

Making Your Tools Work for **YOU**

**BUILDING AND MAINTAINING
AN INTEGRATED TECHNICAL ECOSYSTEM
FOR DIGITAL ARCHIVES
AND LIBRARIES**

Max Eckard



**SOCIETY OF
American
Archivists**

CHICAGO

Society of American Archivists
www.archivists.org

© 2020 by the Society of American Archivists
All rights reserved.

Printed in the United States of America.

Library of Congress Control Number: 2020942499

Book design by Tyler Crumrine
Cover design by Kelly Sweeney

ISBN: 978-1-945246-44-9 (print)
eISBN: 978-1-945246-45-6 (epub)
eISBN: 978-1-945246-46-3 (pdf)

TABLE of CONTENTS

Preface | vii

Introduction | xi

- 1 Systems Integration for the Archival Enterprise: What It Is and What It Looks Like | 1
- 2 Designing Your Integrated Technical Ecosystem: Types and Tiers of Systems Integration | 20
- 3 How to Overcome Analysis Paralysis: Choosing Tools and Systems Designed to Integrate | 41
- 4 Now You're Speaking My Language: Common Metadata Standards and Systems Integration | 61
- 5 How to Get Things Done (and Integrate Systems) with APIs | 86
- 6 Systems Integration Methodologies for Other Use Cases | 119
- 7 Systems Integration "in the Wild" | 142
- 8 Now What? First Steps for Planning a Systems Integration Project | 180
- 9 Systems "Integration" for the Rest of Us | 200
- 10 Can We Talk about Your Messy Metadata? | 222
- 11 Supporting Data on the Move in Even "Loosely Coupled" Integrations: Python for Every Librarian and Archivist | 254
- 12 Making Your Tools Work for You | 285

Acknowledgments | 298

Glossary | 301

Bibliography | 313

Index | 326

About the Author | 337

SAA Preview

Preface

Before you delve into this book on systems integration for the archival enterprise, I want to make a couple of things clear from the beginning. First, I am an archivist, and before I was an archivist, I was a librarian. I have never been a developer, a systems administrator, or even, strictly speaking, a technologist in the broadest sense. As such, while this book on managing technical ecosystems in libraries and archives, with a special emphasis on systems integration, might be interesting to developers, systems administrators, or other archives and digital library technology service providers, my primary audience is people like me. In fact, if I'm *really* honest, one of the primary audiences for this book has just been *me*. I am very thankful to have had this opportunity to clarify my thinking around this topic.

That said, the “archival enterprise,” as I’ll define it in chapter 1, is “big” and “bold” and, of course, “sometimes difficult.” It draws on *many* areas of expertise, and this book is decidedly not *just* for archivists. In fact, I intend it to be useful for many other communities: developers, systems administrators, and other technology service providers, but also digital librarians, museum curators, those who work in digital preservation or curation, and others, especially students. These are all communities that can benefit from the general approach

I am espousing as well as the specific integration patterns and methodologies I describe.

Maybe you're a student—and if you are, a very special welcome to you!—or a new professional looking to learn a bit about the technology used in libraries and archives. Maybe you're a librarian or archivist who already uses a handful of tools or systems in your day-to-day work and you would like to know how to use them more efficiently. Maybe you're not technical at all, or you manage a library or archives and need the appropriate vocabulary to converse with your colleagues in information technology (IT); or perhaps, like a participant I had in a recent Society of American Archivists Digital Archives Specialist course I taught on tool integration, you're situated in IT, but looking to gain insight into the core practices and principles that you implement in conjunction with archivists, curators, or librarians. I have been or interacted with all of these people, and so whoever you are, I'm glad you are here.

Second, inherent difficulties exist in writing a book about technology. Already in the year or so it has taken to write this book, the technological landscape has shifted in significant ways. As this manuscript approaches publication, I can think of two examples:

- 1) The ArchivesSpace Technical Advisory Council (TAC) Integrations sub-team, which I both participated in and led, has shifted its thinking away from “tiers” of systems integration (as we had conceptualized them during my tenure on that sub-team and as they appear in chapter 2) to “methods” and “types” of systems integration. This is a good development, and it is incredibly gratifying when new people come along and take something you've helped create and make it better.
- 2) Stanford University Libraries' Institute of Museum and Library Studies (IMLS)–funded “Lighting the Way” grant, focused on the integration of systems for the discovery and delivery of archives and special collections (as opposed to more “backend” integrations, like many of those I will discuss in this book), is also an exciting development. The project, which I am participating in and advising on, centers around convening a series of national meetings in 2019–2020 to build consensus around strategic and technical directions for effective integration of systems supporting archival discovery and delivery, producing an integration handbook and project whitepaper that I very much look forward to reading and implementing. More than that,

it intends to do all of this as inclusively as possible and with an eye toward developing, as the proposal puts it, a “forward-looking agenda describing an *ethical, equitable, sustainable, and well-integrated* future for access and use of archives and special collections,” even beyond the technological considerations that are the primary focus of this book.

Likewise, any reference to specific tools, systems, or integration methodologies—and I certainly make a few of these in the book—is very likely to be out of date or at least incomplete in just a few short years. I knew this going in, which is why I have attempted throughout the book to focus on higher-level trends in technology—like systems integration—somewhat independently of their particular manifestations, for example, in a specific tool or system in any given archival workflow. That being said, I still found it hard to talk about technology or the theory behind it without using practical examples and talking about specific and particular tools and systems. While I’ve attempted to hold this tension in what I hope is a fairly sophisticated and ultimately helpful manner, as to the impact or longevity of this work, only time will tell. I am also well aware of the fact that, even if I get the technological discussion right, the technology component itself is relatively easy compared to other components of the archival enterprise—which also include organizational components like people and the allocation of resources—even if it does get a disproportionate amount of attention.

Finally, what I have attempted to write here is simply an introduction, an overview. Taking such a broad trajectory necessarily means that I haven’t been able to explore every aspect of systems integration. Instead, I will make my claim more modest: this book is by no means definitive or exhaustive and, knowing that I haven’t been able to anticipate every eventuality from the start, I think it’s best if we all treat this whole endeavor as somewhat provisional. I’d also like to acknowledge up front that I’ve only ever worked for large or very large academic institutions, with all of their undeniable privileges. It’s hard for me to think outside of that box, and yet I know that some of the underlying assumptions I make will reflect this environment and that at least some of what I propose in this book will not work for everyone.

In the end, my sincere hope is that this book serves as a catalyst for discussion, even pushback and criticism, and above all, that it contributes to a deeper, profession-wide exploration of an important topic that is not going away anytime soon.

SAA Preview

Introduction

These days, archivists, digital curators, and system developers make use of many different tools and systems in their day-to-day work. All of them specialize in one function or another, but there are no “silver bullets.” No single piece of software covers all aspects of the archival enterprise and managing all of these systems—especially the handoffs of data and metadata between them—can be overwhelming. The good news, at least when it comes to technology, is that many systems these days are designed to connect. Systems integration can “knit” disparate systems together to support a wide variety of archival needs, creating efficient workflows from accession to ingest to access.

In the next twelve chapters, I’ll give an overview of the what, why, and how of systems integration and its role in “knitting” disparate systems together to support a wide variety of workflows in the archival enterprise. If you’re starting your systems integration from scratch, I hope that this text will serve as a guide to beginning with the end in mind (i.e., to future proof it as much as possible); if you’re starting from an existing workflow that makes use of siloed systems, I hope that this text will help you get those systems to work together, better. Either way, I’ve learned from experience that the process of building and, in the longer term, maintaining, an integrated technical ecosystem (regardless of

the particular product, i.e., a particular integration between systems) becomes a lot easier when one follows the principles and practices outlined in this book.

Chapter 1 begins by contextualizing systems integration and introducing the topic, including an example from my own work at the Bentley Historical Library. In chapter 2, I cover various types and tiers of systems integration that are important to consider when designing an integrated technical ecosystem, as well as an absolutely fundamental and central concept to keep in mind when data and metadata are on the move: system(s) of record.

Chapter 3 takes a look at tool selection for systems integration and previews what exactly it is about certain systems that makes them capable of “playing nicely” with other systems. I then take deeper, more technical dives into the various ways that librarians and archivists can get systems to “talk” to one another, including common metadata standards in chapter 4, application programming interfaces (APIs) in chapter 5, and, finally, other methodologies for more varied use cases, including data interoperability protocols like the Simple Web service Offering Repository Deposit (SWORD), command line interfaces (CLIs), plug-in architectures, and image APIs like the International Image Interoperability Framework (IIIF) in chapter 6.

Chapter 7 might be one of my favorites. It takes a look at various examples of systems integration “in the wild” with case studies on diverse integrations (e.g., integrations designed to accomplish varied archival tasks, integrations done by institutions of varying sizes and types, integrations featuring both open-source and proprietary software, etc.). Each case study concludes with some of my own analysis and, to end the chapter, I highlight common themes that help to contextualize the case studies. These analyses and the conclusion, I hope, help to ensure that these discussions illustrate larger techniques or trends without being limited to the particular set of software applications used by the institutions where the archivists I interviewed work.

Chapter 8 discusses next steps for those who have decided to undertake a systems integration project, and chapter 9 explores “systems integration” for the rest of us. This chapter, and the two more “hands-on” chapters that follow (which include cleaning and reconciling metadata in chapter 10 and computer programming in chapter 11) are devoted to what steps you can take now to manage the “seams” or handoffs between systems that support your current workflows *on the way* to a more robust systems integration; they are about supporting data and metadata on the move, even if you’re not quite ready for a fully automated, tightly coupled systems integration.

Finally, in chapter 12, I conclude by examining some of the broader issues I perceive to be at the intersection of libraries, archives, and technology, and how a book on systems integration ultimately fits into that context. By the time you reach the conclusion, if nothing else, I hope that reading the text and reflecting on what you've read, perhaps aided by the discussion questions at the end of each chapter, will have increased your *critical imagination*—your ability to imagine the possibilities your technical ecosystem might afford—and that you feel empowered and equipped to begin experimenting with integrating your own systems to support your own archival workflows.

Scattered throughout the text you will also find a number of reflections on supplementary topics like computational thinking, the important “humans plus” role of humans in computer-assisted archival workflows, project management, version control, and impostor syndrome. I hope you find these enlightening as well.

MAX ECKARD