Acquisitions Everywhere: Modeling an Acquisitions Data Standard to Connect a Distributed Environment

Eric M. Hanson
North Carolina State University Libraries, emhanso2@ncsu.edu

Paul W. Lightcap
Florida State University Libraries, plightcap@fsu.edu

Matthew R. Miguez
Florida State University Libraries, mmiguez@fsu.edu

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

Part of the [Cataloging and Metadata Commons](#), and the [Collection Development and Management Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Eric M. Hanson, Paul W. Lightcap, and Matthew R. Miguez, "Acquisitions Everywhere: Modeling an Acquisitions Data Standard to Connect a Distributed Environment" (2015). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316250>

Acquisitions Everywhere: Modeling an Acquisitions Data Standard to Connect a Distributed Environment

Eric M. Hanson, Electronic Resources Librarian, North Carolina State University Libraries

Paul W. Lightcap, Collection Strategist and Head of Acquisitions, Florida State University Libraries

Matthew R. Miguez, Metadata Librarian, Florida State University Libraries

Abstract

Acquisitions functions remain operationally crucial in providing access to paid information resources, but data formats and workflows utilized within library acquisitions remain primarily within the traditional integrated library system (ILS). As libraries have evolved to use distributed systems to manage information resources, so too must acquisitions functions adapt to an environment that may include the ILS, e-resource management systems (ERMS), institutional repositories (IR), and other digital asset management systems (DAMS).

This presentation is intended to articulate a vision for applying standards-based practice—as already employed for resource description—to acquisitions functions in a variety of metadata schema and systems. Utilization of standards will be demonstrated in the proposal of a core acquisitions element set that can exist in any system, with proofs of concept including demonstration of the element set within MODS, JSON, and how it may be reflected within the ILS and ERMS.

Building on these proofs of concept in recording interoperable acquisitions data will be an explanation of possible applications, including an exploration of more robust support for semantic web technologies. In particular, this presentation will explore how this element set could utilize published linked datasets, such as the North Carolina State University Organization Name Linked Data and Global Open Knowledgebase (GOKb) linked data service, to provide more accurate and efficient identity management.

Introduction

While substantial work has been done over the past decade to develop and implement standards for resource description, electronic resource management, and discovery within libraries, the realm of acquisitions data and functions—that core set of financial information and accounting practices crucial to growing and maintaining the bulk of what any library counts as its collections—has yet to be comprehensively addressed. Acquisitions functions remain complex for practitioners, demanding adherence to local procurement rules and standard accounting practice, with a continually increasing demand for, on one hand, instantaneity and flexibility, and on the other, transparency and consistency. Perhaps exactly this combination of complexity of practice and the hope for a single, comprehensive solution, has until now composed a seemingly

insurmountable obstacle to standards application within the acquisitive world.

Instead of a total solution that fully accounts for competing demands and exception-ridden workflows, this proposal defines and extends the use of a common core set of elements already part of daily acquisitions work through application across the many systems currently used to manage all forms of paid content. The inspirational models for this work can be found in overlapping communities of practice as have developed the Program for Cooperative Cataloging's BIBCO Standard Record metadata application profile and the Dublin Core Metadata Element Set. This core set of acquisitions elements intersects with work already done as part of the NISO Cost of Resource Exchange (CORE) recommended practice, (<http://www.niso.org/publications/rp/RP-2010-10.pdf>); this work both

supports and extends the CORE efforts, though, in that it proposes a somewhat shifted element set, makes explicit—through proofs of concept—applications for this element set across different systems beyond the ILS/ERMS/vendor triad, and demonstrates initial steps forward in utilizing semantic web technologies. The use of this proposed element set is intended as the first step in an iterative approach, acknowledging that any initial application must cross a low threshold in order to move toward normalizing certain crucial acquisitions data, expanding the use of this data in various library and resource management systems, and laying the groundwork for an evolving conception of distributed acquisitions data and workflows.

The Legacy Challenge, Compounded

As the next-gen library services platform (LSP) market has matured and concerns about age and functionality of the legacy ILS have increased, many libraries have already begun to move ahead with the next round of systems migrations, (Breeding, 2015a; Breeding, 2015b), or at least to prepare for such a move. Considering an ILS migration resurfaces worry about challenges faced in past migrations, topmost of which is concern about data loss. This is particularly noted as impacting acquisitions data such as resource cost, fund, and vendor (Farrell & Truitt, 2003; Butler & Larmore, 2006). While there is some indication that progress has been made in maintaining such legacy data during recent migrations to library services platforms (Fu & Carmen, 2015), it should be noted that the underlying connection between description and acquisitions metadata remains somewhat tenuous.

Many libraries have attempted to address these weak connections between descriptive and acquisitions data by utilizing not only an ILS, but also an ERMS. Through combined reporting from both systems—and, indeed, duplicating data stored within these systems—libraries have developed a point-of-need solution for recorded data that is otherwise missing or difficult to extract. Beyond this, though, libraries have needed to capture and report on data that cannot

be adequately mapped within and reported from either the ILS or ERMS, and so have further duplicated data with the introduction of additional helpmates, such as separate spreadsheets or external databases. In sum, even with the development of new technologies to manage the largest volume and highest portion of expenditures—traditional print and electronic resources—but no accompanying efforts to standardize and deeply link an acquisition element set, libraries have built up an environment of additional systems and substantially duplicated data, without any capability for mutual understanding and interoperability.

Compounding technical worries of data loss and data structure are substantial environmental shifts: the emergence of data-driven culture, a growing emphasis on evidence-based practice, and an ever-increasing scrutiny on expenditures, especially within public institutions. Put succinctly, the demands to immediately and consistently report on expenditures for licensed and purchased resources, correlated to discipline and use, continue to grow. Such an expectation goes far beyond accountability to the citizenry for public expenditures; as experience over the past decade within college and university libraries amply shows, libraries are more regularly expected to report on total expenditures in support of departments or majors, including historical reporting to show trends or past years' oversights. Years of strained funding and implementation of different budget models within parent organizations such as activity-based, performance-based, or responsibility center budgeting, have amplified the demand for robust expenditure analysis mapped to use and areas served. Libraries have been called on to advocate for their funding based on credible data—or, in the case of budget reductions, to produce meaningful analyses that support strategic cuts.

All of these competing concerns and demands alone still would not necessarily prompt rethinking our practices. If resources remained in traditional formats and were acquired through traditional means, perhaps the most effective solution would be to redouble focus on library

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns="http://www.loc.gov/mods/v3"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3"
  xmlns:flvc="info:flvc/manifest/v1"
  xmlns:acq="http://someURI.host"
  xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-4.xsd">
  <extension>
    <flvc:flvc>
      <flvc:owningInstitution>FSU</flvc:owningInstitution>
      <flvc:submittingInstitution>FSU</flvc:submittingInstitution>
    </flvc:flvc>
    <acq:acq>
      <acq:invoice>
        <acq:customerID>http://viaf.org/viaf/125490841</acq:customerID>
        <acq:vendorID>https://www.lib.ncsu.edu/id/onld/00000728</acq:vendorID>
        <acq:invoiceID>001A</acq:invoiceID>
        <acq:invoiceDate>2015-08-15</acq:invoiceDate>
        <acq:orderID>8675309-1</acq:orderID>
        <acq:orderFiscalCycle>2016</acq:orderFiscalCycle>
        <acq:orderType>One-time</acq:orderType>
        <acq:totalPaid currency="USD">127.50</acq:totalPaid>
        <acq:listPrice currency="USD">150</acq:listPrice>
        <acq:invoiceLine id="001">
          <acq:fundID>FAKABTL</acq:fundID>
          <acq:accountNumber>123456</acq:accountNumber>
          <acq:materialType>application/pdf</acq:materialType>
        </acq:invoiceLine>
      </acq:invoice>
    </acq:acq>
  </extension>
  <titleInfo lang="ita">
    <nonSort>Il</nonSort>
    <title>secolo</title>
    <subTitle>Gazzetta di Milano</subTitle>
  </titleInfo>
</mods>
```

Figure 1. Example of element set as an acquisitions extension in a MODS record.

Summary	Full Description
Title:	Il secolo: Gazzetta di Milano.
Type of Resource:	text
Genre:	newspaper
Issuance:	serial
Frequency:	Daily
Date Issued:	1885-04-10
Publisher:	Sonzogno
Place of Publication:	Milano
Extent:	4 p. : ill. ; 59 cm.
Language(s):	Italian
Summary:	This issue contains Le colonie; Attraverso i secoli: Arti vecchie sempre nuove: Corrispondenza Telegrafica del secolo: Sul segreto e sulla libert� del servizio telegrafico; Il comunicato degli eredi Garibaldi; Il Giudizio del "Fascio" ; La politica di Mancini; L'abbandono di Caprera; La battaglia nell'Afghanistan; I membri del Ministero francese; Quello che i governi fanno del denaro pubblico; Il Crematolo di Firenze; La questione agraria nel Mantovano.
Identifier:	FSU_ilsecolonewspaper_04101885 (IID)
Subject(s):	FSU_ilsecolonewspaper_04271885 Ohnet, Georges, 1848-1918 La signorina di compagnia 54 La grande maniera 6
Use and Reproduction:	Use of this item is provided for non-commercial, personal, educational, and research use only. Florida State University Libraries is providing access to these materials for educational and research purposes and makes no warranty with regard to their use for other purposes. The written permission of the copyright owners and/or other rights holders such as holders of publicity and/or privacy rights is required for distribution, reproduction, or other use of protected items beyond that allowed by fair use or other statutory exemptions see Title 17, U.S.C.. For information about the copyright and reproduction rights for this item, please contact Special Collections & Archives, Florida State University Libraries, Tallahassee, Florida: https://www.lib.fsu.edu/departments/special-collections-archives .
Owner Institution:	FSU

Figure 2. Same record displayed within the FSU Digital Library, demonstrating public masking of acquisitions data.

support management and access to this varied content. Libraries have moved well beyond purchase of print and electronic books and journals. Collections now include purchased and licensed digitized manuscripts, satellite images, videos, datasets, and many other digital artifacts (whether born digital or digitized), that libraries must describe and host. Neither the integrated library system nor the ERMS is the appropriate repository for such digital objects or their metadata; instead, libraries have developed institutional repositories and other digital asset management systems to serve this purpose.

A Lightweight Standard

Standards, as acknowledged within the semantic web and metadata communities, are crucial in facilitating interoperability, enhancing reliable retrieval, controlling duplication, and making migration simpler. We propose a lightweight standard, not a deviance from predominant current practice, in the form of a set of core acquisitions elements.

services platform functionality, emphasizing integration and centralization. The scholarly record has changed, though, as has the way that libraries support the research and educational efforts of their constituencies. Libraries have expanded the types of information resources that may compose collections, and added systems to

This element set was conceived as lightweight in application, with no substantial technological impediments to implementation. It was modeled from significant data elements extant within the systems already in use to manage content: the ILS, ERMS, IR, and digital library (DAMS), with examples ranging across content types. The

intention was to demonstrate how such an element set, if standardized, could not only allow for robust meta-analysis but also lay the foundation for moving forward acquisitions practice to use linked data, support all content types and systems, and start to transition from ILS-centric practice.

Acquisitions metadata is managed in both formal and ad-hoc systems. This leads to duplication of information as data is replicated within different systems to answer different questions. This differs from resource description practice where efforts are directed at creating interoperable standards to minimize the duplication of effort. Beyond an increase in effort, maintaining multiple data siloes can lead to “data drift” as information in one silo is updated but not refreshed in others. This problem is being addressed in the resource description community through the use of semantic web technologies. A strategy specific to acquisitions would be to bundle acquisitions metadata into the resource description. This way, wherever the record moves, its acquisition information moves with it.

How can acquisitions data be bundled to resource description? The information serialization technologies XML and JSON can be of service. Both technologies are system independent, and widely understood by the information and technology communities. Their widespread use has been supplemented with a variety of solution packages for every common programming language and standardized schema to solve wide-ranging problems from warehouse and inventory management to bibliographic description. These advantages have spurred the resource description community to develop a kaleidoscope of description standards built in XML.

The system agnosticism of XML and JSON means they can be ingested and exported from any system that can access and understand the rules of the format.

This is a fixed set of data mostly useful for reporting, but by standardizing and fixing the information to the resource, the need for ad-hoc data tracking practices could be eliminated. The

data needed for financial reporting and analysis will move with the resource description when systems change. A publicly maintained and documented serialization standard ensures long-term access to the encoded data, and efficient analysis of data across distributed systems.

A few use cases can demonstrate the immediate utility of adopting a standardized acquisitions metadata element set. Current practices preclude acquisitions analysis at the level of consortia, even if expenditure data were able to be shared, as data recording practices are too different, and the systems for storing the data have no common language. Performing high-level analysis requires too much data clean-up and reconciliation to be feasible. If a standard element set were adopted, this process would be greatly simplified, even among members storing data in different systems.

Another use case would be performing meta-analysis against collections held in distributed systems. The ERMS, ILS, and DAM have different strengths for content management and access. As a result, paid content is hosted in an increasing number of systems, which leads to additional systems being used to track acquisitions data. For instance, supporting a STEM program might require purchasing datasets, digital images, and interactive licensed content along with more traditional library acquisitions. Common external systems such as spreadsheets are used to track all of this spending, and correlate it to resources. With acquisitions data bundled into the resources, collecting this data for aggregate analysis would require only a few queries against the different hosting platforms.

As the conversation around acquisitions metadata develops, a core element set may be extended to answer more complex questions such as cost-per-use or other dynamic data. The community can also look to resource description for other insights. Open Archives Initiative (OAI-PMH) is a simple web-service for standardizing the remote harvesting of descriptive metadata. In a distributed information environment, a similar service can be developed to standardize acquisitions information collection across

Resource Identifiers (URIs) that can exist independently from the labels used by each system. There are many potential sources of URIs for organizations, including the Virtual International Authority File (VIAF), Library of Congress Name Authority File (LCNAF), International Standard Name Identifier (ISNI), DBpedia, and the North Carolina State University (NCSU) Organization Name Linked Data (ONLD).

The NCSU ONLD set is based on the NCSU Organization Name Authority, a tool used to manage the variant forms of names for serial and e-resource publishers, providers, and vendors in E-Matrix, a locally developed electronic resource management system. This dataset was enhanced with links to related descriptions in VIAF, LCNAF, ISNI, DBpedia, and Freebase. The NCSU ONLD served as the seed data for organizations in the Global Open Knowledgebase (GOKb), which is an open data repository that contains key publication information about electronic resources as represented within the supply chain from publishers to suppliers to libraries. It is also used by the AgriProfiles project (formerly AgriVIVO) to help manage organization names.

The NCSU ONLD can be used to transition a set of organization labels into a set of URIs for those organizations. In the NCSU ONLD, each organization is assigned a URI and a preferred label, as well as storing a list of alternate labels. With this information, the NCSU ONLD can be used to crosswalk different labels coming in from different data sources to a single URI that identifies the organization. By checking a list of organization names against the NCSU ONLD, a list of ONLD URIs could be generated for any organization labels that matched against either the preferred or alternate labels. Additionally, any VIAF, LCNAF, ISNI or DBpedia URIs that are provided through the ONLD could be used as identifiers for any of the matching organizations.

By describing customers and vendors using URIs rather than labels, it is much easier to build organization-oriented reports based on elements from multiple data sources. Integrated library system (ILS) payment data can be compared to the list price on vendor title lists for calculating the cost savings associated with different vendors. The amount for content covered by a particular vendor license could be generated using licensing data from an ERMS and payment data from the ILS. Usage data from an ERMS could be analyzed by subject area based on call number data from the ILS and proprietary subject classifications from a vendor title list. With a thoughtful and consistent approach to identity management that utilizes URIs across multiple data sources, more complex scenarios could be devised that would have been unthinkable in previously siloed systems.

The Next Step Forward

The proposed core element set, applied within different schema and utilizing URIs where possible, is but a first, tentative step toward standardizing acquisitions data. The authors understand that development and application of standards within the large community of practice of resource acquisitions necessarily requires a broad base of participation to further develop the proposal. It is the hope that this initial suggestion may prompt participation from acquisitions and metadata practitioners to further develop use cases, continue to refine the proposed core element set, and build additional proofs of concept across myriad systems to fully explore the utility of standards development for acquisitions. A solicitation for participation to achieve these goals will be shared widely in the coming weeks across acquisitions and metadata communities, but directly contacting the authors to indicate interest in moving this project forward is also encouraged.

References

- Breeding, M. (2015a). Library systems report 2015. *American Libraries Magazine*. Retrieved from <http://americanlibrariesmagazine.org/2015/05/01/library-systems-report/>
- Breeding, M. (2015b). Library services platforms: A maturing genre of products. *Library Technology Reports*, 51(4). <http://dx.doi.org/10.5860/ltr.51n4>
- Butler, T. L., & Larmore, D. P. (2006). Biz of acq—The conversion problem: One state's migration of acquisitions data. *Against the Grain*, 18(3). Retrieved from <http://docs.lib.purdue.edu/atg/vol18/iss3/29>
- Farrell, K. T., & Truitt, M. (2003). The case for acquisitions standards in the integrated library system. *Library Collections, Acquisitions, & Technical Services*, 27(pp. 483–492). <http://dx.doi.org/10.1016/j.lcats.2003.09.012>
- Fu, P., & Carmen, J. (2015). Migration to Alma/Primo: A case study of Central Washington University. *Chinese Librarianship*, 40, 1–14. Retrieved from <http://www.iclc.us/cliej/cl40FC.pdf>