Agile-ish: Bringing Agile and Scrum into Project Management for Digital Collections Metadata
Linda Ballinger

Abstract

Implementations of agile values and principles are increasingly seen in project management beyond their original home in software development. Most library projects drawing on agile and scrum, agile’s most popular methodology, have needed to adapt these principles and methods to varying degrees, but most have been in environments similar to software development. The Pennsylvania State University Libraries’ cataloging department was interested to see if agile and scrum approaches could be successful in managing a metadata project involving an ad hoc team, composed of members volunteering part of their time to the project, and inexperienced in the work needed for the project. While the Penn State Libraries project used extensively modified versions of agile and scrum, we have concluded that using these principles and methods, even if adapted, can greatly improve the process and the outcome of many projects.

Introduction

Implementations of agile and scrum are increasingly seen in project management beyond their original home in software development. While the two terms, “agile” and “scrum,” are used almost synonymously in much of the literature, and in popular works online, it is useful to distinguish between them. Agile is sometimes referred to as a methodology, but it is better understood as a set of values and principles that describe an ideal organizational culture. Scrum is agile’s most commonly used methodology, or framework, but it is just one of several ways to achieve an agile work environment. Most library projects drawing on agile and scrum have tended to involve software development and to include information technology departments or personnel. Even with these similarities in environment, libraries have often needed to adapt traditional agile and scrum project management approaches. Most of the library teams described in the literature were similar to typical scrum teams, in that they were intended to be long-term and composed of experienced members. But they were frequently unable to follow the ideal agile practice of letting team members be entirely dedicated to one project at a time. The Pennsylvania State University Libraries’ cataloging department was interested to see if agile and scrum approaches could be successful in managing a project involving an ad hoc cataloging team, composed of members volunteering part of their time to the project, and inexperienced in the work needed for the project.

Agile in Libraries

Agile tends to be called many things: a strategy, a process, a mindset, a framework. It has a similar range of definitions in the library literature. Critchlow, Friedman, and Suchy (2010); Cervone (2011); Niemi-Grundström (2014); and Shein, Robinson, and Gutierrez (2018) all refer to agile as a “method” or “methodology.” Chang (2010) highlights agile as “a development and management framework for innovative projects” (672) and as “a leadership philosophy that encourages teamwork, self-organization and accountability” (675), a description that comes closer to the intentions behind agile.
At its core, agile is a set of values that can be applied to multiple ways of working. Agile grew out of the efforts of various software development communities seeking “an alternative to documentation driven, heavyweight software development processes” (Highsmith 2001). In 2001, a small group of representatives from these communities came together and drafted a “Manifesto for Agile Software Development,” which outlined the values they shared:

We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

**Individuals and interactions** over processes and tools

**Working software** over comprehensive documentation

**Customer collaboration** over contract negotiation

**Responding to change** over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

© 2001, the Agile Manifesto authors. This declaration may be freely copied in any form, but only in its entirety through this notice.

(Agile Alliance 2001 [emphasis in original])

These values were, according to the signatories, “based on trust and respect for each other and promoting organizational models based on people, collaboration, and building the types of organizational communities in which we would want to work.” (Highsmith 2001). From the start, agile emphasized the quality of the process in project management, as well as the product. Of the twelve “Principles behind the Agile Manifesto,” drafted by the same authors shortly after the manifesto, half are about how an agile team works together:

- Business people and developers must work together daily throughout the project.
- Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
- The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
- Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
- The best architectures, requirements, and designs emerge from self-organizing teams.
- At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

(Agile Alliance 2001, “Principles” page)

Like the manifesto authors, some libraries felt the need for an alternative way to manage projects. Critchlow, Friedman, and Suchy (2010) describe the drawbacks of traditional methods such as waterfall, which “[look] to have ideas complete, approved, and set in stone before moving forward” (Critchlow, Friedman, and Suchy 2010, ¶ 1). This approach to planning tends to take a long time and involves large numbers of representatives, since product requirements that are not included at the beginning of the project are difficult to introduce while the project is underway. Cervone (2011) adds that “requirements definitions are often so labor intensive and protracted that the requirements for the project have changed before development even
begins.” (18). The product or service being developed is not complete enough to be demonstrated until the end of the project, so feedback on the entire product happens near the end, making it very difficult and expensive to address that feedback. As libraries became more involved in developing innovative and user-responsive products and services, they also started looking to agile for ways to achieve their goals.

Library implementations of agile have tended to have a technology connection, with goals such as developing innovative web applications (Chang 2010), mobile websites (Critchlow, Friedman, and Suchy 2010), institutional repositories (French et al. 2016), or inventory management systems (Bales, Fox, and VanNevel 2017). But adoption of agile is gaining ground even for projects that are not technology-based. Many libraries have found that even partial implementation of agile practices have contributed to the successful completion of a variety of projects (Dulock and Long 2015). One library even attributed their project’s success to practices they only afterwards realized were characteristic of agile, explaining that they gravitated to these practices naturally (Shein, Robinson, and Gutierrez 2018).

Libraries usually have to bypass or modify some agile values and practices, even in implementations that closely resemble the software development context. Most, if not all, libraries found it impossible to entirely dedicate the time of team members to the project. Cervone (2011) emphasized that “no outside influence should be allowed to interfere with the work of the scrum team” during a sprint (20). But most of the authors write about needing to balance project work with other responsibilities.

In spite of needing to make adjustments, the libraries experimenting with agile project management were satisfied with the outcomes of their projects. Many pointed to the positive effects on staff motivation and morale. Chang (2010) described her team as “energized by the non-traditional active development and learning environment” (674). In French et al. (2016) the authors each noted significant improvements in their work and in their job satisfaction. Most of the library literature highlights the importance of communication and observes the benefits of agile project management approaches in this respect. Cervone (2011) references the lesser significance of “copious project documentation” in an agile project, in favor of “direct communication with partners in the development process” (19). Dulock and Long (2015) also emphasize communication over documentation, adding that communication “result[s] in less administrative overhead as well as increased accountability and trust between team members” (7).

Scrum in Libraries

Agile inspired the development of many different methodologies to implement its values and principles, with scrum being the most popular. Like agile, scrum has acquired many different definitions, but the developers of scrum stated that “scrum is not a process, technique, or definitive method. Rather, it is a framework within which you can employ various processes and techniques.” (Schwaber and Sutherland 2017, 3). That framework is characterized by scrum’s roles, events, and artifacts. Although Schwaber and Sutherland also state that partial implementations of scrum are not scrum (19), most libraries have found it necessary to adapt the roles, events, and artifacts that characterize scrum.

Scrum teams are composed of three roles: the product owner, the scrum master, and the development team. The product owner represents the stakeholders in the outcome of the project, and provides expertise in the requirements of the product or service being built. The scrum master coaches the other team members in agile and scrum processes and removes obstacles for the team. The development team is composed of the individuals who perform the
day-to-day work of the project. Ideally, the development team is cross-functional, so that all team members can perform all the activities needed for the project. Their time is devoted exclusively to one project at a time, and they are co-located to promote face-to-face communication.

These roles are intended to be performed by different individuals, but for the libraries reporting on their implementations of scrum, most found that some team roles had to go unfilled or be assigned to one person filling multiple roles. Cervone (2011) notes that project managers often fill the role of scrum master, while unit managers serve as product owners (20). Having a specific person for the role of scrum master was one of the things that most libraries needed to adapt to their own environment. French et al. (2016) initially questioned whether they could implement a scrum project at Virginia Tech when they had no one to serve as scrum master. Their project was successful without anyone specifically filling this role, but from French’s description of how she participated as Director of Digital Research Services, she filled most of the responsibilities of the role, while also acting as product owner (French et al. 2016, “Amanda French’s Thoughts: Implementing Scrum” section). Chang (2010) describes her responsibilities in their agile project as “[having] to provide the vision and technical skills as well as plan and coordinate the implementation” and “to be an enabler and facilitator…” (679). These responsibilities fit both the product owner and scrum master roles.

In scrum, work is done in sprints, which are units of time allotted for work on the project. These units of time are intended to be short—two weeks is a frequently used schedule. The sprint includes four events: sprint planning, daily scrums, sprint reviews, and sprint retrospectives (Schwaber and Sutherland 2017). Sprint planning enables the team to select and organize their work in a way that is responsive to change, since it is done one sprint at a time. Daily scrums are very brief meetings (15 minutes is common) conducted to give every team member an opportunity to communicate what they did the day before, what they will be doing that day, and what obstacles need to be cleared to do that work. The sprint review takes place at the end of the sprint and reflects on all the work that was done during the sprint, and often includes a demonstration of the product to stakeholders. Finally, the sprint retrospective provides the team with time to reflect on how their work went during the sprint, and what changes could be made to improve.

Scrum artifacts serve as the project’s primary documentation. They include the product backlog, the sprint backlog, and the increment (Schwaber and Sutherland 2017). The product backlog documents all the requirements for the project as a whole. Each sprint also has a sprint backlog, where the team identifies tasks to be done during that sprint. Finally, the increment is the result of the work that was done during the sprint. As one popular work on agile explains, “Whether the product is a website or a new house, the product increment should be complete enough to demonstrate its working functionality.” (Layton and Ostermiller 2017, 75).

Working in Teams at Penn State

The Pennsylvania State University Libraries’ cataloging department has a long history with self-organizing teams (Bednar, Brisson, and Hewes 2000). In the early 1990s, Penn State promoted an institution-wide adoption of Total Quality Management (TQM) and Continuous Quality Improvement (CQI). As part of this initiative, the Libraries’ cataloging department transitioned from a multi-tiered hierarchical organization to a team-based model that is still in place today. In this model, the department head provides administrative and supervisory control, as well as overall leadership for the department. But the day-to-day management of work is shared among the members of seven format-based teams, which include faculty members who serve as their team’s resource person, but not as their supervisor. Leadership of these teams is
rotated among all the team members. In addition to managing their team’s workflows, team members “assume responsibility for their own self-development” and their ability “to function well in a participatory and collaborative environment” (Bednar, Brisson, and Hewes 2000, 270). This reorganization took considerable time and effort, since “the shift from a hierarchical organization based on authority to a learning organization based on collaboration and empowerment was a radical shift in orientation for staff…” (Bednar, Brisson, and Hewes 2000, 251). But working in self-directed teams is now well integrated into the department’s organizational culture, making experimenting with agile practices for a new kind of team less of a reach.

When the Penn State Libraries began to contribute metadata for their digital collections to the Digital Public Library of America (DPLA) in 2015, metadata creation and management for digital collections was primarily the responsibility of the recently hired metadata librarian. The metadata librarian organized the first metadata remediation project using a group of volunteers from the cataloging department, the Special Collections Library, the Architecture and Landscape Architecture Library, and the Department of Art History. While this project met its immediate goals, there was little time to reflect on how the project went or how to continue metadata remediation work beyond this initial project. There was not enough regular work of this kind to justify establishing another permanent team in the cataloging department, but there was more work than one person could do. In addition, there was interest in expanding non-MARC metadata experience throughout the department, rather than confining it to a single team.

In 2017 the metadata librarian proposed forming the Digital Projects Metadata Team (DPMT), an ad hoc group that could be convened as needed to assist in remediating metadata for existing digital collections in preparation for contributing them to the DPLA. Team members were volunteers from the cataloging department and the Special Collections Library. The metadata librarian provided the team with training in updating and enhancing descriptive metadata and working in the digital repository system. For this project, the metadata librarian supplemented training by creating collection-specific documentation with instructions for each field. Creating this documentation took a significant portion of the metadata librarian’s time and the workload was difficult to sustain. The team completed remediation work for a few additional collections before other responsibilities took over the time of the metadata librarian, and the DPMT went on an extended hiatus.

Implementing Agile and Scrum at Penn State

In early 2019 the metadata librarian reconvened the DPMT, this time with the additional intention to experiment with agile and scrum practices to see if they could benefit project management in the cataloging department. Eight members of the cataloging department and one member from the Special Collections Library volunteered for this new project. Most of the team members had served on the DPMT in 2017, but three were new volunteers. The new DPMT functioned as the development team for the new project, though it differed from traditional scrum development teams in several ways. Scrum development teams are typically long-term teams, composed of members who will work together in their team arrangement over the course of many projects. Also, members are ideally assigned exclusively to the team, and are not expected to perform other work outside the team or the project. The DPMT, however, was composed of members devoting only part of their time to the project, and were meant to disperse between sprints. More significantly, the DPMT did not begin as a cross-functional team of experts, in which members share their individual areas of expertise so that all team members can contribute to all team activities. Instead, the entire team was participating in the same training while they were working on the project. Even the continuing members of the DPMT considered themselves to be novices at working with non-MARC metadata, and felt too out of
practice to want to start a new remediation project without additional training. So all the team members attended training sessions before the first sprint, and continued to learn on the job throughout the project.

Other scrum team roles also required significant adaptation. For this experiment, the metadata librarian had to serve as both product owner and scrum master. As the product owner, the metadata librarian populated and maintained the project backlog, defined the acceptance criteria, assessed the quality of work based on the acceptance criteria, and provided training and documentation. Since the metadata librarian was the only team member familiar with agile and scrum, she also served to some extent as scrum master. She advocated for the experiment with scrum, provided training in scrum principles and practices for the development team, and handled any problems getting in the way of the team's work. Throughout the project, the metadata librarian was assisted by the Metadata Implementation Specialist, a new position in the cataloging department that had been filled a few months before. The metadata specialist’s contributions during the project formed part of her own training; as she gained skills, she was able to serve as backup to the metadata librarian. As discussed above, other libraries faced similar challenges in filling all the roles in a scrum team. Having a manager serve as both product owner and scrum master is a common library adaptation. In libraries where agile and scrum practices are already familiar to the team members, having all the members contribute to the activities of a scrum master is also common.

The Penn State Libraries project was organized into two-week sprints, in which team members spent two to three hours a day working on metadata remediation. The original plan was to schedule these hours to occur at or near the same time of day, in order to provide as much coordination as possible within the team, since they could not be completely dedicated to the project or co-located in the same room. However, the team members’ schedules were too staggered and too full of other responsibilities to allow this kind of coordination. Daily standup meetings (i.e., daily scrums) were scheduled during the sprints, and were kept to fifteen minutes. This minimal obligation made it possible to find a time that most of the team members could commit to for most of the days of the sprint. At the standup meetings, each team member was asked to share the same three topics: 1) what they did since the last standup, 2) what they planned on doing before the next standup, and 3) what obstacles needed to be cleared to get the intended work done. The term “standup meeting” was used instead of “daily scrum” simply because it was already commonly used in the Libraries. The practice of standing for such meetings is intended to enforce the short duration of the meetings, but of course, standing was not actually required of everyone. Each two-week sprint was followed by a two-week break, when all the team members were able to focus on their regular work, then return refreshed for the next sprint. No set number of sprints was planned; rather the project was allowed to continue until the work was completed. The entire project ended up lasting three sprints, or ten weeks.

Other scrum events, especially sprint planning and the sprint review, were significantly modified for this project. Sprint planning was done by the metadata librarian and the metadata specialist before each of the three sprints. Planning for the first sprint included writing training documentation and conducting two training sessions: one to introduce agile and scrum principles and practices, and the second to bring all the volunteers up to speed on remediating descriptive metadata and working in the digital repository system. Training documentation was kept to a minimum, in an attempt to avoid the bottleneck that writing collection-specific instructions had created in earlier projects, and to encourage in-person and on-the-spot shared training. The documentation that was provided was a version of the Libraries’ data dictionary for descriptive metadata, with the addition of more help text and examples.
Sprint planning also included selecting the tasks to be done over the course of the sprint. Because the team members were in the beginning stages of learning the skills needed for the project, the metadata librarian and the metadata specialist selected collections from the repository and assigned them to each team member. For the first sprint they assigned very small collections, in order to give team members a chance to practice new skills without being overwhelmed. It was also possible for most of the team members to complete their first collection within the two weeks of the sprint, giving them an early sense of accomplishment. For the first sprint, there was no sprint review, but after the second sprint, the team conducted an adapted sprint review, in which they demonstrated their newly remediated collections in the test metadata aggregator to show how they had improved discovery for their collections. This demonstration took place at a cataloging department meeting, so the team was also able to talk about their experiences during the project with the rest of the department. Each sprint ended with a typical sprint retrospective, in which the team discussed among themselves how the sprint went, how they felt about the experiment with scrum practices, and what they would like to do differently for the next sprint.

The primary scrum artifact used in the project was a product backlog. Here, the product was the metadata resulting from the remediation project as a whole. Before the first sprint, the metadata librarian and the metadata specialist identified all the digital collections needing metadata remediation, and added them to the product backlog. Work was organized at the collection level, rather than as distinct tasks, since editing is done by collection in the digital repository system. Time or effort estimates were not attempted, since the team members were still learning, but time spent on each collection was tracked to enable estimates in the future. The product backlog was organized on a paper-based kanban board. Kanban boards are simple tools to visualize work in progress, typically composed of columns representing distinct stages of the work, and markers representing specific tasks, or other subsets of the work, that can be moved from column to column as work progresses. While there are several online tools that can organize and visualize work this way, an analog kanban board was chosen for its simplicity and ease in maintenance. The board was divided into Backlog, WIP (work in progress), Review, and Done columns, and sticky notes were made for each collection needing remediation. A separate sprint backlog was deemed unnecessary for the project, since the work to be done was not broken down into individual tasks. Each remediated collection was its own increment, where the product was the updated and enhanced metadata itself, and its effectiveness in making the resources being described findable, accessible, and understandable.

Results and Reflection

Working with a team of novices rather than experts meant that standups differed from traditional daily scrums in that they consisted mainly of team members asking questions of the metadata librarian and the metadata specialist. Since the metadata specialist was also still learning, standups were therefore unavoidably dominated by the metadata librarian answering questions and providing on-the-spot training. In spite of that, the team thought the standups were worthwhile. At the first sprint retrospective, team members noted that the standup meetings were the primary way they interacted with each other, and that the meetings gave them the chance to compare their questions and difficulties with each other. While the metadata librarian and metadata specialist were available to answer individual questions anytime the team members needed, team members often brought questions to the daily standup knowing that other team members would have similar questions. So the standups served as a way for the metadata librarian to convey consistent information and assistance to the entire team at once. They also gave team members a greater sense of being a team. One team member
wrote, “The stand-ups provide more immediate feedback and is helpful with not only answering questions some of us have or will have, but also with making it feel like a team—it’s nice to see the issues others are dealing with in their projects.” (Leah Oakes, email to author, March 15, 2019).

Surprisingly, the paper kanban board was one of the most popular elements of the project. Team members enjoyed using it, as it gave them a tangible way to measure their progress. Team members moved their collection marker from WIP to Review when they were ready to have their collections checked by the metadata librarian and metadata specialist. Once the review was finished the metadata librarian either moved the collection marker to Done, or back to WIP after meeting with the team member to give feedback and suggestions for further work. The board was placed in a high-traffic area of the cataloging department, providing transparency for the scope and progress of the project for both team members and department colleagues. The kanban board also attracted a lot of attention from colleagues from other departments passing by, and often served as a conversation starter about the project.

By the end of the first sprint, the team completed metadata remediation for four collections. While this result did not represent a large number of metadata records changed, the first sprint was very successful in terms of training and team building. During the sprint retrospective, team members expressed satisfaction with using scrum techniques and continued interest in practicing them. They did ask for a little more documentation, but not the collection-level instructions used in the past. The metadata librarian and metadata specialist met the request by enhancing the training version of the data dictionary with more explanatory text and examples. Development team members also asked whether the metadata librarian and metadata specialist were still taking on too much work in sprint planning and in reviewing each team member’s work. The whole team discussed possible changes in the workflow as the team members became more experienced. Among the ideas considered for future projects are peer review of metadata work and involving the entire scrum team in sprint planning.

At the final sprint retrospective, team members discussed how to improve the sprint experience for all team members. Several members of the development team felt that they worked better when they dedicated longer periods of time to metadata remediation work, so they tended to work all their sprint hours over a couple of days. This preference actually aligns better with preferred scrum practices, in which teams are dedicated to one project at a time, to reduce task switching. Rubin (2012) describes the extra costs in efficiency and quality of work when team members have to multitask (350). Team members also felt that two weeks spent between sprints was too long to reinforce new skills. The original time limit of two to three hours per day, and the pattern of two-week sprints interspersed with two weeks of non-project work, were intended to minimize interfering with the regular work of the department as much as possible. The time between sprints was especially helpful for the metadata librarian and metadata specialist, since it gave them plenty of time to balance sprint planning and metadata review with other responsibilities. Given the development team’s preferences, however, and the frequent advice against task switching, future projects might be planned to test different sprint schedules. Perhaps two-day sprints, repeated weekly would be better for projects that include a considerable amount of training and skill building. While the team members would still be alternating between project work and regular work, they would be doing so in a working rhythm that provided time needed to focus on each task, while still being mindful of departmental needs and competing responsibilities.

By the end of the project, the team completed metadata remediation for twenty-one collections, and more than doubled the number of collections that Penn State contributes to DPLA and PA Digital, our local DPLA Service Hub. This level of productivity might have been
possible without using agile and scrum principles and practices, but it is doubtful whether the team would feel that the pace of work was sustainable. By dividing the work into sprints, with breaks in between, team members were able to focus on the project during project hours, without worrying that their regular work was being neglected. In an environment where dedicating scrum teams exclusively to their project is impractical, the working rhythm of sprints can still offer benefits, especially when the unique needs of the team members are well balanced. Likewise, the other scrum events and artifacts can provide just enough structure to make planning and implementing a project more successful and easier to maintain than traditional project planning techniques, even if these events and artifacts are significantly adapted.

While the Penn State Libraries project used an extensively modified version of scrum—to the point that some might argue that it would not count as scrum—we have concluded that using scrum practices, even if modified, can greatly improve the process and the outcome of metadata projects, even those with unusual scrum teams. We would also argue that our project was, for the most part, agile. Our project benefited from adopting agile values, especially those that encourage making processes, tools, and documentation as lightweight as possible, in favor of focusing on the people involved and the ways they interact to produce quality work. We were also positively influenced by the agile principles, striving to make an ad hoc team of trainees work and feel like a team of “motivated individuals” who can “get the job done” with the right support.

Linda Ballinger (lmb49@psu.edu) is Metadata Strategist in the Penn State University Libraries

**Published:** September 2019

**References**


Acknowledgements: The author would like to thank Susan Lane, Metadata Implementation Specialist, and the members of the Digital Projects Metadata Team for all their work in making this project a success.