

FOLIO Special Collections & Archives Working Group (FOLIO SC&A WG) White Paper (August 2018)

The FOLIO SC&A WG was formed in late 2017-early 2018 in response to questions about FOLIO from the SC&A community and at the suggestion of the FOLIO Product Council (PC). The formation of the WG also reflected the general recognition that traditional Integrated Library Systems and even comprehensive Library Management Systems are not good solutions for special collections and archives; and that the needs of the SC&A community need to be represented in the FOLIO project. The WG defined itself as “an exploratory pre-SIG” in recognition of its preliminary nature and to distinguish its work from the other FOLIO SIGs.

Charter

The WG’s charter was drafted and adopted in late January 2018. It proposed four specific areas of activity:

1. To identify unique needs of SC&A departments that are within the scope of the FOLIO Project and/or relevant to FOLIO SIGs (e.g. the Metadata Management SIG).
2. To determine the degree to which these needs are being met by existing open-source SC&A software applications.
3. To explore whether existing SC&A software can be adapted, modified, and/or incorporated into the FOLIO code base.
4. To report back to the FOLIO PC on its findings, with recommendations. Report date: May 1, 2018.

Work to Date

The WG has met weekly since late January 2018, with a break in July 2018. The group made modest progress on the first two points in its charter—specifically, to identify unique needs of SC&A departments that are within the scope of FOLIO and/or relevant to FOLIO SIGs; and to explore the degree to which these needs are being met by existing SC&A software. It has yet to address the third—that is, to explore whether existing SC&A software systems can be connected to FOLIO or eventually integrated into the FOLIO code base. The WG submitted a [progress report](#) on its work, with recommendations, to the FOLIO PC on May 1, 2018.

The best source of information on the WG’s work to date is its [Wiki page](#).

SC&A Needs and the FOLIO SIGs

SC&A materials and workflows differ from traditional library materials and workflows. For example: archival materials tend to be one-of-a-kind and to require customized descriptions. These descriptions may represent content in aggregate form, rather than individual items. Description may also involve complex hierarchical relationships (e.g. analog to digital, digital to other digital derivatives and/or formats, and other parent-child relationships). Furthermore, SC&A departments deal with a wide variety of materials, ranging from manuscripts and photographs, to legacy media (e.g. film stock, reel-to-reel tapes, cassette tapes, and obsolete removable storage media), to textiles, art works, and other 3D objects. Physical items may have associated records (e.g. for conservation treatments), as well as digital surrogates. Increasingly, SC&A departments are assuming responsibility for the stewardship of electronic records, digital video, and other born-digital content. Finally, special collections and archives have access policies that are fundamentally different from library access policies. Most library materials circulate; archival materials don’t.

For these and other reasons, traditional library software systems and the MARC format are not good solutions for describing, managing, and providing access to special collections and archives. The unique nature of SC&A requirements implies specialized software applications and/or complex integration with

other systems. To some degree, these requirements are reflected in a number of extant open-source and commercial software applications that are specifically designed for use with archival materials. Examples include [ArchivesSpace](#) (and its precursors [Archivists' Toolkit](#) and [Archon](#)), [AtoM](#), [ArcLight](#), [Aeon](#), and [CALM](#). The number and variety of software applications in this area suggests that there may still be room for a comprehensive, end-to-end software solution that addresses more or different aspects of the SC&A workflow. There is certainly a need for better interoperability among library management systems, discovery systems, and archival management systems. The FOLIO Project seems like a good venue in which to explore these questions.

The FOLIO SC&A WG has identified a number of SC&A-specific issues and desired features. They map fairly closely to existing FOLIO SIGs, especially the Resource Management and Metadata Management SIGs. They also connect, to a lesser degree, with the Resource Access, Accessibility, and System Operations and Management SIGs. Here are some items from the SC&A WG's wish-list, organized by the relevant SIGs.

1. Resource Management SIG

- Better ways to manage related collections in other repositories.
- Better ways to handle cataloging issues that are specific to SC&A departments (e.g. provenance).
- Better ways to manage gifts (e.g. in the FOLIO Acquisitions module), a process that is much more common in SC&A departments than in general libraries.
- Better ways to manage relationships among digitized and digital collection files (digitized analog, generations, directory structures).
- Tools for electronic records management (e.g. e-mail archives).
- Ways to handle algorithms in an archival setting. These have been defined by Cliff Lynch as "large, complex socio-technical systems (often today shorthanded as 'algorithms') that centrally involve some mixture of personalization, opaque rules, and machine learning components" (<http://firstmonday.org/ojs/index.php/fm/article/view/8097>).

2. Metadata Management SIG

- Ability to accommodate DC, MODS, METS, EAD, ISAD, ISAD (G), PREMIS, and other metadata schemas that are widely used by special collections and archives.
- Better descriptive modules and management tools for born-digital archival collections (especially born-digital video and complex digital files).
- Better tools for cataloging rare books and manuscripts (i.e. beyond MARC).
- Better ways to manage time-based media collections with complex relationships (i.e. analog to digital migrations), born-digital file migrations and normalizations, and now "data" collections and data subsets. Our current systems make these complex relationships difficult to track and surface to staff and patrons.
- Additionally, when SC&A departments publish digital collection content online, staff must invest a lot of time in creating MARC metadata (856 fields) and digital content. Example: the American Folklife Center (AFC) at the Library of Congress has modified its repository ingests to create some very minimal descriptive records, along with automatically generated PREMIS metadata. Minimal description, such as collection name and component, helps to mitigate some of the efforts and systems accessed for searching for content for patron access. The AFC also has a homegrown application that has historically maintained records for analog to digital relationships on a file-level basis, but this is being phased out. This particular metadata issue may be out of scope for FOLIO, however.
- Better ways to synchronize metadata across different systems. There is considerable effort spent creating, maintaining, and/or crosswalking metadata across various systems that describe the same set of materials. For example, for a given archival collection, we may create a detailed collection inventory in ArchivesSpace, a collection-level bib record in ILS (MARC), and records for

born-digital and digitized items in our repository. The hope is that FOLIO can provide better options for synchronizing metadata across systems to reduce redundant data entry and ongoing maintenance of the same metadata across systems. For more background on this question, see this 2015 blog post:

https://blogs.library.duke.edu/bitstreams/2015/09/18/metadata_synchronization/

3. Accessibility SIG

- Ways to record and provide support for restrictions on access, often date-driven (sensitive materials, copyright, donor agreements, etc.), particularly relating to born-digital materials but ideally across all collections.
- Discovery (at LC). The LC AFC primarily creates accessible collections via two levels of ILS records: a collection “stub” record and, if called for, ILS records for each and every component (in our case, usually the oral histories and interviews a collection is composed of) of a collection. AFC also creates accessible finding aids for a small subset of its collections. The Civil Rights History Project (CRHP) is an example of an LC collection that has been optimized for access using every tool in LC’s arsenal: <https://www.loc.gov/collections/civil-rights-history-project/about-this-collection/>. As part of the AFC, the Veterans History Project (VHP) has a unique website, whereby a MySQL database serves as the source for data for VHP’s stand-alone online presentation: <http://memory.loc.gov/diglib/vhp/html/search/search.html>. This is the most efficient way for users to access materials; VHP serves a fraction of its collections through the Library’s finding aid system.
- Discovery (at Auburn). The Auburn University Libraries is currently using an extensively customized version of VuFind as its discovery tool. The discovery interface includes MARC records from the Voyager ILS and DC records from CONTENTdm. Auburn also uses DSpace and (to a much lesser degree) Omeka as digital content management systems, but records from those systems have not been harvested into and indexed in VuFind. We are interested in a universal open-source discovery tool that can index and serve up records from all of our repositories, including an archival management system (we are currently testing ArchivesSpace and AtoM).
- Ways to shield or suppress Personally Identifiable Information (PII). Example: the Library of Congress’ Veterans History Project collects a significant amount of PII that is stored only in an internal database accessible only to staff, and would need to maintain that restriction in FOLIO or a future SC&A system.

4. Systems Operations and Management SIG

- Interoperability between FOLIO and existing SC&A systems (e.g. ArchivesSpace).
- A calendaring system (e.g. for restricted or embargoed collections).
- Modules for recording long-term conservation and preservation activities.
- Better transmission methods (documented APIs) between existing SC&A systems and FOLIO modules (e.g. ability to talk between an SC&A system and the FOLIO Acquisitions module).
- Connections (hooks, APIs) between archival management systems and large-scale discovery systems, both open-source and commercial.

5. Internationalization SIG

- Multilingual back-end and user interfaces (e.g. German, French, Spanish)
- Compliance with the EU General Data Protection Regulation (GDPR) and other data-protection laws