

December 15, 2020

Kristin,

Thank you for offering to consider this request for developer time for work on bulk APIs. I've tried to break this presentation up into parts for easier review.

What are bulk APIs. Bulk APIs are APIs that are intended to process large data sets, such as millions of records, both quickly and efficiently. They accomplish this by processing records in batches, often a thousand at a time.

Why do we need them. We need them for migrating data from legacy library systems into Folio, and to support the bulk data operations that librarians often perform on a daily basis in their production systems. It is hugely wasteful of people's time to try to do bulk operations without such APIs because you cannot submit a job, check the result, make a correction, and resubmit in any kind of natural workflow. You spend all of your time waiting for the job to complete. In the case of data migration, this often takes days.

The history. We have been trying in our Data Migration group to get bulk APIs developed for over two years. We have an epic and a series of jiras that provide the necessary specifications and rationale to proceed, but nothing has happened. The project has never been ranked high enough to warrant the allocation of developer time to complete it.

Why nothing has happened. Bulk APIs are dead in the water because the system in place for allocating developer resources only prioritizes the expressed individual needs of its institutional users, which are often focused on what is critical for them to go into production.

How are bulk APIs different? Bulk APIs are different because they are at least as much a common good as they are a value to individual institutions. They provide the Folio infrastructure with tools that make it more efficient in resource usage, more flexible in its programming solutions, and more performant in its operations. Whether migrating to Folio from older systems, or using it in day-to-day operations, they save time, and they improve the quality of the user experience by making Folio easier to work with. In terms of the goals of the Strategic Vision statement, based on our experience, they should noticeably increase the adopters' user satisfaction, and perhaps also the likelihood that eighty libraries will adopt it by 2022 by making the process of adopting Folio more rewarding.

What are the problems? Bulk API development does not fit well into the Folio development model, where work is done either centrally or within individual modules. Bulk APIs require both. They need RMB support to code the basic functionality, but also module developer support to apply that functionality to the data flows of their particular modules. There is a lot of inefficient communication overhead involved in getting developers to coordinate on a common implementation.

A possible solution. Bulk API development could be made more affordable by simplifying the development effort. The developer time to implement bulk APIs could be greatly reduced by having a single developer, working with the RMB team, design and implement a model set of bulk APIs for a single module, and then apply that model set to the remaining modules that need them. We have identified around twelve in all. This approach would eliminate all of the communication overhead and ramp up time that independent module developers would require to become familiar with the project, which would likely be much higher than the time actually needed to implement the code. It should also improve the quality and uniformity of the code. In addition, much of the RMB work has already been done, and some usable bulk API code has already been written, so that much of the development of the template code could leverage existing work. I'm also hopeful that its application to the other modules would be straight forward and could be done with minimal modifications.

We previously presented this project to the Product Council, who referred us to Capacity Planning, who voted it down, based on its lack of priority among implementing institutions. So, I thought I would try here to provide a wider context for why it is needed, in the hope that it may be useful in your thinking about it.

Anyone interested in the technical requirements for bulk APIs that support data migration can refer to our epic, <https://issues.folio.org/browse/UXPROD-2818?jql=labels%20%3D%20bulk-api>.

Thank you for considering this request, and please let me know if you have any questions.

Regards,

-dale

Dale Arntson
Date Migration Group Convener
Developer and Database Administrator
University of Chicago Library