

# The Library of Alexander, Christopher

## Project Summary

This project is a significant audit of Minnesota Public Radio's metadata and schema workflow and datasets, resulting in documentation of audit findings and focused recommendations based on testing.

In the late 1970s the architect Christopher Alexander, with Berkeley's Center for Environmental Structure, published a trio of books on designing livable spaces. The first book, *The Timeless Way of Building*, is the philosophy—it proposes a new theory of design—the second book, *A Pattern Language*, is the tool by which one applies that philosophy, and the third book, *The Oregon Experiment* are the philosophy and the tools put into practice. This NDSR project take inspiration from Alexander's approach and philosophy; the trio of books are a comprehensive whole, yet a person could pick up any of the books, or read the bolded or italicized text of any pattern set, and know how to build an intelligent, comprehensive, life-giving space.

The micro-projects of MPR's NDSR—titled Schema, Public Record, Disambiguation, and Marketplace Archive—will result in levels of documentation, recommendations, and implementations that can have distinct applications and will together create a comprehensive approach to managing MPR's archive metadata. Throughout the micro-projects, the resident will apply the Alexandrian approach: assess the environment, design and apply patterns to improve the environment, and through these observations recommend a structure-preserving and structure-enhancing transformation that will be able to unfold continually within current MPR ecosystem. The result will hopefully be processes or procedures which will actually generate living structure—vibrant within the MPR metadata environment and in the other structures that metadata touches, improving upon MPR's digital initiatives, innovations, and relationships in the public media and preservation sphere.

It is tackling the question that Christopher Alexander poses in his 1996 presentation at the ACM Conference on Object-Oriented Programs, Systems, Languages and Applications (OOPSLA):

“How could one possibly get ahold of the massive amount of construction that is taking place on earth and somehow let it be generated in a good fashion and in a living fashion? And this is not merely because of the scale problem, but also one of the characteristics of any good environment is that any part of it is extremely highly adapted to its particularities.”

Just replace “construction” with “data”—how do we manage the massive amounts of data within and without our world, adapting to uniqueness of creation and use, creating a dataset and, moreover, an environment that is morally profound, that will

actually make human life better by way of bringing our approach into the system? In the context of this residency, the question is how can the archive harness and present our metadata to make the experience of discovery and use more profound? What environment will we set up that simultaneously deals with structure and has its roots in human emotion—particularly the human experience of wondering, searching, and listening. The work of this residency a giant leap towards answering that question.

### **Specific Project Goals and Objectives**

- Observation and education - Work with Archive team and meet with constituents
- Schema Reconciliation - Analysis and recommendation for data normalization
- Public Access - Efficient and controlled publishing for web, partners and aggregators
- Disambiguation - Cleaning and RDF tagging for people, places and events
- Marketplace – Assess the collection and initiate a plan for preservation and reuse.

### **Project Timeframe and Deliverables**

Months 1-4  
Schema

As an intro to MPR and MPR archive systems, the resident will spend the first months of the residency working with the metadata schemas used across the company, including the archive database (Eddy) and standards (PBCore), the CMS (Itasca and Drupal), blogs, RSS feeds, and internal production and broadcast systems (Dalet, Jukebox, Scheduler, etc.), examining all in the context of internal use and external collaborations and partnerships. The intention of this chapter of the residency is as follows: the resident is to 1) become familiar with the breadth of MPR production, broadcast, and archive systems, goals and needs; 2) Assess schema disparities, consistencies, and inconsistencies across and within systems; 3) Recommend fixes, based on testing, with an architectural program that includes a migration plan and program of execution. This schema audit will result in a recommendation for a metadata workflow that is clean, consistent, and comprehensive, promoting the gathering and sharing of data internally and externally and leaving the MPR Archive in a much stronger position to manage, share, and receive metadata.

Months 5-7  
Public Record

Building on the knowledge gained during the schema audit, the resident will lead a two-part project managing the Archive's public data and metadata. Part One: assess the Archive's public/published metadata and plan for the migration of that metadata into the internal archive database (Eddy). Part Two: assess workflow

required for publication of Eddy records. At this point, the majority of MPR's public data and metadata has no internal link back to the archive database; the online holdings and the internal archive are disconnected save from a manual search and rescue. Part One of this project will connect MPR's public story and program metadata with the internal archive records. This will include working in the archive's Drupal website (<http://archive.mprnews.org/>), as well as planning for the migration of data from the greater CMS environment (Itasca and Drupal) into Eddy and assessing migration needs for MPR's legacy websites with a looming kill date. Part Two will use that knowledge to establish workflow that supports the publication of Eddy records, taking into consideration current archive protocols and the requirements and ideas of users (e.g. the newsroom) and supporters (e.g. Eddy developers in MPR's Software Applications Group). This project will be the next big step in connecting the archive to the outside world; application of the resident's recommendations will bring a significant amount of unique data into the archive, reconcile published records with internal records, and lay the groundwork for consistent, clean data exchange between the internal archive and the public archive.

#### Months 8-9 Disambiguation

The final months of the residency will focus on the disambiguation of archive metadata. The archive's metadata is dirty from years of migrations, assumptions, and inconsistencies. The dataset of names (people, places, etc.) is particularly in a state of disrepair. The resident's task is to propose a workflow and environment for cleaning and controlling name-based metadata with methodology that can efficiently and effectively be applied to the over 300,000 existing records and records yet to come. As an outcome of the disambiguation project, the resident will provide 1) assessment of existing name-based metadata, 2) a manifest documenting the true or standardized names referenced within and without the archive—the master key, 3) a proposed workflow for executing such a massive metadata restoration, based on current archive resources and research of potential archive tools such as Open Calais, Watson speech-to-text and OpenRefine. Project will include an investigation into RDF tagging with output stored in a PBCore schema for both disambiguation and improved search (SEO).

#### Ongoing Marketplace Archive

Throughout the residency, the resident will assess the Marketplace archive and propose a detailed plan for fully integrating past and present Marketplace data and assets into the MPR Archive database and supporting systems. Returning to Christopher Alexander, this is The Oregon Experiment aspect of the residency, putting the resident's theories and tools into practice by applying them to a discrete scenario.

In short: 1) Assess all the elements and moving pieces (ENPS, David, Dalet, CDs, Drupal, etc.) of the Marketplace archive from the start of the program to present day, aligning pieces for Eddy ingest, Dalet storage, or server storage (depending on the need). 2) Propose workflow to get metadata and audio into Eddy; 3) Test workflow and update project plan with expectations of the time and resources it would take to execute the plan. In completing this project, the resident will usher a set of shows from start to finish, gaining first-hand experience in the very real world scenario of inheriting data and a range of archive formats and systems, and developing a working plan to herd it all into a comprehensive, usable, living archive.

### **Resources Required for Project**

The resident will be supported by the project mentor (Margaret Bresnahan), the Director of Broadcast Operations (Kyle Wesloh), two digitizers (Scott Adamson and Richard Rasch), and software application specialist (Kyle Swanson), all of whom are committed to providing the resident with training and resources in support of the resident's time at MPR and their work on the MPR Archive. In addition, the resident will have access to anyone at APM|MPR who could shed light on the resident's archive projects or support the resident throughout their time here. Connections to significant stakeholders—such as the Digital Products and Broadcast Operations teams—will be made early in the residency.

For the duration of the residency, the resident will have their own workspace, computer, and phone, and other tools that support the work of the resident. The resident's cubicle will be alongside the rest of archive team, in the heart of the Broadcast Technology & Operations department. The resident will have sufficient hardware and software to support their work, including access to relevant servers, systems, and databases supported by MPR.

### **Project Context**

The resident will be supported by the project mentor (Margaret Bresnahan), the Director of Broadcast Operations (Kyle Wesloh), two digitizers (Scott Adamson and Richard Rasch), and software application specialist (Kyle Swanson), all of whom are committed to providing the resident with training and resources in support of the resident's time at MPR and their work on the MPR Archive. In addition, the resident will have access to anyone at APM|MPR who could shed light on the resident's archive projects or support the resident throughout their time here. Connections to significant stakeholders—such as the Digital Products and Broadcast Operations teams—will be made early in the residency.

For the duration of the residency, the resident will have their own workspace, computer, and phone, and other tools that support the work of the resident. The resident's cubicle will be alongside the rest of archive team, in the heart of the Broadcast Technology & Operations department. The resident will have sufficient

hardware and software to support their work, including access to relevant servers, systems, and databases supported by MPR.

### **Required Knowledge and Skills for Residents**

- Self-motivated, resourceful, detail-oriented, accurate, thorough, and highly organized
- Demonstrated knowledge of database organizational systems
- Knowledge and interest in the emerging technologies that are shaping the fields of library, archives, collections, preservation and public access to archive content
- Digital Literacy: active user of web, mobile and networking technologies and services; able to inform decisions concerning new technologies from a user perspective
- Familiarity with PBCore or DublinCore
- Interest in system architecture
- Knowledge of news, current events and classical music
- The ability to work independently and collaboratively
- Ability to communicate clearly and effectively both orally and in writing
- Demonstrated time-management skills with proven ability to meet deadlines

### **Preferred Knowledge and Experience of Residents**

- Experience in creating and editing HTML documents and Web content
- Content management, software application bug tracking, and end user training
- Experience with OCR and speech recognition software
- Familiarity with production and CMS systems such as Dalet, ProTools, and Drupal
- Understanding of digital media formats including standards for the preserving audio, video, text and images
- Experience creating organizational systems and processes for users at various skill levels
- Demonstrated ability to analyze, implement, and complete large, complex projects and a strong desire to lead developments in the emerging field of digital archiving

### **Post-Residency Project Integration Plan**

When the residency is completed, the resident's recommendations will be directly applied to archive system renovations including but not limited to: normalizing metadata flow throughout our facility; connecting with third party endpoints like AAPB, DPLA, the Public Media Platform (PMP), and the Public Radio Satellite System (PRSS); and redesigning the way archive is positioned within MPR's production environment. An example of the latter is the archive's relationship to MPR's CMS tools. Currently there is no machine-based connection between MPR's public website data and the internal archive data; for example, a story published at [mprnews.org](http://mprnews.org) or [archive.mprnews.org](http://archive.mprnews.org) is devoid of any association with the master record that may or may not exist in the internal archive database. With the resident's series of recommendations, particularly those developed during the

Schema and Public Record portions of the residency, we will begin the CMS-Archive integration by consolidating the public archive portal website ([archive.mprnews.org](http://archive.mprnews.org)) with the internal database, applying the Disambiguation recommendations, and republishing the metadata to the public website. Feeds will be created off the backend of this new archive environment, enabling the archive to fluidly share and receive data from and with partners and platforms such as those described above. This is a holistic application of the resident's recommendations, establishing a format that makes publishing easier, cleaner, and more inter-connected.

With similar cohesion, the work of the resident will lay the groundwork for a number of archive projects, including:

- Preparing the archive for RDF, Semantic Web, and entity extraction.
- Preparing the archive for more effectively sharing metadata with MPR partners, including the American Archive, the Digital Public Library of America, and the Public Media Platform.
- Preparing the archive to be more visible and viable in internal initiatives, such as publically available archive selections curated by MPR staff.
- Inform the work we're doing to interconnect disparate internal schemas, and the projects that work supports.
- Inform database development (Eddy 3.0)
- Inform public projects – crowdsourcing games, data representation/presentation, etc.

In a myriad of ways the resident's work will help reposition the archive closer to MPR's audience and content producers, moving towards a freely accessible, intelligently interconnected, and agile data environment.

### **Additional Support**

In addition to the aforementioned support, MPR will provide the resident with a subsidized public transportation pass (unlimited rides at roughly \$26/month) or a discounted parking pass, depending on the resident's needs. The resident will also have staff privileges such as an employee discount at the MPR gift store and select events, and will be invited to attend any free events made available to the general MPR staff. Any other requests, needs, or concerns will be happily attended to.