

# Designing a Migration Path - Environmental Scan

This page will be used to collect materials for an environmental scan of literature related to software upgrades and migrations as well as planned or recommended Fedora 3.x - Fedora 4.x upgrade projects.

## Sources

- [Reframing Open Source Repository Upgrades](#) by Erin Tripp (dataset - <https://osf.io/s3qx6/>)
- General migration advice "Open Source Repository Upgrades: Top Advice from Practitioners": 1) [Slides](#) 2) [Speaking Notes](#)
- [Digital Collections Survey Report](#) by Bridge2Hyku Project Team
- [Fedora and Digital Preservation Survey](#) by Fedora Leadership Group
- Jonathan Rochkind Blog post "On the present and future of samvera technical architectures" <https://bibwild.wordpress.com/2018/08/28/on-the-present-and-future-of-samvera-technical-architectures/>
- [Breaking Up With CONTENTdm: Why and How One Institution Took the Leap to Open Source](#) by Gilbert & Mobley
- [Hitting the Road towards a Greater Digital Destination: Evaluating and Testing DAMS at the University of Houston Libraries](#) by Wu, et al.
- [A Clean Sweep: The Tools and Processes of a Successful Metadata Migration](#), Anna Neatrou, Jeremy Myntti, Matt Brunsvik, Harish Maringanti, Brian McBride & Alan Witkowski
- [Objectivity Data Migration](#), Marcin Nowak, Krzysztof Nienartowicz, Andrea Valassi, Magnus Lubeck, Dirk Geppert
- [Outside The Box: Building a Digital Asset Management Ecosystem for Preservation and Access](#), Andrew Weidner, Sean Watkins, Bethany Scott, Drew Krewer, Anne Washington, and Matthew Richardson
- [Are we still working on this? A meta-retrospective of a digital repository migration in the form of a classic Greek Tragedy \(in extreme violation of Aristotelian Unity of Time\)](#), Steve Van Tuyl, Josh Gum, Margaret Mellinger, Gregorio Luis Ramirez, Brandon Straley, Ryan Wick, Hui Zhang
- [The Devil's Shoehorn: A case study of EAD to ArchivesSpace migration at a large university](#), Dave Mayo and Kate Bowers
- [The Semantics of Metadata: Avalon Media System and the Move to RDF](#), Juliet L. Hardesty and Jennifer B. Young
- [Massive Newspaper Migration — Moving 22 Million Records from CONTENTdm to Solphal](#), Alan Witkowski, Anna Neatrou, Jeremy Myntti and Brian McBride
- [Taking Control: Identifying Motivations for Migrating Library Digital Asset Management Systems](#), Ayla Stein, Santi Thompson
- [Deploying Islandora as a Digital Repository Platform: a Multifaceted Experience at the University of Denver Libraries](#), Shea-Tinn Yeh, Fernando Reyes, Jeff Rynhart, Philip Bain
- [A Doomsday Scenario: Exporting CONTENTdm Records to XTF](#), Andrew Bullen
- Berghaus, F., Blomer, J., Cancio Melia, G., Dallmeier Tiessen, S., Ganis, G., Shiers, J., Simko, T. (n.d.) CERN Services for Long Term Data Preservation. Retrieved from [https://cds.cern.ch/record/2195937/files/iPRES2016-CERN\\_July3.pdf](https://cds.cern.ch/record/2195937/files/iPRES2016-CERN_July3.pdf)
- [International Linked Data Survey for Implementers, 2018 Report](#), OCLC Research
- [From Silos to Opaquenamespace: Oregon Digital's Migration to Linked Open Data in Hydra](#), Julia Simic, Sarah Seymore
- [Understanding Metadata Needs When Migrating DAMS](#), Ayla Stein, Santi Thompson
- [Time, Money, and Effort: A Practical Approach to Digital Content Management](#), Christine Wiseman, Al Matthews
- [Who gives a DAM?: The Iterative Process for Assessing Digital Asset Management Tools](#), Bailey, Bondurant, Buckner, Creel, duPlessis, Huff, Mel goza, Mosbo, Muise, Potvin, Sewell, Wright
- [Spinning Communication to Get People Excited about Technological Change](#), Suzanna Conrad
- [Overly Honest Data Repository Development](#), Colleen Fallaw, Elise Dunham, Elizabeth Wickes, Dena Strong, Ayla Stein, Qian Zhang, Kyle Rimkus, Bill Ingram, and Heidi J. Imker
- [Developing Institutional Research Data Repository: A Case Study](#), Zhiwu Xie, Julie Speer, Yinlin Chen, Tingting Jiang, Collin Brittle, Paul Mather
- [Migrating to an Open Source Institutional Repository: Challenges and Lessons Learned](#), Devin Soper, Bryan Brown
- [Migrating an IR to New Technology: Opportunities, Challenges, and Decision-Making Processes](#), Simone Sacchi, Eva T. Cunningham

## Summary

Repository upgrades and migrations are quite common, and the literature covers several important aspects of this process: motivations for undertaking a migration, the difficulty of migrations, the possible benefits of a migration, and advice for those looking to undertake a migration in the future.

A common motivation for repository migrations is the cost of a commercially licensed product. [Gilbert and Mobley](#) were facing an increased cost to their CONTENTdm license due to reaching the item limit of their current tier, and [Stein and Thompson](#) cited license and maintenance fees as one of the main drivers of repository migrations based on survey data. Issues with the commercial platform itself, from performance and scale limitations ([Neatrou et al.](#), [Witkowski et al.](#)) to a lack of flexibility with regard to file and metadata formats ([Gilbert and Mobley](#), [Wu et al.](#)), were also key motivators. Finally, better support for digital preservation ([Stein and Thompson](#), [Berghaus et al.](#), [Fallaw et al.](#)) and linked data ([Wu et al.](#), [Stein and Thompson](#)) rounded out the top motivators in the literature.

There are many factors that make migrations difficult, but there is one primary problem category throughout the literature: metadata. [Van Tuyl et al.](#) cite metadata remediation as the biggest time sink during their migration project, and many others ([Bridge2Hyku Team](#), [Gilbert and Mobley](#), [Neatrou et al.](#)) present case studies that involve significant time spent on metadata normalization, de-duplication, and remediation. This speaks to a related difficulty often cited in the literature: inconsistent or "messy" source data. The process of mapping metadata from one repository system to another would be much simpler were it not for the fact that many legacy systems tend to have metadata quality problems in the form of custom local fields, duplicate fields, and misspelled entries.

There is a great deal of migration advice to be found in the literature, based primarily on lessons learned from migration projects. [Tripp](#) summarizes much of this advice into four categories: planning, metadata normalization, migration, and verification. Each of these categories is represented in the rest of the literature; [Nowak et al.](#) undertook a great deal of planning for their migration project, while [Simic and Seymore](#) invested a lot of time in large scale metadata normalization prior to migration. The migration phase itself was often accomplished with a combination of scripts and manual intervention, and the same is true of the verification step.

## Common Themes

- Motivations for migration
  - Commercial license costs
  - Lack of flexibility
  - Staff investment vs. licensing fees
  - Performance and scale issues
  - Better integration with other applications/services
  - Support for linked data
  - Support for digital preservation
- Migration difficulty
  - Custom metadata fields
  - Inconsistent data
  - Different data models
  - Metadata mapping: e.g. MODS XML to RDF
  - OSS documentation is not always complete/accurate
- Migration benefits
  - Metadata improvement/enrichment
  - Skills development
  - Streamlined workflows
  - Enhanced discovery via metadata enrichment
- Migration advice
  - Importance of communication
    - Engaging with stakeholders, collecting feedback, reporting on progress
  - Working with a representative sample
  - Requirements, scope
  - Normalize metadata before migration
    - But carefully scope this effort
  - Iterate, spot check
  - Agile methodologies
  - Contingency planning: staff turnover, learning curve, no single points of failure
  - Need for clear roles and responsibilities
- Repository requirements
  - Flexible object types and metadata
  - Batch ingest (e.g. from a spreadsheet)
  - Large community
  - Modularity
- Status of Fedora
  - Most still using Fedora 3
  - Plans to migrate but few timelines
  - Samvera/Fedora has major performance issues
  - Is the value of Fedora worth the complexity it introduces in the Samvera stack?
- Tools
  - Several examples of tools developed to aid/automate migration activities