The Craft of Database Curation: Taking Cues from Quiltmaking *

No Author Given

No Institute Given

Abstract. Data migration within library, archive and museum collections is a critical process to maintaining collection data and ensuring its availability for future users. This work is also an under supported component of digital curation. In this poster we present the findings from 20 semi-structured interviews with archivists, collection managers and curators who have recently completed a data migration. One of our main findings is the similarities between craft work and migration practices in memory institutions. To demonstrate these similarities, we use quiltmaking as as a framework. These similarities include the practice of piecing multiple systems together to complete a workflow, relying on community collaboration, and inter-generational labor. Our hope is that by highlighting the craftful qualities already embedded in this work we can show alternative best practices to data migration and database management. This is in an effort to get a broader understanding of what a successful data migration can look like.

Keywords: Craft · maintenance · database curation

1 The craft of long-term database maintenance

Database migration in library, archive and museum (LAM) collections is a critical, yet under-supported, part of digital curation. Though database systems are meant to provide a stable home for digital holdings and catalogs, in reality, they must be continually updated to ward off obsolescence, or to meet changing user needs. Inevitably, databases must be migrated from one system to another. This may happen rarely under an individual information professional's watch, but will happen many times over the life span of a collection, and often is such a time intensive process it seems never ending (reference anonymized for review).

Database obsolescence doesn't happen all at once, though, and migrations are rarely quick or easy; in our research to date, we've found that they can sometimes last years, particularly in under-resourced institutions. Between and during migrations, LAM database curators often find themselves needing to "satisfice" [19, 23] their data needs via skillful, craftful work-arounds to obsolescence. One participant referred to this process as creating a "patchwork quilt... of data

* Supported by organization x.

systems": an idiosyncratic, crafted-together approach to databasing that supplements the "official" but soon-to-be-obsolete data system with *ad hoc* support from other external systems.

In this poster we illuminate the craftful qualities of database work by extending our participant's metaphor: we use the craft of quiltmaking as an analytical lens to highlight craft processes in database curation and maintenance. We present common strategies curators, collective managers, and archivists use to weave, patch and piece their systems together, across long spans of time, and around roadblocks and obsolescence. This also includes relying on community collaboration and fostering a tradition of creating inter-generational data objects. Our hope is that by highlighting the craftful qualities already embedded in this work we demonstrate new and effective database maintenance and migration practices.

Barley and Orr describe technical work as existing somewhere between craft and science: between skillful manual labor and mental labor [2]. Others in the information sciences have written on the role of craft in human-computer interaction, data curation, and the data sciences [1, 5, 14, 16, 17, 20–22]. Some of the affordances ascribed to craft work in this literature are creativity, resistance to obsolescence, and openness to repair. In relation to the age of planned obsolescence [26] and fast technology, this is a very different notion of what our information systems can be.

This craftful perspective is largely missing from existing literature on data migration, which is typically written from a computer science and engineering lens [3, 4, 10, 12, 24]. This research tends to prescribe best practices in line with normative views of database practices. These normative approaches tend to assume a well-resourced computing environment in which databases can be quickly updated or customized to changing needs, and in which databases are used primarily for information search and retrieval rather than the on-going curation of a collection (digital or otherwise). By centering craft in our work, we surface database practices that have otherwise been under-studied. We also give alternative approaches to data work via a more feminist lens [8] by showing how non-normative data practices can be more effective than "best" practices in some contexts.

2 Methods

We are developing case studies of database migration at a range of LAM institutions. In this poster we present findings from two of these cases: a large university-based natural history museum seeking to migrate all it's subcollections' databases to a unified system; and an archival database software user community, working to adopt a new database after theirs became obsolete. We conducted 15 interviews with collection managers and curators in the natural history museum and 6 with archivists from the software user collaborative. All were actively completing a migration, or had recently completed a migration. The interviews were coded in NVivo 12, a qualitative coding software.

In interviews, we asked about what factors drove the decision to migration, the experience and challenges faced during the migration process (and in the inevitable period of limbo after), and what they would do different in future data migrations. This set of questions gave space for our participants to share practices they currently utilize, and give advice for future migrations as well. We used inductive coding methods to identify themes in the data [15]. In both cases, participants described "quilting," "patching," or "sewing" systems together to cope with a protracted migration process. In our final round of coding, we explicitly adopted quiltmaking as an analytical lens, and identified the following types of quilt-like database maintenance practices.

2.1 Quiltmaking as an analytical lens

We chose quiltmaking as a lens both because our participants' used this metaphor for this work repeatedly, and because of the known parallels between this craft and computational work. Quiltmaking requires similar skills as database work, such as breaking a large project into manageable tasks, following complicated step by step instructions, and completing repetitive tasks. Computational thinking curriculum calls these skills decomposition, algorithmic thinking, abstraction, and pattern recognition [18, 25].

Like databases, quilts are functional objects, but ones that can vary in complexity and design: they were originally created out of necessity, but have since evolved into pieces of art and textiles that can depict scenes and patterns, or create optical illusions. Creating these visualizations require specific measurements and stitching that follows complex patterns and instructions. Common quilt patterns such as the pinwheel, the eight point star, and the double wedding ring patterns create optical illusions, but require strict adherence to a pattern of colors and block design to successfully complete the desired visual. *Figure 1* is an example quilt pattern and one can see the abstraction required to follow this and produce a finished product.

Databases are often similarly constructed following a pattern: for instance, an Entity Relationship Diagram [6], illustrates the relationships between different classes of data. *Figure 2* is an example of an entity relationship diagram published by Specify, a natural history museum databasing system adopted by one of our cases. Just as the quiltmaker abstracts the pattern into a finished product, a database designer abstracts an ERD into machine readable code, which is then read and interpreted by relational database software.

Quiltmaking and database construction obviously rely on different skills and media, but we provide these examples to highlight how they nevertheless do have some skills and practices in common, including patternwork, a reliance on abstraction, and skillful design. Below we further explicate the parallels between data and craft work via quiltmaking.

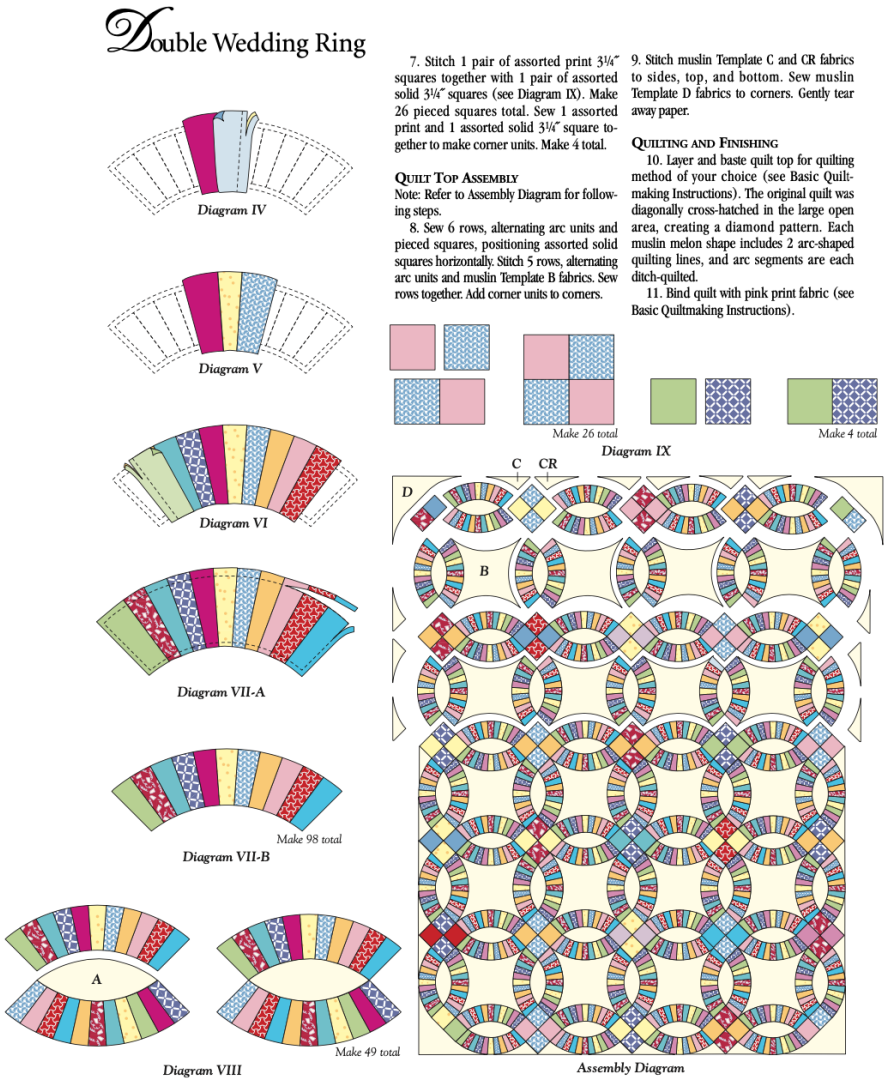

Fig. 1. Quilt pattern for a Double Wedding Ring Quilt [9]

Collector				
In Specify, the person or persons who collected a specimen. Collector records are stored in the Agent table.				
Field	**Type**	**Length**	**Description**	**Flags**
IsPrimary	java.lang.Boolean		Is Primary	Required
OrderNumber	java.lang.Integer		The position in the list of collectors.	Required
Remarks	text	4096	Additional comments.	
Text1	text	65535		
Text2	text	65535	Text2	
TimestampCreated	java.sql.Timestamp		The timestamp the record was created.	Required
TimestampModified	java.sql.Timestamp		The timestamp the record was last modified.	
Version	java.lang.Integer		Version	
YesNo1	java.lang.Boolean		Yes No1	
YesNo2	java.lang.Boolean		Yes No2	
Relationships				
Type	**Target Table**	**Field name**	**Description**	
many-to-one	Agent	agent	Agent	
many-to-one	Collecting Event	collectingEvent	Collecting Event	
many-to-one	Agent	createdByAgent	Created By Agent	
many-to-one	Division	division	Division	
many-to-one	Agent	modifiedByAgent	Modified By Agent	

Fig. 2. Example of schema documentation provided by the database Specify. This is the documentation for the table Collector. [7]

3 Craft in curation

3.1 Piecing things together

One of the primary actions in quilting is the attachment of multiple pieces of fabric into a single textile. This is done by stitching one piece of material to another – both to create a final piece with multiple layers, and to repair the quilt over time.

While many LAMs wish for a single unified database to meet all their needs, we have found that in practice, this "single" system must be woven from a multiplicity of data "surfaces." Despite the wide range of LAM-specific database systems (both proprietary and open access), even robust database platforms require augmentation through other information systems. 11 out of 20 participants described using multiple databases, collection management systems, platforms, and even spreadsheets to complete their workflows. This practice can seem precarious, but we found that it actually *adds* to the functionality of the collection overall. Like a quilt, the whole of these pieced together systems is stronger than the sum of its parts.

That said, when it comes time to migrate these complex systems, it isn't as straightforward as moving from one system to another. Rather, database curators must figure out how to move all of these pieced together platforms forward, combining some, but also continuing to use legacy systems as well. Often they must reconstruct past database curators' quilting methods to move the system forward.

One interview participant who manages a biological collection noted this in describing the multiple types of data they manage, and how each is in various stages of being added into the centralized database. They stated, "We have several other ancillary collections. We have photographs of specimens and localities. Those are not formally incorporated into the catalogue...We also have some older field notes that are that are physically there, but may or may not be digitized... We have a variable number of quantitative environmental data, like water quality...they're more easily incorporated into the database immediately". This complex work of piecing together multiple data systems is a critical component of data migration.

3.2 The databasing bee

Sewing a quilt by hand is a time and labor intensive process, and therefore was historically often done by multiple (female) members of a family [11], or in parties or "quilting bees" to finish a single quilt faster than one would be finished alone [13]. Similarly, we find that database curators often build their own communities to bring collective knowledge to their databasing projects – a databasing bee, so to speak. They draw on each other's knowledge to avoid common pitfalls, and to build interoperable systems that follow community norms and best practices.

Most of the archivists we spoke with were employed in small archives where there was only 1 or 2 full time staff members, so there was a strong reliance on

their professional community outside of their institution to support their work. All of the archivists we spoke with belonged to a user collaborative for the software Archon. Archon had announced its sunset in 2014, but many archives are still utilizing the software. This user collaborative, which had started when the software was still regularly updated, had changed its role over time to provide community support towards eventual data migration. Members of this user collaborative had already funded an update to the softwares open source code, but were also in continual contact discussing potential archival systems that members were considering migration to.

At the natural history museum, database curators were unified by their employment by the same institution – but they also worked with community members outside of their university to draw on others' "lessons learned." This led to some tension initially; the university administration wished to unify the 45 different databases under one umbrella, despite disciplinary differences and norms. Conversations went on for at least three years between collection managers, museum directors, library staff, and the university IT department trying to decide on a singular collection management system that would work for everyone. Eventually the museums split into using two systems: the biological collections went into the system Specify and the cultural collections into Collective Access. While the museums did not wind up using the same system in the end, the community they built around databasing during this attempted migration was critical to eventually figuring out solutions that fit everyone.

3.3 Creating inter-generational data objects

Both quilts and databases are intentionally multi- or inter-generational objects. The structure of quilts is such that they can last for decades – sometimes even centuries if patched and cared for properly. Quilts often transition from being functional objects to family heirlooms.

In a similar manner, databases are passed down from one generation of curators to another. Many of our participants had inherited their collections databases from a predecessor – and are working to create a digital object that will be robust enough for future generations. Often, curators did not know who originally created the databases they worked with. They consequently had to reconstruct their history, often to understand how mysterious remnants of database-structures past came to be in their current systems.

This inter-generational quality can sometimes make migration more complicated. One participant described having to modify the metadata schema after a migration because the migration added many redundant fields: "It was a big process...deciding what fields in there were redundant. A lot of it were vestiges from the old database when they weren't doing things properly. And I tried to clean it up." Another described having to normalize text: "Everything was in uppercase. There was no lower case. When we converted that over we had to figure out how to do that. I don't remember how that was done, but we eventually got it out of all uppercase".

We highlight this theme because these remnants from past systems were often spoken of as a data entry problem that needed to be corrected, or as an obstacle after the migration that needed to be addressed. We compare this with the inter-generational qualities of a quilt, where patches and evidence of past use are signs of character and seen as a more positive aspect of the quilt. We note that these remnants of legacy data are the reality in most migrations, and that it isn't always feasible to erase all of them. In thinking of this work like a patchwork quilt, we believe its important to note that it can still be a successful migration if these evidences of past use remain.

4 Conclusion and future work

By using quiltmaking as an analytical lens, we're able to highlight craft practices and narratives embedded within database work, which are often overlooked. Craft, re-use, resourcefulness, and DIY solutions all can have positive implications on migration work. Secondly, by demonstrating how this work is actually being completed, even if it doesn't align with best practices on what a data migration should look like [3, 12, 24], we are bringing in a more complete understanding of this work. A main goal of our project is to highlight best practices in data migration work, and this poster shows a range of practices that could be useful in justifying workflows for other memory institutions contemplating a data migration.

Further, in spring 2022 we will be hosting a virtual workshop to share our findings of this research, brainstorm next steps, and most importantly begin to build a critical community of those interested in continued research in this space of research data maintenance and migration. Our work has shown that information professionals working on data migration desire community support in this work, and the workshop will be an important first step in building awareness of the prevalence of these activities. Our hope is that by sharing this poster with the iConference community we will be able to recruit interested parties in participating in this workshop which will take place shortly after the conference.

References

1. Bardzell, S., Rosner, D.K., Bardzell, J.: Crafting quality in design: integrity, creativity, and public sensibility. In: Proceedings of the Designing Interactive Systems Conference. pp. 11–20. DIS '12, Association for Computing Machinery, Newcastle Upon Tyne, United Kingdom (Jun 2012). https://doi.org/10.1145/2317956.2317959, https://doi.org/10.1145/2317956.2317959
2. Barley, S.R., Orr, J.E. (eds.): Between craft and science: technical work in U.S. settings. Collection on technology and work, IRL Press, Ithaca, N.Y (1997)
3. Brodie, M.L., Stonebraker, M.: Migrating legacy systems: gateways, interfaces & the incremental approach. The Morgan Kaufmann series in data management systems, Morgan Kaufmann Publishers ; IT/Information Technology, San Francisco, Calif. : [S.l.] (1995)

4. Buneman, P., Cheney, J., Tan, W.C., Vansummeren, S.: Curated databases. In: Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '08. p. 1. ACM Press, Vancouver, Canada (2008). https://doi.org/10.1145/1376916.1376918, http://portal.acm.org/citation.cfm?doid=1376916.1376918
5. Cheatle, A., Jackson, S.J.: Digital Entanglements: Craft, Computation and Collaboration in Fine Art Furniture Production. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. pp. 958–968. ACM, Vancouver BC Canada (Feb 2015). https://doi.org/10.1145/2675133.2675291, https://dl.acm.org/doi/10.1145/2675133.2675291
6. Chen, P.P.S.: The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems **1**(1), 9–36 (Mar 1976). https://doi.org/10.1145/320434.320440, https://dl.acm.org/doi/10.1145/320434.320440
7. Consortium, S.C.: Specify Schema (2021), https://www.specifysoftware.org/specify-6-schema/
8. D'Ignazio, C., Klein, L.F.: Data feminism. Strong ideas series, The MIT Press, Cambridge, Massachusetts (2020)
9. Ferguson, N.: Vintage Double Wedding Ring. McCall's Quilting - Vintage Quilts pp. 8–10 (1999), https://s31968.pcdn.co/wp-content/uploads/2017/10/FREEDoubleWeddingRing$_{Q}uiltPattern.pdf$
10. Henry, S., Hoon, S., Hwang, M., Lee, D., DeVore, M.: Engineering trade study: extract, transform, load tools for data migration. 2005 IEEE Design Symposium, Systems and Information Engineering pp. 1–8 (2005)
11. Higgs, E., Radosh, P.F.: Quilts: Moral Economies and Matrilineages. Journal of Family History **38**(1), 55–77 (Jan 2013). https://doi.org/10.1177/0363199012470063, http://journals.sagepub.com/doi/10.1177/0363199012470063
12. Maatuk, A., Ali, A., Rossiter, N.: Relational Database Migration: A Perspective. In: Bhowmick, S.S., Küng, J., Wagner, R. (eds.) Database and Expert Systems Applications, vol. 5181, pp. 676–683. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85654-2$_5$8, $http$: $//link.springer.com/10.1007/978-3-540-85654-2_5 8, iSSN$: $0302-9743, 1611-3349 SeriesTitle$: $LectureNotesinComputerScience$
13. MacDowell, M.: Quilting Parties. In: Encyclopedia of Recreation and Leisure in America, vol. 2, pp. 166–167 (2004)
14. Mentis, H.M., Rahim, A., Theodore, P.: Crafting the Image in Surgical Telemedicine. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. pp. 744–755. ACM, San Francisco California USA (Feb 2016). https://doi.org/10.1145/2818048.2819978, https://dl.acm.org/doi/10.1145/2818048.2819978
15. Miles, M.B., Huberman, A.M., Saldaña, J.: Qualitative data analysis: a methods sourcebook. SAGE, Los Angeles, fourth edition edn. (2020)
16. Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q.V., Dugan, C., Erickson, T.: How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–15. ACM, Glasgow Scotland Uk (May 2019). https://doi.org/10.1145/3290605.3300356, https://dl.acm.org/doi/10.1145/3290605.3300356
17. Owens, T.: The theory and craft of digital preservation. Johns Hopkins University Press, Baltimore (2018)

18. Papert, S.: Mindstorms: children, computers, and powerful ideas. Basic Books, New York, 2nd ed edn. (1993)
19. Prabha, C., Silipigni Connaway, L., Olszewski, L., Jenkins, L.R.: What is enough? Satisficing information needs. Journal of Documentation **63**(1), 74–89 (Jan 2007). https://doi.org/10.1108/00220410710723894, https://www.emerald.com/insight/content/doi/10.1108/00220410710723894/full/html
20. Rosner, D.: Critical fabulations: reworking the methods and margins of design (2018), http://ieeexplore.ieee.org/servlet/opac?bknumber=8555178, oCLC: 1045208731
21. Rosner, D.K., Fox, S.E.: Legacies of craft and the centrality of failure in a mother-operated hackerspace. New Media & Society **18**(4), 558–580 (Apr 2016). https://doi.org/10.1177/1461444816629468, http://journals.sagepub.com/doi/10.1177/1461444816629468
22. Rosner, D.K., Ryokai, K.: Reflections on craft: probing the creative process of everyday knitters. In: Proceeding of the seventh ACM conference on Creativity and cognition - C&C '09. p. 195. ACM Press, Berkeley, California, USA (2009). https://doi.org/10.1145/1640233.1640264, http://portal.acm.org/citation.cfm?doid=1640233.1640264
23. Simon, H.A.: Theories of Bounded Rationality. In: Decision and organization. North-Holland Pub. Co., Amsterdam (1972)
24. Vassiliadis, P.: A Survey of Extract–Transform–Load Technology:. International Journal of Data Warehousing and Mining **5**(3), 1–27 (Jul 2009). https://doi.org/10.4018/jdwm.2009070101, http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jdwm.2009070101
25. Wing, J.M.: Computational thinking. Communications of the ACM **49**(3), 33–35 (Mar 2006). https://doi.org/10.1145/1118178.1118215, https://dl.acm.org/doi/10.1145/1118178.1118215
26. Zallio, M., Berry, D.: Design and Planned Obsolescence. Theories and Approaches for Designing Enabling Technologies. The Design Journal **20**(sup1), S3749–S3761 (Jul 2017). https://doi.org/10.1080/14606925.2017.1352879, https://www.tandfonline.com/doi/full/10.1080/14606925.2017.1352879