Best Practices for Digital Repositories in Public Libraries

Talking Books Proof of Concept, 13 March, 2013

Report Author:

Maryann Kempthorne,
Accessibility and Learning Specialist,
BC Libraries Cooperative

This report was prepared for:

The Provincial and Territorial Public Library Council
Conseil provincial et territorial des bibliotheques

Table of Contents

Background

Public libraries open service doors to adults and children across a vast spectrum of information need. In the reading lives of many print disabled Canadians libraries are reliable and welcoming service centres for access to quality material and interaction about desirable titles to request for reading and research needs.

Cognizant of media shift and end-of-life on many formats libraries are set to build a bridge and plan effective collections for digital delivery that are:

- Well selected to meet standards of currency for collection and reliability for interest;
- Well described such that every book find its reader;
- Unrestricted in their downloads so that access is maximized, modernized and mapped;
- Measured and assessed for use and re-use in themselves;
- Integrated to library services overall change to format-on-demand;
- Effectively housed in technically sound, reliable digital libraries with responsible retention and disposition principles; and
- Extensible to a broad community of like-minded public (or other sister) library ventures through open data links and connections for subscribed content.

The NNELS repository project is one more in the series of migrations for library format specialties to better meet community need whilst bringing libraries beyond walls and traditional catalogues. As we work to develop quality digital access for visually impaired Canadian we do so mindful of how the fairness and openness of this access can translate to a universal design for library services.

Pan-Canadian Repository

Repositories in public libraries are scarce and accessible audio-on-demand systems rarer still. After 10 years of repository developments in K-12 and higher education and a lot of evolution on the archival side, public library repository services are still fairly new, and cross-institutional repositories are breaking fresh ground. As a first venture into a {consortial} public digital library of accessible content, we build around repository best practices that rely on:

- Simple metadata standards designed for interchange;
- Open source community software development for growth;
- Secure, lightweight and simple backup and replication systems; and
- Sustainable capacity targets and predictable storage costing.
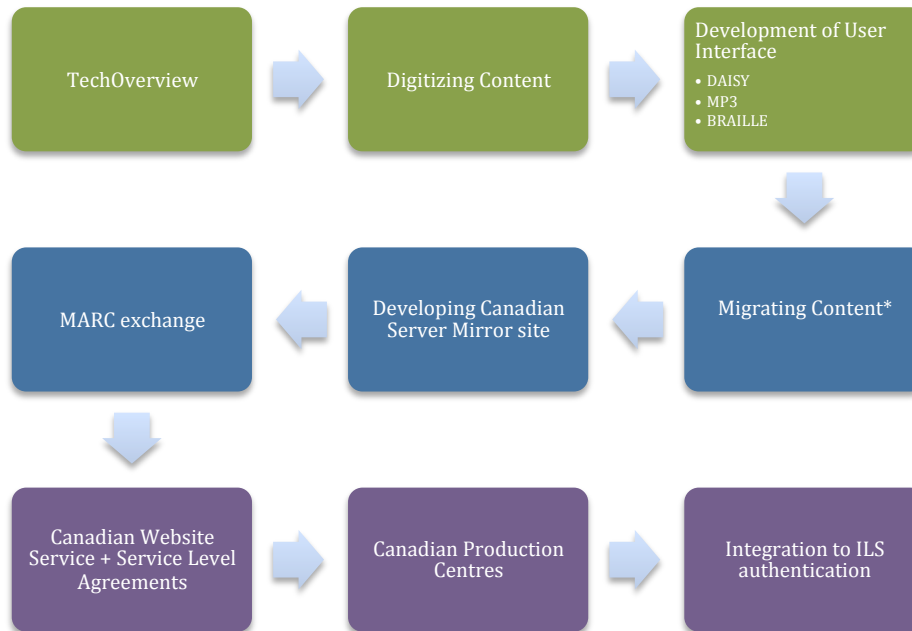
Component Approach to Repository Design

There are three main components to the proposed design of an audio-on-demand repository for public library service.
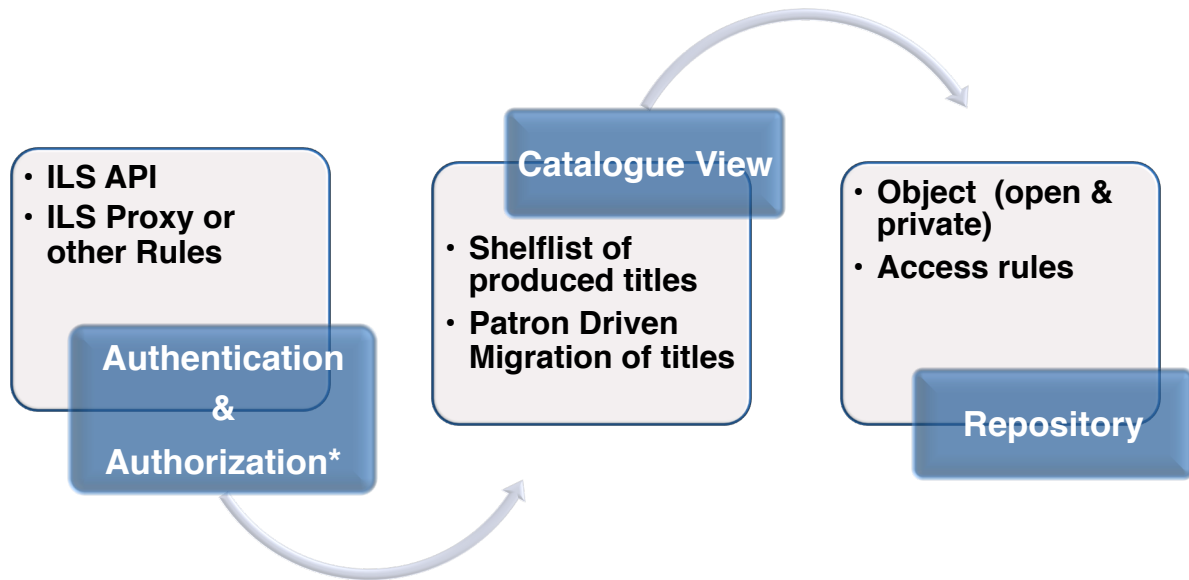
* The authorization of users in the CALS repository restricted to patrons with authorization to access copyright exempt materials. These users will authenticate themselves through their public library service centre by providing proof of eligibility in accordance with the local certification practices of their public library. Provincial and Territorial Library Branches may play some role in moderating membership certification and acting as chief verifiers for these borrower types.

Phases of Repository Development

To achieve a repository in a short period of time a phased approach to hardware, software, development and testing is planned.

```
TechOverview  →  Digitizing Content  →  Development of User Interface
                                          • DAISY
                                          • MP3
                                          • BRAILLE
                                                        ↓
MARC exchange  ←  Developing Canadian  ←  Migrating Content*
                  Server Mirror site
      ↓
Canadian Website       →  Canadian Production  →  Integration to ILS
Service + Service Level    Centres                authentication
Agreements
```

Outcomes

- 2,000 audio on demand fiction talking books immediately, potentially supplemented by a subscription service. Longer term 20,000-90,000 titles on the 'shelflist';
- Local production knit to international, strategic sharing;
- Production of 5% narrated, 95% synthetic voice (format sequence above) books;
- Canadian content for exchange, selected new title content and migration of content (the latter on-demand);
- ILS agnostic authentication for a library built repository;
- Scalable transferable repository framework up to 75,000 users in year one
- Independent use and institutional in the short term and 200,000 users over 3 to 5 years;
- Compatible to linked (federated) digital library access for other audio book formats and sources;
- Copyright advice will be sought around the appropriate format shifting of materials under current Copyright Act along with international controls of content in the catalogue.

The NNELS network: Content harvesting and Collection Sharing

Much of the progress of the CALS prototype was made possible by the generous assistance of the Commonwealth Braille and Talking Book Catalogue of the Association for the Blind of Western Australia. The longstanding spirit of co-operation of this organization guides much of our path ahead. Choosing to focus initial effort into an accessible web interface for users, building administrative tools, and opting for patron-driven migration of existing items are key drivers for the NNELs repository. Shelf list construction will prioritize DAISY format first, and transform digital braille and MP3 formats before other content types.

Linking to other sources, open data and integration of emerging subscribed content is a cornerstone principle of the project. In an environment of media shift the precise use case for this repository is accepted as finite. An optimized NNELs repository will offer content that is currently in demand with strong potential reuse and a direction toward seamless interaction of library services regardless of sector or content provider. The core of work in year one is to:

- Devise the content network, workflows and priorities;
- Negotiate excellent and up-to-date data models for similar content and publish their crosswalks;
- Digitize content and engage in the world digital archiving movement
- Rollout technical infrastructure that is cost-effective, well planned and will scale readily.

Content sources are expected to iterate at the fastest rates and a network to support the library as a digital delivery platform is the most durable outcome we can hope to achieve.

Working from the provincial and territorial library perspective, each jurisdiction should identify a local network of technical and collection specialists that can provide feedback on the tools and content. Software updates are planned on an annual basis and collection updates are targeted as monthly in the first year and weekly after that. Currently, we are seeking to establish production centres up to five strategic locations nationally.

Producing Digital Content

The costs of equipment for digitization of books have lowered substantially over even a few years ago. Combining affordability of more commonly used hardware with the strong development of open source software from the DAISY consortium (and other software developers) production of digital content by library agencies groups and by individual libraries themselves is newly reachable. Digitization is inherent in the operation of today's library and the NNELS repository brings the mechanics of makerspaces to library operations at a core service level.

Each production centre can get underway with a capital investment of less than $10,000. A train the trainer model based in recorded, re-useable self-paced training materials is recommended as repository operations build out. Shared community of practice and bug tracking will be published to a shared web space. Documentation of this type can be published via a wiki knowledge base ready to support de-centralized participants and capitalizing on the transferrable skills of professional, paraprofessional and clerical library staff.

The work of regional/post-secondary systems sets optimal standards for production. NNELS project productions will be designed for a choice of system platforms and use the most widely available hardware. The College and Institute Library Services (CILS) centre at Langara College, Vancouver, is on-board to assist the NNELS project as we develop a suite of tools and digitization processes. Informed by the post-secondary education network for production the project will benefit from CILS agreement to share procedures and experience with our production sites.

Why Open Source?

The Canadian Accessible Library System (CALS) project relies on open source tools and an agile development approach. Coming to the repository environment at a time of major library technology crossroads this is taken as a pragmatic approach – agility is something intrinsic to public library service. Basing our build on modular services rather than monolithic proprietary paths offers an opportunity to blend, and respond to, the diversity of participants in our context and on an internationally collaborative stage. With an eye to fostering a pattern of repository infrastructure that can reach toward multi-sector partnering and accessibility-centric metadata open-source community software development offers the most strategically sound and cost-effective course of action.

Risks and Opportunities

Authentication

Libraries have customarily relied on the authentication options of their Library Integrated Library Systems (ILS). With the evolution of subscription and open digital assets in libraries, these ILS authentication tools prove inadequate. Like other digital library resources the NNELS repository seeks to leverage the patron data management of the ILS without using it specifically. Preliminary investigations of the operations of current repository software solutions suggest that this is do-able in the medium-term.

Investing in API development for a wide variety of library systems ensures that the repository authorizes library cardholders' swift and immediate access to the content they are entitled to via the library. In the first case for the shared contents of a NNELs repository and subsequently for other resources including OpenLibrary, accessible content from related K12 institutions and/or higher education sources as well as subscribed content that meets similar needs. This aspect of the project presents a prescient opportunity for reuse of the authentication and authorization tools outcome of the investment in this area.

Service level commitments

The users of talking books in our proof of concept initial stakeholder engagement will use a catalogue to allow users to self-select materials. As with all library catalogues, expectations of data held and configurations for searches opinions vary wildly. For most repositories the reminder is wisely held that selected contents need not necessarily rely on traditional automated environments. The two keys to a usable data set is operations for catalogue users to access well selected materials under a process to set up personal profiles and that a strict logic is in place to populate the repository with valid and rich bibliographic data that can be effectively cross-walked among our collaborators. In the initial phases of the project managing the metadata for the objects will have intensive tasks. Updating this repository to the rapidly changing data standards for libraries in extensive flux for 2013 will also ensure that accessibility, multilingual and interoperability standards will be best met out.

User group expectations are that the digital library be fast, reliable and available 24/7 with only intermittent service windows and that content be as fresh and updating as the contents of current physical and other regular library services. Centralizing some staffing in the short term to establish this reliability will be a critical commitment in the first three years.

Recognizing the importance of addressing the current shortfall of accessible library services, we intend to launch a stable version of the repository as soon as

possible, with the caveat that users can expect better performance as the system is optimized.

Preservation and Security of Objects

The long-term preservation of the digitized objects for NNELS is a much more theoretical undertaking than repositories of, for example, university special collections or other institutional repositories.

While ideally, systems of original content plan for and are designed for the complexity of version and ongoing format migration, the repository content in the immediate term for NNELs it not intrinsically archival. This collection is intended to be a living and well-tended library. Initial design places the enabling of liberal access for library users before deep systems for retention. To seek a more complex archival and digital preservation approach will require a long-term sustained and top-down approach to administration and collection management which might not be most fully in keeping with the spirit of a mosaic approach to access nationally and potentially across the Commonwealth or Open web.

Timelines for Project Development

	Start Date	Completed
Technical Overview	November 15, 2012	January 30, 2013
MARC Analysis of current Canadian Collections	January 2013	March 15, 2013
Digitizing Content – Contract to CILS partners (seasonal)	February 2013	July 31, 2013
Production exchange with CBTBC	February 2013	Ongoing
User Interface Development	March 2013	July 15, 2013
Canadian Library Production Centres	March 2013	March 2014
Repository MARC Exchange with public libraries	July 2013	Ongoing
Library ILS integration	April 2013	August 2013
Canadian Website Services Launch	(beta testing July)	September 2013
Federated collections	January 2014 proposed	

References

"BC ELN Connect 10.1 May 2012. Institutional Repositories in British Columbia
" Accessed February 22, 2013. http://eln.bc.ca/newsletter/201204/article.html.

CNIB. (2012) Reading re-imagined: national digital hub to support Canadians
with print disability.

Cramer, Tom. (2011) Hydra: a technical and community framework for
customized, shared repository applications.

LeVan Ralph, OCLC Research: ContentDM vs. DSpace vs. Fedora.

"Patron Driven IR Development | Summit." Accessed February 22, 2013.
http://summit.sfu.ca/item/10810.