difficult than ingestion of e-journal content. Typically, the original PDF of a chapter (or an entire book) and its accompanying XML metadata are placed on a publisher's FTP site and subsequently accessed by the preservation initiative. Other methods of collection include data harvesting through regular crawling or dedicated **ONIX** feeds. A publisher should always send test content first to allow for investigation of any issues that might arise in existing workflow procedures.

The different initiatives discussed herein manage e-data differently. For example, **CLOCKSS** accepts and preserves source and presentation files. The content is preserved as it came from the publisher; the content is not manipulated or normalized. **Portico** works with each publisher to develop a customized preservation plan, and does subsequently normalize some publisher content, depending on the content type. I'm told that, compared to e-journal content, eBooks are usually more standard with good metadata. **E-Depot**, while not normalizing the content itself, chooses to modify the metadata into its proprietary standard. The **DNB** also changes the publisher metadata into its own preferred form. **Tobias Steinke**, who is responsible for the preservation mission of the **DNB**, tells me that the need for publishers to deposit unprotected files is crucial. **Digital Rights Management (DRM)** measures render the preservation process extremely difficult. I was pleased to hear that **Springer's** eBook files are very easy to work with because they are **DRM-free**.

When discussing the technical issues of eBook preservation efforts, it should be noted that some larger institutions or consortia opt for local hosting of eBook content, an option supported by most large publishers. This preservation method, however, may be beyond the reach of smaller libraries.

As noted above, technical issues are not the greatest challenge to preserving eBook content. In contrast to most journal publications where the publisher or a professional society retains copyright over published content, book contracts often contain a provision that requires

against the grain people profile

Heather Ruland Staines

Global eProduct Manager SpringerLink
Springer Science + Business Media
233 Spring Street, New York, NY 10013
Phone: (646) 201-8474

<heather.staines@springer.com>

www.springer.com; www.springerlink.com

BORN & LIVED: I was born into graduate student housing at **Iowa State University** in Ames, Iowa, but I grew up in Columbus, Ohio. I still consider Columbus to be my hometown. While I've lived on four continents (including Tokyo, London, and Wellington, NZ), my longest stint (in Connecticut) began in graduate school and stretches to the present (with a four year interlude in Richmond, Indiana).

EARLY LIFE: I was raised to be an Ohio State Buckeye, and I never appreciated what Columbus had to offer when I was still living there.

FAMILY: My husband **John** teaches in the English Department at **John Jay College (CUNY)**. I have two young sons, **Liam** and **Ethan**, who keep me busy as a homework assistant, cheerleader, and unofficial ultimate fighting referee.

PROFESSIONAL CAREER AND ACTIVITIES: I entered academic publishing after two years as a Postdoc at **Yale University**, where I directed the **United Nations History Project for International Security Studies**. **Greenwood Publishing Group**, an early creator of eBook and database products, proved an excellent training ground for electronic publishing. During my last three years there, I was Editorial Director for the **Praeger Security International** imprint. In March 2008, I joined **Springer** as their first Global eProduct Manager for **SpringerLink**. In this role, I work with sales and marketing around the world, in particular our Asian offices.

IN MY SPARE TIME I LIKE: I read thrillers, try my hand at writing some science fiction, sing a little karaoke, and spend way too much time on the Internet (research for work, of course!).

FAVORITE BOOKS: **John Irving's** *A Prayer for Owen Meeny*, **Dan Simmons' Hyperion**; and anything by **Lee Child** or **Harlan Coban**.

PET PEEVES/WHAT MAKES ME MAD: People who take too long in airport security and anyone who hides a candy wrapper under my couch.

PHILOSOPHY: Presume that other people are at least as nice as I am, until proven otherwise.

MOST MEANINGFUL CAREER ACHIEVEMENT: Launching my **PSI** imprint at **ALA Midwinter** in San Antonio in 2006. Weather problems kept our keynote speaker from attending the launch party, so I had to do the presentation instead. Thanks to an understanding crowd, a beautiful venue, and my trusty colleague **Adam**, the talk went off without any further hitch!

GOAL I HOPE TO ACHIEVE FIVE YEARS FROM NOW:

To travel to two additional continents (Africa and Latin America — I'll give Antarctica a pass for now!) and to get invited to share my ideas at Foo camp.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS:

As an acquisitions editor (and historian) at heart, I hope (and believe) that print books will never go away. I do, however, see an increasingly e-environment in publishing. I see publishers moving more into the provision of services and the boundaries between content producers and distributors blurring even more. 🐾

copyright to revert to the author in the event of a publisher bankruptcy or other development that renders the book unavailable to the marketplace. It is this provision that obligates a publisher to take extra care with eBook content that might be "triggered" some day. Going forward, publishers can add contractual language that explains participation in preservation initiatives, but publishers need to take

extra care with previously published content in accordance with the contract language.

Portico, with its library membership model, could choose to release triggered eBook content to their customers who have previously subscribed to that content; however, this would mean they treat triggered eBook content differently from triggered e-journal content, which

continued on page 32

is available only to **Portico** subscribers. The **CLOCKSS** model, on the other hand, makes triggered content publicly available to all under a Creative Commons license and will not recreate subscription access. While one could argue that a trigger event would be extremely unlikely, as most larger publisher's catalogs would be hosted by a successor, such an argument risks the creation of a scenario where eBook content is preserved but cannot legally be triggered, a situation which hardly fulfills the ultimate goal of digital preservation. Preservation of content within a library setting, such as that offered by the **KB** or the **DNB**, avoids the copyright issue but requires a scholar to travel to the Netherlands or to Germany to view needed titles.

A Commitment to Preservation

"There is always a question on archiving at each presentation I give, and rightly so" **Cynthia Cleto** notes. "Since the eBooks have a unique ownership model — customers that purchased a copyright year have perpetual access to that content — preservation becomes a concern. Customers want assurance that they will be able to access what is probably their largest library of eBooks in calm as well as turbulent times." **Craig Van Dyck**, Vice President of Global Content Management for **Wiley-Blackwell**, emphasizes that **Wiley's** strategy to preserve eBook content mirrors their approach to preserving journal content. "We are working with other parts of the industry (libraries, publishers, preservation archives, industry associations, technical experts) to come to terms with the issues, and to determine the best approaches."

The case for preserving eBooks in their digital form is a good one. In time, print archives physically deteriorate. Books can become damaged or be lost. A digital copy is more durable and takes up less space. Despite the challenges, **Springer** remains committed to finding a satisfactory solution in the near future. Knowing that **Springer** has robust preservation measures in place, our customers can rest assured when adopting our eBook content. As part of the **CLOCKSS** outreach committee, I speak regularly with publishers about their preservation strategies — or lack thereof. Recently, I have been describing a new **CLOCKSS** pilot project to ingest eBook content during 2009. As **Springer's** contact for **Portico**, I have proposed conference panels on digital preservation that feature the perspectives of the publisher, the library, and the preservation initiative. (Please look for us at **NASIG** in June). These opportunities naturally raise even more questions: can we preserve databases, electronic supplementary materials, whole Websites, files formatted for mobile devices?

We are still in the early stages of defining a comprehensive digital preservation strategy, one that requires the efforts of different entities with varying models. It is an exciting time to be working in publishing, alongside dedicated librarians and forward-thinking preservation initiatives. Future generations of researchers depend on the success of our collective efforts.

Tracking E-journal Preservation: Archiving Registry Service Anyone?

by **Peter Burnhill** (Director, EDINA National Academic Data Centre, Causewayside House, 160 Causewayside, University of Edinburgh, Edinburgh, EH 9 1PR, Scotland, UK; Phone: +44 131 650 3301/2; Fax: +44 131 650 3308) p.burnhill@ed.ac.uk edina.ac.uk

The idea for a registry of archived scholarly publications has featured in various digital preservation and archiving discussions. In 2003 **Maggie Jones** highlighted the need for clarity on what the various digital preservation agencies were doing.¹ In 2006, **Kenney** et al went further and recommended a registry that would indicate which agencies were preserving which journal content, one that could be used to identify gaps in publisher or content preservation coverage.²

JISC, the agency for the UK higher education that funds initiatives such as these, acted on this and commissioned a "scoping study for a registry of electronic journals that indicates where they are archived."³ Having interviewed a range of stakeholders in the UK, including representatives from national and university libraries, publishers, and archiving organizations, **Sparks** et al (2007) concluded, "Almost everyone agreed that there was ... an overall lack of information about where e-journals were archived, but more particularly, the difficulty of finding the information across a range of sources."⁴

There was, however, a lack of consensus on the scope of the registry. There were differences of view relating to timing, implementation and sustainability, and at least one archiving organization wished funding to go more directly to sustaining archiving per se.

On the matter of organization, the scoping study suggested the registry should be attached to something else that already existed in order to leverage existing organizations and infrastructure. A pilot project was recommended, followed by phased development of the registry using **SUNCAT**, the UK serials union catalogue,⁵ as the possible master list against which to compare the current and planned "holdings" of archiving services.

Piloting an E-journals Preservation Registry Service (PEPRS)

The next step for **JISC** was to commission a pilot registry service from **EDINA**,⁶ the UK national academic data centre based at the **University of Edinburgh** responsible for **SUNCAT**. **EDINA** opted to partner from the outset with the **ISSN International Centre**, with which **EDINA** has had a long-standing association.⁷

Funding for a two-year e-journals preservation registry service pilot (nicknamed **PEPRS**) began

in August 2008. The aim of **PEPRS** is to investigate, build and pilot an online facility that would enable librarians and policymakers, principally in the UK but also worldwide, to ascertain the archival provision for e-journals, especially of scholarly work published in e-journals, and to identify the gaps in such provision.

Preliminary Thoughts on Design and Re-considerations of Scope

A registry for e-journal preservation would correlate what is being done by each preservation agency for each known e-journal. This puts the focus on (1) metadata for e-journals and (2) metadata for each agency and archiving action, both of which are addressed briefly below.

The intention is not just to build and keep a register but also to deliver a set of registry services, so the **PEPRS** project must establish the functionality for a registry service, including the review and testing of user requirements, with implicit consideration of just who constitutes the primary use communities.

A registry needs to be accurate, up-to-date and comprehensive in coverage in order to be effective and command respect, as well as meet specific requirements and functionality. The choice of data model and architecture (3) are critical in determining that these matters of quality can be met.

This registry and the provision of its basic services must be designed to survive for the long run, like its subject matter, digital preservation. This implies (4) a business model that is sustainable over the long, one aspect of which should be low cost.

(1) Metadata on e-journals

A system of persistent and internationally accepted identifiers is clearly a good thing for a registry. The inclusion of the **ISSN International Centre (ISSN-IC)** as partner in the **PEPRS** project is as well. The **ISSN-IC** co-ordinates the **ISSN Network** which manages the international standard numbering system for serials, of which e-journals are a proper subset.

It could be argued that any e-journal worth preserving ought to have an ISSN.

A registry that made use of the metadata hosted in the ISSN serials database would have a critical mass of serial titles for project purposes, likely representing a good majority of the world's scholarly publications, including open access journals.⁸

The total number of e-journals is unknown but could be said to be growing. Fortunately, in recent years the **ISSN-IC** has made e-journals a priority for inclusion in the **ISSN Register** and has already issued over 60,000 identifiers for e-serials.

continued on page 34