IJDC | *Peer-Reviewed Paper*

Getting to Beta: Building a Model Collection in a World of Digital One-Offs

Kathryn Gucer
University of Maryland

Kristina Adams
National Agriculture Library

Chuck Schoppet
National Agriculture Library

Ricardo Punzalan
University of Maryland

Abstract

Libraries and archives are increasingly producing subject-based digital collections alongside, but separate from, their main digital collections. These smaller projects are often treated as digital one-offs; they are created, launched, promoted, and then largely forgotten. The authors of this study argue that small-scale digital collections should instead be treated as test cases for their institutions' main digitization programs. Because they are lightweight and have relatively low stakes, these collections get pushed through the system quickly and can illuminate its workings and shortcomings in a snapshot form. The authors treat their own experience in developing the Animal Welfare Act History Digital Collection at the National Agricultural Library as a case study in using a digital collection to test and revise an institution's digitization program. In so doing, this study suggests how agile projects like the AWAHDC can be core components in digital curation policies and their implementation.

Received 16 January 2018 ~ *Revision received* 2 December 2018 ~ *Accepted* 10 December 2018

Correspondence should be addressed to Kathryn Gucer, 4th Floor, 4130 Campus Drive, College Park, MD 20742. Email: kathryn.gucer@gmail.com

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v13i1.554

Introduction

In September 2015, the Animal Welfare Information Center (AWIC), based at the National Agricultural Library, embarked on a project to create a subject-based digital collection: the Animal Welfare Act History Digital Collection.[1] For AWIC and the library, where the authors of this paper work and conduct research, this resource is a new tool in fulfilling an important mission of the center to inform the members of the community regulated by the Animal Welfare Act about the intent and history of the act. At the outset, we imagined the collection would be a free, full-text, and fully searchable Internet resource of 900 US government publications in PDF. But, while the library had produced a similar collection the year before, the participants in this earlier project had left no record of the digital production process for us to follow. As we talked to these participants about their experience, the root cause for this absence of documentation emerged: there was an underlying view among them that these small-scale digital initiatives were secondary to, and even a distraction from, the library's main digital program, the National Agricultural Library Digital Collections (NALDC). The latter had been in place since the mid-2000s and contained over 84,000 items at the time that we launched the Animal Welfare Act collection. To many staff members, "little projects" like our subject-focused digital collection were an extra burden on already heavy workloads. Nor was it clear how these digital projects fit into the larger vision for digitization at the library and, by extension, why staff should prioritize them.

In this article, we argue that libraries should treat digital projects like the Animal Welfare Act collection as core elements of their programs in digital curation, not as digital one-offs. These small-scale projects uniquely point out the strengths and weaknesses of their host institution's main digitization program. Because they are lightweight and have relatively low stakes, these collections get pushed through the digitization pipeline quickly and illuminate its workings and shortcomings in a snapshot form. They can thus serve as diagnostic tests of systems that are otherwise hard to assess because they are constantly evolving with the pace of technological change and with changes in staff and their expertise. Most importantly, these projects enable otherwise disconnected staff to get "on the same page" by understanding their interconnected roles in the digitization process, especially as these roles shift. Here we lay out a production plan that treats the backend production of our collection as a model for the testing, evaluation, and recording of a library's digitization system. Specifically, we spell out a repeatable three-step procedure grounded in the single largest lesson we learned throughout the backend design and implementation of this digital collection – that is, document, don't forget, the process as it unfolds:

Step 1: Document the Process
Step 2: Test the System
Step 3: Revise the Document

What follows is a case study in three parts. In the Background section, we describe the context to our decision to use the Animal Welfare Act collection to test and

1 For easier reading, we will refer to this collection as the Animal Welfare Act collection in the rest of this paper.

document the digitization process at the library. We then detail the three-step procedure that emerged from this test in the Procedure section, providing precise lessons learned at the end of each step in order to spell out relevance of these details to digitization teams more broadly. In the Reflection section, we expand on the importance of the model as whole for the National Agricultural Library and for other libraries with similar digitization programs.

Background

In this section we describe the context behind our decision to model a process of digital production and review on the Animal Welfare Act collection. As indicated above, we based our decision on an unexpected lack of documentation about the digitization process at the library. When we embarked on the Animal Welfare Act collection, we thought that we would model our collection, and the process of digitizing it, on another subject-based digital library that had been developed and launched in the previous year: the Historical Dietary Guidance Digital Collection.[2] This collection, the first of its kind at the library, was sponsored by another information centre at the library, the Food and Nutrition Information Center (FNIC), and developed with the help of two scholars at the University of Maryland's College of Information Studies. The Dietary Guidance collection team had produced a valuable scope of their collection at the beginning of the project (James and Punzalan, 2014), which we used as a model to scope the contents and rationale behind the Animal Welfare Act collection. But as we moved from defining our collection to implementing its production, we found that no such documentation existed for the digitization of the earlier collection. By "digitization," we mean the process by which an item chosen for inclusion in a digital collection at the library moves from its origins as a physical item on the shelves to its ultimate destination in the library's digital repository, fully converted to an electronic format and attached to the metadata that best facilitates its discovery.

Looking back on the Dietary Guidance collection with the staff who had been involved in the project, we learned that there were two main reasons for the absence of a global perspective on digitization at the library:

- **Communication Barriers**: Although the earlier project had been completed and launched, we learned that not everyone involved in it fully understood each other's role(s).[3] Like the departments in many libraries, we learned, the National Agricultural Library's divisions were siloed, a situation that did not lend itself to effective communication, cross-departmental engagement, or a collective understanding of how the digitization process worked.

- **Technological Change**: Although the backend software (i.e., Islandora) for storing collection items and making them discoverable had been in place for over four years, the digitization team at the library had not yet used this technology to create digital collections. The library had adopted Islandora as a backend interface to the library's digital repository, known as the unified repository, four years earlier and, in that period, had used Islandora to build the library's article citation index, PubAg. With the Animal Welfare Act collection,

2 Henceforward we will refer to this collection as the Dietary Guidance collection.
3 We will describe these departments and their particular perspectives in more detail in the Procedure section.

 doi:10.2218/ijdc.v13i1.554

the digitization team would use this important tool to create a digital collection for the first time.

These factors significantly shaped our decision to treat the Animal Welfare Act collection as a model within the library. We would use the collection to get, and document, a global perspective on digitization at NAL, documenting the whole system as we digitized a particular collection.

Another important context for our decision to model the Animal Welfare Act collection in this case study was the absence of such a model in library and information studies. Although small, subject-based digital collections have become a focus in recent scholarly publications (Barton, Dixon, Skopelja and Javed, 2013; Beard, 2017; James and Punzalan, 2014), these studies do not model digital production or the evaluation of digitization workflows. While there were guides on how to build digital collections more generally (Lee, 2002; Purcell, 2016; Zhang and Gourley, 2008), these books did not address our specific needs. Aimed at libraries building large-scale diverse programs from the ground up, they had little to say about sending a small, relatively uniform, collection through a complex legacy system. Questions covered extensively in these guides – such as selecting materials for digitization from large diverse collections, digitizing diverse formats, choosing institutional software and metadata schemas, and determining best practices in these and other areas – were not relevant to us because these issues had already been decided and tested by the library. Conversely, the questions we were facing were addressed minimally in this work, if at all. These questions centred on getting the loosely connected participants in this pre-existing system up-to-date and on the same page so that they could interact with each other with maximum efficiency and effectiveness. We were also exploring which features and workflows in the existing, and evolving, technology would work best for us and the library more generally.

Although the literature did not provide us with an exact model or guide for our work, it did suggest that such a model could be useful outside of the National Agricultural Library and also how we might go about creating one. For one thing, the literature suggests that libraries are very much in need of tools for modelling their programs and policies in digital curation. As of 2016, only 26% of institutions in the Association of Research Libraries (ARL) who responded to Alexandra Dressler's request for them had policies in digital preservation, a term encompassing the processes in digital curation addressed by our model. Even more to the point, Dressler concludes that the policies she did receive reveal a "gap" in the understanding of "exactly *how* this work would be addressed and *who* would be completing such work" (Dressler, 2017). She surmises that this situation is "likely quite common" at many "institutions with regard to digital preservation efforts" and repeatedly notes that policy in these institutions lags behind practice because of challenges that resonate with those we describe above – specifically, "rapid change in technology," "staff expertise," and "a lack of understanding and supporting documentation/policies." The model we detail below addresses this gap between theory and practice directly by providing a step-by-step guide for staff to create workflows of *actual* practice that can be revised collectively to produce *best* practice (the stuff of policy) and a common understanding going forward. Also, because it is iterative and flexible, our model directly addresses the problem of "rapid change in technology," enabling staff to perform diagnostic tests frequently and relatively easily.

Furthermore, a number of studies show that a preference for small-scale digitization projects in libraries is emerging just as many mid- to large-size libraries and museums

are re-evaluating their digital programs after the first two decades of life (Bertacchini and Morando, 2011; Hochstein, Gemoets and Goshorn, 2014; Kar, 2016; Sadler, 2016). This suggested to us that the time was ripe to use digital projects like the Animal Welfare Act collection to draw lessons for digital curation more broadly. With this in mind, we borrowed elements of a methodology described in some of these studies: agile development. Originating in the software industry, agile development emphasizes flexible, unencumbered, and responsive teamwork in the design and implementation of digital applications (Dingsøyr, Nerur, Balijepally and Moe, 2012). Following the library-based examples to this approach in Dulock and Long (2015), Niemi-Grundström (2014), and Cervone (2011, 2012), we incorporated three of the 12 principles in the agile methodology into our digitization project:

- **Small team(s)**: We engaged the remaining staff members who had been involved in the Dietary Guidance collection (and who would also be involved in our project) as a team in the production of a scoping document covering what was known about the digitization process at the library.
- **Lean mentality**: We made our collection as small as possible so that its size and complexity would not be an unnecessary drag on the digitization process. Where the Dietary Guidance collection had been comprised of 900 digital items at its launch, the Animal Welfare Act collection was comprised of 200 digital items.
- **Incremental, iterative approach**: We decided to soft-launch the Animal Welfare Act collection as a beta version in the expectation that we would later expand the collection in phases with the lessons learned from the first round of production in mind.

Each step in the procedure that we lay out below corresponds to one of these principles, as articulated in the Manifesto for Agile Software Development. Step One, in which the team comes together and documents the system as it is understood, is based in the principle that "the best architectures, requirements, and designs emerge from self-organizing teams." In the library's environment, where knowledge was compartmentalized, we recognized the need for a small team made up of members from all sections involved in digitization and unified by a "shared understanding of," and crucially, a "commitment to solving," the problem (Niemi-Grundström, 2014). Step Two, in which we use our collection to test this system and evaluate the first set of results, embraces the idea that "simplicity – the art of maximizing the amount of work not done – is essential." Hence our effort to scale down the collection from 900 items in multiple formats (e.g., .jpg, .pdf, and .mpg) to 200 items in one format (.pdf). This procedure privileges the "simple design" of the test subject (the collection) in order to achieve "technical excellence" (Dingsøyr, Nerur, Balijepally and Moe, 2012) in the system as quickly as possible and proceed to the next round of testing with this success as a foundation. Finally, Step Three incorporates the agile emphasis on frequent self-evaluation, in which "the team reflects on how to become more effective, then tunes and adjusts its behaviour accordingly." For us this idea meant fixing and documenting the solutions to the problems *as we encountered them*, instead of taking on the "monolithic" task of trying to anticipate all possible problems in advance (Cervone, 2012).

 doi:10.2218/ijdc.v13i1.554

Step One: Document the Process

Step One was to identify and engage all remaining staff in the **documentation** of digitization workflows describing the process for the Dietary Guidance collection. We sought to get everyone involved in digitization on the same page about who did what, how, and when in the process. Experienced librarians have long cautioned that collaboration among the diverse parts and players involved in digital production is essential to building digital collections (Hunter, Legg and Oehlerts, 2010) and the Animal Welfare Act collection would be no exception. We would have to unite disparate staff members in three of the four divisions at the library – the Information Products Division; the Information Systems Division; and the Document Production Division – into one small but cohesive and flexible team. This team was composed of 12 staff members from the following departments within these divisions:

1. **Animal Welfare Information Center**: the center's Program Leader, who was a subject specialist in Animal Science and a postdoctoral scholar in digital curation based at the University of Maryland, College of Information Studies.

2. **Cataloguing**: Two cataloguers who are experts in machine-readable cataloguing (MARC) standards and have deep experience creating MARC records for the library's online catalog, Agricola; and one librarian with expertise in converting MARC records to Metadata Object Description Schema (MODS) and ingesting them into the unified repository.

3. **Digitization**: Two digitization librarians and two cataloguers who work for the Internet Archive,[4] an external contractor involved in digitization at the library to scan physical items, create metadata for them, and transfer them to the Internet Archive's servers.

4. **Data Management**: The manager of the library's digital asset management architecture (Fedora/Islandora) and two managers of the automated indexer (Solr). Collectively, they harvest the digital items and their metadata from the Internet Archive and other sources, transfer them to the library's digital repository (an instance of Fedora Commons known at the library as the unified repository), and manage the dynamic indexer.

As mentioned in the Introduction and Background sections, these groups had worked together on the NALDC and the Dietary Guidance collection. But, we learned, no new items had been added to the NALDC in four years and during that time significant changes had been made to the software for ingesting and preserving items in the unified repository. Specifically, the library had incorporated a new staff interface, Islandora, to the unified repository. Islandora is a digital asset management system that provides librarians with a user-friendly tool to create and maintain complex digital asset systems (i.e., everything from scientific datasets, to article citations, to audio/video collections). We will go into more detail about Islandora and the reasons for the policy shift below, but the important point about this new tool to be made here is that it would be used to create digital collections at the library for the first time with the Animal Welfare Act collection. Because they would be based on a digitization process that was outdated, the first workflows our team produced would be incomplete and, at times,

4 The Internet Archive is a non-profit digital library that offers free universal access to books, movies, and music.

incorrect. We would therefore have to treat these initial workflows as working documents, using them to establish a common understanding among all the team members at the outset. We would later revisit and revise the workflows (ultimately, the goal of Step Three in our model) as our knowledge became more complete.

With the idea of documenting and modelling the Animal Welfare Act collection in mind, we distributed rough and highly provisional mock-ups of workflows to all the team members involved, admitting up front that these documents were flawed and in need of as much modification and amplification as the team could provide. Since we wanted to promote full communication from all these departments, we met with them separately to discuss these materials and invited them to modify our mock-ups in advance. This procedure enabled everyone to communicate fully and openly and led to enough consensus to set the digitization process in motion. Most importantly, producing the workflows immersed us, the authors of this paper and the project leaders (who were not digitization experts), in the details of digital production and preservation at the library and, thereby, illuminated parts of a process that were obscure, if not completely unknown, to us when we started. At the end of Step One, we had a working document that offered us and the rest of the digitization team a global view of the digitization process for the Animal Welfare Act collection.

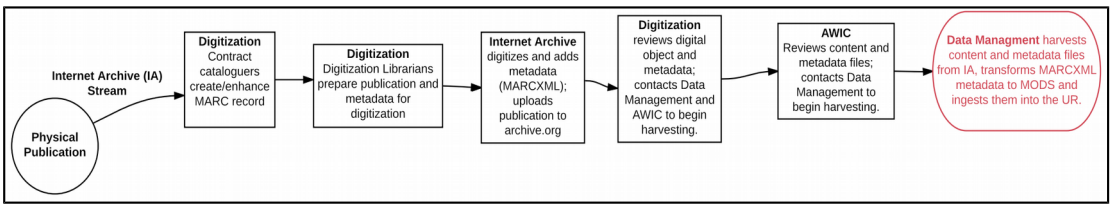

Figure 1. Simplified detail from initial workflows.

A simplified version of these initial workflows is presented in Figure 1.[5] The red box indicates the part(s) of the process that were the most mysterious to us after we had distributed and revised the workflows. Although the entire document was subject to change, we suspected that we would have to revise this part of the process considerably later on. Significantly, these workflows highlighted to us how much of the digitization process hinged on generating metadata files and converting them into and out of different digital formats (e.g., MARCXML, MODS). This insight was an early sign of the complications to come, as we investigate further in Step Two.

Lessons Learned

1. **Let the team emerge organically:** Although we did not realize it at the time, when we first began documenting the existing digitization process at NAL, we were simultaneously building an engaged, committed team for revising that process. The 12 staff members who ultimately comprised our team were a subset of the larger group of people whom we originally solicited for feedback on the digitization process. As we engaged in the back-and-forth interactions with the larger group, a sense of the team emerged organically among those most

5 The "before" and "after" views of the full workflows are available in the Digital Repository of the University of Maryland (DRUM): http://hdl.handle.net/1903/21103

 doi:10.2218/ijdc.v13i1.554

interested in understanding and solving the problem. This experience suggests that libraries should follow the agile preference for ground-level self-organizing, rather than groups organized by leadership from the top down.

2. **Don't let the perfect be the enemy of the good:** As the incomplete picture of the process in the red box suggests, we did not walk away from Step One with a perfect understanding of the system. This gap in our knowledge led us to debate about whether to hold back with the digitization of our collection. But we ultimately decided to proceed with a much smaller batch of documents than originally planned. With this manageable subset, we believed, we would be able to work through the snags and grey areas in the process as we encountered them. We therefore recommend letting go of perfection. Digitization teams will find that this flexible attitude has at least two benefits: i) it keeps the team from getting bogged down in details, with the related risk of pursuing unnecessary work; ii) it also maintains a sense of momentum among the team members.

Step Two: Test the System

Step Two was to run the lean collection through the digitization pipeline documented in the workflows and evaluate the results. As we suspected, this first round produced mixed results. When we accessed the collection in the Islandora interface, we found that the metadata which was supposed to make the Animal Welfare Act collection discoverable was so incomplete and, at times, incorrect that the collection was effectively unusable. One recurring error in the metadata illustrates this overall problem: the absence of a source field and the relevant data in many records that resulted from searching the collection. This error can be seen in Figure 2, the record in the user interface for an item in the Animal Welfare Act collection: "Bill Green's Efforts for Animal Welfare," a two-page excerpt from the government serial, *Congressional Record*. This item contains the remarks in favor of expanding the US government's protection of animals, made by Bill Green, a member of the House of Representatives, on June 2, 1982.

Figure 2. Errors in "Bill Green's Efforts for Animal Welfare."

There are multiple obvious errors in the metadata in Figure 2, but the one we will focus on is the absence of a key piece of information about this item: its source. Although there are fields for the title, author, subject, year, and the content file, there is neither a source field nor the data that field should contain (in this example that missing data is *Congressional Record*, 1982, v. 128), even though our users would need this information in each record and all members of our team expected it to be there. Once

we saw this error, we acted quickly while the digitization process was still fresh in every one's mind. We convened brief, sometimes impromptu, meetings with individuals from the groups on the team to identify where in the process these errors had crept into the system.

Ultimately, we traced the origins of this error to a hazy understanding among the team members about how Islandora, the open source software used for ingesting digital objects into the Animal Welfare Act collection, works. Islandora is made up of three core components: a Drupal interface for its backend users; the Fedora repository used for storage (at NAL, the UR); and the Solr index, which uses the metadata to make the collection searchable. In the digitization workflow above, Islandora is primarily used by the staff in Data Management at the end of the process for ingesting the content and metadata files created, at the beginning, by the cataloguers in Cataloguing and Digitization into the unified repository (see the red box in Figure 1). The results of our trial run pointed to a disconnect between these two subgroups of our team. The cataloguers who designed and created the metadata for "Bill Green's Efforts for Animal Welfare" did not fully understand how their work related to that of the technologists in Data Management who transformed this metadata into a format that the Solr index could use. As the workflow shows, the cataloguers created a MARC record for "Bill Green's Efforts for Animal Welfare" which they then uploaded as a MARCXML file to the Internet Archive (archives.org). After this upload, the staff in Data Management harvested this file from archives.org, and ingested it into the unified repository.

What the red box in the workflow *does not* show is that, during ingest, Islandora automatically generates a second metadata file in another format, Metadata Object Description Schema (MODS), from the original MARCXML file. It is this second MODS file that the Solr index uses to search the items in the collection and to display the results in the records like the one for "Bill Green's Efforts for Animal Welfare" in Figure 2. The cataloguers who designed and created the metadata for this item not did understand this workflow in Islandora and, more specifically, the complex ways that Islandora handles the MARCXML that results from their work. Specifically, after the staff in Digitization uploaded the MARCXML file to archives.org, the staff in Data Management harvested these files from archives.org, and ingested them into the unified repository using the Islandora interface. But the cataloguers in Digitization had unwittingly put the data about the source of this item into a field in MARC, which was in turn moved to a MODS field (known as a "data element") that the Solr index had not been configured look for. Unable to find the data element for the source of "Bill Green's Efforts for Animal Welfare" in the MODS file, Solr omitted the source field from the record for this item. As the multiple errors in this one record suggest, this kind of mismatch in the conversion from MARC to MODS data elements – known as crosswalking – happened repeatedly. In fact, all of the items in the AWADHC had been adversely affected by this error.

Lesson Learned

1. **Embrace Failure:** "Fail fast," and the related term "pivoting," is a motto among proponents of the agile method: the idea of working with and from failure to get to success. Librarians often favour caution and meticulous planning over taking risks, however calculated. Our experience in Step Two argues for embracing risks and the failures that may follow them. The results of the first round of digitization were not at all what we wanted or needed, but we quickly

 doi:10.2218/ijdc.v13i1.554

adjusted our response to learn – and learn *how to* learn – from this failure in the metadata conversions. We therefore suggest building an expectation of failure into the process of modelling digitization projects like the AWAHDC in the knowledge that the fear of failure can hold the team back from learning and creating new paths forward. Calculated risks have a number of benefits. They inure the team to uncertainty and build resilience in the face of its consequences. They can also save time by focusing the team on actual problems in the system, not imagined ones.

Step Three: Revise the Document

At this point, Step Three was clear: revise the workflows to document the processes that were hidden in and/or inadvertently left out of the workflows. At the ground level, we knew that we would have to revise the MARC-to-MODS conversion process so that it consistently transferred data into MODS elements that the Solr index could be configured to access and use. In the case of "Bill Green's Efforts for Animal Welfare," the source information would very likely have to go in the MARC 773 field (Host Item Entry), where the series title (*Congressional Record*), volume (128), and date (1982) could be parsed into separate, named subfields. Thus separated in the MARC record (and in the subsequent MARCXML file), Islandora would be able to crosswalk these data into a MODS element in which they were similarly parsed (see these corrections in the MODS file for "Bill Green's Efforts for Animal Welfare" in Figure 3). But, while understanding these particular details was necessary to correct the errors in this first round of digitization, it was too specific to include in the amended workflows. The latter would have to serve as a model for future digitization projects at the library, which would have varying kinds of data, and data elements, in their metadata files. In order for them to be useful, the workflows would have to be broad and flexible enough to accommodate these and other variations.

Figure 3. Detail from corrected MODS file for "Bill Green's Efforts for Animal Welfare."

With this larger, more adaptive, perspective in mind, we changed the workflows to address explicitly how the data elements in MARC were chosen, and later in the process, evaluated. If incorporated into digitization practices more broadly, these procedures could be core elements of digitization at the library, not just the Animal Welfare Act collection. Accordingly, we revised the workflows to:

1. Reflect the need for communication between the cataloguers in Cataloguing and the technologists in Data Management early in the digitization process.

2. Incorporate quality control procedures in Data Management to perform checks on the metadata after ingest.

We inserted the first change into the beginning of the workflows specifying that the MARC cataloguers communicate with the Solr experts at the beginning of the process, as they design the metadata templates for the MARC files. Together, these team members would establish that the pathways from the MARC to the MODS files were clear and that the conversions would produce clean, desirable metadata that could be accessed by the Solr index. The second change went into the red box at the end of the workflows. As mentioned earlier, this box contained the parts of the process that were the most mysterious to us after we had distributed and revised the workflows. On the advice of the Data Management staff, we decided to use quality control checks that are built into Islandora to review representative samples of metadata periodically in the digitization process. These checks add an evaluative layer to the ingest process, enabling the staff in Data Management to review the implementation of the MARC-to-MODS conversion. The next round of digitization in the Animal Welfare Act collection would, in part, be a test of these controls and their usefulness. Overall, these changes to the workflows (see the green text in Figure 4) were relatively minor. But they represented a significant advance in the communal knowledge of the process and confidence in the team's ability to work together. We were now ready to repeat steps one through three in the second round of digitization, a repeat of steps one through three with a new batch of collection items. In this second round, we would put what we had learned in the first to test and further refining our knowledge.

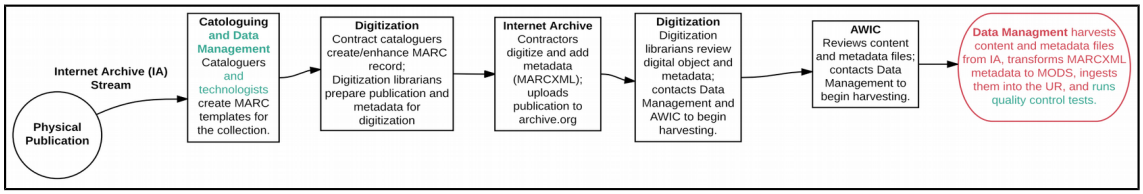

Figure 4. Revised workflow from Step Three.

Lesson Learned

1. **Focus on the forest, not the trees:** It is a commonplace that people immersed in a project can easily lose sight of its larger significance, so involved in the details that they "can't see the forest for the trees." Throughout steps one and two, and the first part of the process described in step three, we found ourselves caught up in the technical details of the project, as the many acronyms for metadata standards and software dotting the text above suggest. This focused, ground-level perspective worked its way into the first drafts of the workflows in more ways than one, though it is not obvious from the versions presented above. For instance, we used the personal names of the staff members involved (e.g, "John Doe"), instead of their official roles in the process and at the library (e.g., "Contract Cataloguer") and the explanations were sometimes too detailed for someone outside of the team to understand. Nonetheless, by the time we reached

 doi:10.2218/ijdc.v13i1.554

this third step, the cumulative effect of reflecting on our experience in writing subtly prompted us to step back from these details and see in them – or rather *past them* – to their larger significance for the digital curation at the library. We therefore recommend paying attention to these "a-ha!" moments, especially when they widen the team's perspective on the process unfolding in the work before them. In abstracting from the particulars in one project to a generalized perspective, the team can build longevity into the model and produce a record that other teams can follow.

Reflections

In this section, we reflect more broadly than above on the importance of this model for the National Agricultural Library and other libraries with longstanding and expanding digitization programs. At the National Agricultural Library, the biggest impact of our effort to model the Animal Welfare Act collection has been the collective understanding of the digitization process that has resulted from documenting the system. Before we embarked on our project, staff members at the library had a hazy and patchwork understanding of the whole process involved in digitizing collections like the Historical Dietary collection and the Animal Welfare Act collection. The subgroups on our team operated without a clear sense of how their work was instrumental in the larger digitization process. This misunderstanding, and its less-than-optimal results, had led to a lack of confidence in the system and a diminished appetite for embarking on these "little projects." After the first round of digitizing the Animal Welfare Act collection, we can now say that the members of our team have achieved a detailed working knowledge of the digitization pipeline and the process of sending collections through it. While no one understands their colleagues' roles well enough to perform them, each team member knows concretely and specifically that the success of the digitization process hinges their interlinking roles and their ongoing communication about them. Relatedly, the collective creation, implementation, and revision of the workflows – which document who does what and when in the process – means that team members know whom to approach with questions when they arise.

The importance of documenting workflows to the development of effective digitization workflows cannot be understated. This aspect of our model built a much-needed element of self-reflection into the system, creating room for "a-ha!" moments like the one that illuminated the glitch in the MARC-to-MODS conversions of the metadata for "Bill Green's Efforts for Animal Welfare." Nor are such moments only significant at the production level; they can reveal questions whose answers have larger curation policy implications. For instance, the trouble we encountered in converting MARC to MODS metadata raised a number of such questions, including: would it have been better to create new metadata in the more granular MODS instead of converting from a MARC record? The long-held practice at the library has been to create a MARC record automatically for every item, whatever its format, so that it could be represented in the library's Voyager-based catalogue, Agricola, which uses the MARC metadata standards. But the more we grappled with the complexity of the MARC-to-MODS conversions, the more that we found ourselves wondering about the need for MARC metadata in our collection – at least, at the beginning of the process. Originally designed for cataloguing physical items, MARC is less suited to the needs of digital users than MODS because it does not contain an item's description in a granular form – a form, as

we have seen in the example of "Bill Green's Efforts for Animal Welfare," that an indexer like Solr can capture and use. While it may ultimately be important for our users to find records of the Animal Welfare Act collection in Agricola, our immediate priority was Web discovery. For this reason, we began to think through the possibility of shifting the library's policy away from the emphasis on MARC metadata in digital collections development. As we write, this is still an open question and it is not one that our team can decide on our own, since it has wider implications for digitization and collections management at the library as a whole. But without the intense focus on the metadata conversion process in the Animal Welfare Act history collection, the library would not be asking this important question and considering its larger impact.

Finally, as this open question and the process by which we formulated it suggests, our experience points directly toward the ways in which small-scale digitization projects can impact policy in digital curation at libraries that, like the National Agricultural Library, have longstanding and expanding digitization programs. Our experience has shown that libraries are in need of policies that can respond to the ever-evolving digital technologies and the equally fluid human expertise necessary to keep pace with them. Top-down global policies on digitization, like those envisioned by traditional workflow diagrams, will not by themselves capture these fluid conditions. This was the overarching lesson we learned after sending the beta version of the Animal Welfare Act collection through the pipeline initially envisioned in the workflows: the process we were chasing was not stable and, therefore, could not be fixed in a single perspective, at least not for very long. We therefore tailored our iterative, three-step procedure to this reality. When problems arose, while they slowed us down, they did not impede us from identifying and fixing the glitches, revising the process, and moving on to the next round of digitization. The errors resulting from our imperfect knowledge of the MARC-to-MODS conversions, for instance, were manageable. This is the core strength of incorporating small agile projects into library planning and policy: they provide the managers of large complex legacy systems with focused, time-limited tasks that allow for testing and diagnosis of the whole system.

Conclusion

Elizabeth Yakel defines digital curation as "the active involvement of information professionals in the management, including the preservation, of digital data for future use" (2007). If valued for their ability to diagnose and modify digitization programs in libraries, small-scale digital collections can be core tools in this active involvement. What members of the digitization team at NAL initially referred to as "little projects" are located in the critical space where the rubber hits the road – that is, where the ideas behind digital programs in libraries meet the current practical realities of implementing them. They hold out the promise of enhancing, streamlining, and evolving the digital resources libraries offer their users. We recommend taking them up on their offer.

Acknowledgements

The authors would like to thank Adam Kriesberg and Paul Wester, who read and commented on earlier drafts of this study.

References

Barton, A.J., Dixon, B.E., Skopelja, E.N., & Javed, S. (2013). A case study of a semantically enhanced public health digital collection. *Journal of Library Metadata, 13*(4), 367-380. doi:10.1080/19386389.2013.849908

Beard, I. (2017). The eBethArké Syriac digital library: A case study. *Digital Library Perspectives, 33*(1). doi:10.1108/DLP-07-2016-0017

Beck, K. et al. (n.d.). Manifesto for agile software development. Retrieved from http://agilemanifesto.org

Bertacchini, E. & Morando, F. (2013). The future of museums in the digital age: New models for access to and use of digital collections. *International Journal of Arts Management, 15*(2), 60-72.

Cervone, H.F. (2012). Understanding the elements of a digital library project plan: Part 1. *OCLC Systems & Services: International digital library perspectives, 2*8(2), 75-78. doi:10.1108/10650751211236604

Cervone, H.F. (2011). Understanding agile project management methods using Scrum. *OCLC Systems & Services: International digital library perspectives*, *27*(1), 18-22. doi:10.1108/10650751111106528

Dingsøyr, T., Nerur, S., Balijepally, V., & Moe, N.B. (2012). A decade of agile methodologies: Towards explaining agile software development. *Journal of Systems and Software, 85*(6), 1213-1221. doi:10.1016/j.jss.2012.02.033

Dressler, V.A. (2017). The state of affairs with digital preservation at ARL member libraries: A survey and analysis of policy. *Digital Library Perspectives*, *33(2)*, 137-155. doi:10.1108/DLP-08-2016-0030

Dulock, M.J., & Long, H. (2015). Digital collections are a sprint, not a marathon: adapting scrum project management techniques to library digital initiatives. *Information Technology and Libraries (Online), 34*(4), 5-17. doi:10.6017/ital.v34i4.5869

Hochstein, C., Gemoets, D., & Goshorn, J. (2014). Ten years of change: National library of medicine toxmap gets a new look. *Medical reference services quarterly, 33*(4), 428-437. doi:10.1080/02763869.2014.957090

Hunter, N.C., Legg, K., & Oehlerts, B. (2010). Two librarians, an archivist, and 13,000 images: Collaborating to build a digital collection. *The Library Quarterly, 80*(1), 81-103. doi:10.1086/648464

James, C., & Punzalan, R.L. (2014). Legacy matters: Describing subject-based digital historical collections. *Journal of Archival Organization, 12*(3-4), 198-215. doi:10.1080/15332748.2015.1150104

Kar, D.C. (2016). Digital Library of India: An initiative for the preservation and dissemination of the national heritage and rare books and manuscripts collection. *World Digital Libraries – An international journal, 9*(1), 45-60.

Lee, S.D. (2002). *Electronic collection development: A practical guide*. Neal-Schuman Publishers, 100 Newfield Ave., Edison, NJ 08837.

Niemi-Grundström, M. (2014). Developing, evaluating and managing library with agile methods. *Library Management, 35*(6-7), 481-485. doi:10.1108/LM-02-2014-0022

Purcell, A.D. (2016). *Digital library programs for libraries and archives: Developing, managing, and sustaining unique digital collections*. American Library Association.

Sadler, S. (2015). Building a digital library: What to expect as a technology project manager on a library construction project. *Journal of Library Administration, 55*(3), 221-234. doi:10.1080/01930826.2015.1034048

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International digital library perspectives*, *23*(4), 335-340. doi:10.1108/10650750710831466

Zhang, A. & Gourley, D. (2008) *Digital collections: A practical guide*. Chandos Publishing, Oxford, England.