

Knowledge Base and Metadata - Scope and Domain

Introduction

The management of metadata, within a Knowledge Base (KB, or more than one), is an important aspect of the early work of FOLIO. This living document presents our current thinking and proposed plan for ongoing development.

Context

The environment within which metadata and Knowledge Bases exist is rapidly evolving, and no single approach to metadata and knowledge management (or choice of implementation systems) spans institutions, groups, nations or cultures.

A broadening spectrum of resource types, purposes and contexts is putting pressure on the ability to support and move between a wider variety of metadata representations and formats. In part, this means there is no longer a single obvious choice for the format or representation of metadata.

More sharing across institutional boundaries, in response to the burden of an increasing rate of change of metadata and quantity of resources to describe, is affecting attitudes towards authority and ownership.

Document Scope

This document covers the initial scope of the metadata work within FOLIO and is focused on the operational needs of an academic library. It currently only covers some basic aspects of bibliographic and holdings/catalog metadata.

Broad Objectives

Support a Library's Operational Needs

Any resource metadata management system needs to support the operational needs of a variety of Library contexts, some of which are:

- Acquisitions
- Cataloging
- Inventory
- Circulation
- Access
- Discovery

These different contexts might have overlapping yet significantly different needs and workflows, each will use different aspects of a variety of metadata models for resources.

Support Unforeseen Uses of Metadata

Bibliographic and management metadata is increasingly used in a variety of learning, teaching and scholarly communication (amongst others) contexts. We need to be mindful to create a model that supports unforeseen use by new applications.

Some recent examples of this are:

- Article subject area analysis to predict possible funding opportunities and inform research decisions
- Article Processing Charge (APC) handling where libraries are struggling to track the cost of open access venue publication in an auditable way.
- Analytics and reporting on institutional scholarly communication processes, such as the Research Excellence Framework.
- Institutional responsibility to store, publish, and preserve data and other artifacts of the research process.

Demonstrate FOLIO's Architecture

FOLIO's purpose for providing a platform for building library systems expands the technical considerations beyond what is typically present in the development of an Integrated Library System.

As the metadata modules are fundamental parts of a FOLIO system it is important that any reference implementations demonstrate effective use of the platform.

Examples of some of the technical aspects relevant to providing metadata:

- Modular separation of concerns and behaviour (e.g. storage and business logic)
- Integration of a combination of existing external open source and commercial systems
- Interchange (publishing and consumption) of linked data representations of resource metadata

Initial Scope

Where do we start?

The scope and objectives above are broad and cover a lot of areas and topics, and in order to demonstrate incremental progress and to elicit feedback, we need to choose a narrower focus for where to begin.

Our initial work is intended to support the needs of the ongoing FOLIO development and provide basic capabilities needed to operate a library catalog.

We will expand on other areas of bibliographic and management metadata as the work progresses.

Initial Goals

Within the context described above, some initial goals for metadata support with FOLIO could be:

- Reduce repeated cataloguing effort within organisations
- Allow organisations to transition to FOLIO incrementally
- Establish the groundwork for support of data formats beyond MARC and bibliographic standards beyond AACR2 or RDA
- Unify electronic and physical resource management where possible

Initial Outcomes

To begin to achieve those goals, the outcomes we seek are to:

- Support reference and copy cataloging
- Support external bibliographic and management metadata knowledge bases
- Support a wide range of resource types
- Support a wide range of import and export formats and representations
- Support a wide variety of technology choices (both open source and commercial)
- Easily map existing catalogs to external bibliographic or subscription metadata

Initial Deliverables

In order to start having the desired impacts above, an initial set of deliverables could be a system which provides:

- Basic representation of physical monographs for circulation
- Ingesting of an existing inventory of physical monographs
- Basic representation of electronic journals or books (to be decided) and entitlements
- Bibliographic metadata read from an external Knowledge Base
- Management metadata read from an external Knowledge Base

Planning

Below is a short description of the scope of each of the above deliverables.

Basic representation of physical monographs (for circulation)

This is the most basic requirement of any inventory capabilities and will underpin basic circulation of monograph copies.

This deliverable will introduce the concepts of an item and (internal) instance into the domain model.

Ingestion of an existing inventory of physical monographs

This deliverable allows for a collection of physical monographs (possibly represented in MARC21 or MODS) to be imported into the FOLIO inventory.

This will test and demonstrate the initial basic cataloguing capabilities of FOLIO and provide a basis for different models for transitioning to FOLIO based resource management.

Some degree of instance matching and consolidation between items will be included in this work. As will the storage of the original source records which were ingested.

Basic representation of electronic books or journals and entitlements

Both of the current example external Knowledge Base systems primarily contain electronic resources, therefore in order to demonstrate integration with external systems, FOLIO first needs to support some aspects of electronic resources.

This deliverable will likely introduce the concepts of packages, subscriptions and entitlements, and will be the first opportunity to try and unify the resource management models for physical and electronic resources.

We will decide closer to implementation which of electronic books or journals to support first.

Bibliographic metadata read from an external Knowledge Base

This deliverable will provide the ability to read bibliographic metadata (predominantly instances) from an external Knowledge Base system.

How we might start to map local (internal) bibliographic and holdings metadata to global (external) definitions is a major part of this work.

In order to test the design of the interfaces involved in this process, it is prudent to integrate with a variety of existing systems. Choosing one open source system (e.g. GOKb) and one commercial system (e.g. EBSCO EPKB) could provide a good starting point for this.

Management metadata read from an external Knowledge Base

This work extends the integration with an external Knowledge Base to reading management metadata (predominantly items and entitlements, but may also include package, platforms and subscriptions).

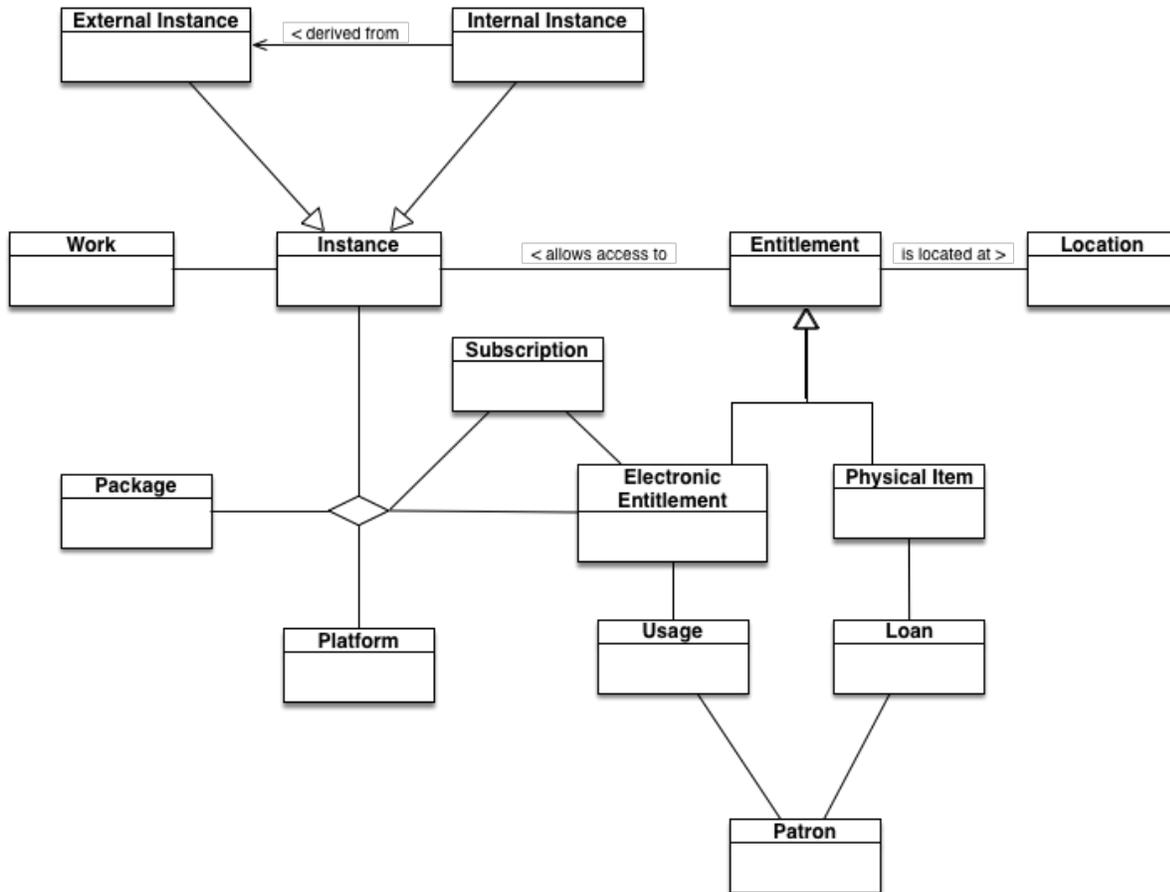
Conceptual Domain Model

An important aspect of this work is in trying to determine a general agreeable set of nomenclature for this domain.

Below is a partial and speculative conceptual domain model, intended to show many of the core aspects of resource metadata (mostly bibliographic and management) and to elicit feedback from the community.

Only some of the concepts within the broader domain are relevant to circulation and access, however they are intended to represent the start of a model that is used in a variety of contexts.

An expansion of the terminology used in this diagram is available in Appendix 1.



Risk Register

All efforts carry some risks, this is a partial list of identified (known) risks surrounding this work.

Summary	Likelihood	Impact
Overly specific bibliographic metadata formats and representations (e.g. too predominantly coupled to MARC21)		

Appendices

1. Terminology

Term	Definition
Usage	Statistics relating to how frequently an electronic resource has been accessed
Authority	The organisation (e.g. Library of Congress) responsible for an established particular form of a concept (e.g. subject or person)
Bibliographic Metadata	Metadata describing a bibliographic resource such as title, author(s), publisher, date and place of publication, edition, standard numbers (identifiers?), subjects, format ^[1]
Electronic Entitlement	An entitlement to an electronic resource
Entitlement	The access granted to an organization, allowing it to a resource
External Instance	An instance whose authoritative representation is owned by an external bibliographic metadata knowledge base
Holdings	All resources contained within or accessible via a given library
Knowledge Base	An external source of metadata related to resources
Identifier (standard number?)	A alphanumeric string used as a unique identifier for a resource. Some identifier types have established validation rules (e.g. ISBN and ISSN) ^[1]
Ingest / Import	The act of importing, or ingesting, and processing information from an external vendor ^[1]
Instance	A material embodiment of a resource, e.g. a particular published form ^[2]
Internal Instance	An instance whose authoritative representation is locally scoped to the organisation whose holdings/catalog contains one or more copies. May be derived from an External Instance.
Inventory	The group of physical items an organization owns (and is entitled to use)
Physical Item	An physical item is an physical copy of an Instance ^[2] Ownership of which effectively entitles an organisation to use it
Loan	The process by which the system: (1) validates whether or not a library user can borrow a library item based on defined attributes and (2) if a loan is permitted, links the item with the patron and applies certain conditions based on policies ^[1]
Management Metadata (or Administrative Metadata)	Data about an information resource primarily intended to facilitate its management ^[1]
Metadata	data that provides information about other data ^[3]
Package	A grouping of instances on a platform offered by a supplier under particular terms ^[1]
Platform	An interface that administers or delivers electronic resources content, or provides a route to the content, to the user ^[1]
Resource	an item that may be collected and/or made available by an organization ^[1]
Resource Management	the practices and techniques used by librarians and library staff to track the selection, acquisition, licensing, access, maintenance, usage, evaluation, retention, and de-selection of a library's resources ^[derived from 4]
Source Record	The original representation of a record ingested from an external source (relates to records which were derived from it)

Subscription	A subscription is an agreement or potential agreement between a 'subscriber' and a 'provider' to gain access to a set of resources, for a period of time, under specific conditions (set out in a license and/or elsewhere), and usually at a specific cost [5]
--------------	---

2. References

1. FOLIO Product Council, Glossary of Terms, available at <https://wiki.folio.org/display/PC/Glossary+of+Terms> (accessed 2016-12-15)
2. Library of Congress, Overview of the BIBFRAME 2.0 model, available at <https://www.loc.gov/bibframe/docs/bibframe2-model.html> (accessed 2016-11-09)
3. Merriam Webster Dictionary, available at <https://www.merriam-webster.com/dictionary/metadata> (accessed 2016-12-15, via Wikipedia)
4. Electronic Resource Management, available at https://en.wikipedia.org/wiki/Electronic_resource_management (access 2016-12-15)
5. KB+ Concepts and Terminology, available at <https://knowledgebaseplus.wordpress.com/kb-support/kb-discussion-documents/kb-concepts-and-terminology/> (accessed 2016-12-15)

3. Assumptions

4. Kabalog

Kabalog (n): Unified data model and infrastructure for conventional bibliographic metadata and E-resource Knowledge Bases.

I generally extend this definition mean an expression of the tension between a desire for the fusion of knowledge base and catalog contexts whilst keeping some of the concepts in those worlds distinct and loosely coupled.