Software Sustainability Challenge: ECOLM and Lute Tablature

Chris Cannam
Particular Programs Ltd
London, UK
chris.cannam@particularprograms.co.uk

David Lewis
Goldsmiths, University of London
London, UK
d.lewis@gold.ac.uk

Tim Crawford
Goldsmiths, University of London
London, UK
t.crawford@gold.ac.uk

ACM Reference Format:
Chris Cannam, David Lewis, and Tim Crawford. 2023. Software Sustainability Challenge: ECOLM and Lute Tablature. In *Proceedings of 10th International Conference on Digital Libraries for Musicology (DLfM'23).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 NATURE AND PURPOSE OF THE SUSTAINABLE TOOL OR RESOURCE

ECOLM (1999-2002) was a project led by Tim Crawford which developed and populated a database of lute tablature encodings with metadata, for scholarly use, queried using a web interface. Subsequent projects **ECOLM II** (2002-2006) and **ECOLM III** (2012) expanded the database and used it for some computational musicological investigations. The resulting database was hosted on a public-facing web server at Goldsmiths, University of London.[1] It is still running today, although nobody is formally responsible for maintaining it.

ECOLM is a relatively small scholarly resource with around 2,000 carefully-curated encodings. A number of other public lute tablature resources exist, of varying size, quality, and consistency (see section 2). These typically face challenges to sustainability similar to those of ECOLM (see section 3). It seems useful to consider those challenges for such resources in general, at least insofar as they address the same audience.

[add more to justify why these resources, especially ECOLM, are worth spending time on]

2 AUDIENCE AND USERS, DISCIPLINES AND SUBJECTS

We are considering here the sustainability of at least five lute tablature datasets: ECOLM itself; **lutemusic.org**[2] by Sarge Gerbode; **mss.slweiss.de**[3] curated by Peter Steur and the late Markus Lutz; a set of scans from **Lute Society publications** curated by John Robinson; and a set of transcriptions of **editions by Pierre Phalèse** curated by Jan Burgers. All of these are either Creative Commons licensed or have been offered by their maintainers as possible constituents of a future combined tablature resource.

These resources serve a spectrum of audiences. At one end, lutemusic.org aims at performers and includes edited transcriptions with relatively little scholarly metadata or editorial comment. At the other, ECOLM was aimed at computational musicologists and prioritises diplomatic facsimiles and transcriptions that preserve original scribal idiosyncracies.

In this review we are particularly interested in sustainability for musicology and other academic purposes. To this end, we asked three exemplary users of online early-music resources—a musicologist, a computational musicologist, and a lute teacher—for their views about them in order to understand scholarly expectations.

Briefly, they agreed on the importance of trust and provenance, particularly in knowing about the quality of transcription and level of editorial intervention in a resource. There was some consensus about the value of simple search with subsequent refinement, of a cleanly-designed results layout including inline incipits, and of API and data provision. Resources regarded as worth studying included DIAMM[4] (Digital Image Archive of Medieval Music), the Vihuela Database,[5] the Josquin Research Project,[6] and RISM[7] (Répertoire International des Sources Musicales) which is a near-ubiquitous entry point for musicological queries.

3 CHALLENGES FOR SUSTAINABILITY

A number of highly typical challenges apply to subsets of the datasets we are considering: curation and maintenance by individuals or small groups of enthusiasts, in some cases of retirement age; maintenance in limited periods of spare time, perhaps following initial short-term funding; data management using ad-hoc methods or private systems that are not accessible to third-party reproduction; lack of data export facilities or support for common interchange formats.

[expand this]

4 FUTURE DIRECTIONS

We have identified three alternative directions for sustainable development, as follows.

4.1 "Enhanced ECOLM"

In this approach, we retain the relational data schema of the existing ECOLM, which is the only one of the datasets under consideration to have a formal schema, and provide ETL (extract, transform, load) data loaders for other sources. We then publish the schema, data dumps, and automation to rebuild or mirror the data, along with the code of our query interface and encourage others to attach their own interfaces or tools to it.

Advantages of this approach include the ability to preserve existing code and to use original ECOLM records as a reference. The existing schema is detailed and fairly effective, provides appropriate structure, and reflects some good domain-specific decisions. Relational data import is a well understood field, and we could focus on user interfaces and data conversion rather than any novelty of data

[1] http://doc.gold.ac.uk/isms/ecolm/database/
[2] https://lutemusic.org/
[3] https://mss.slweiss.de/
[4] https://www.diamm.ac.uk/
[5] https://vihuelagriffiths.com/
[6] https://josquin.stanford.edu/
[7] https://rism.info/

DLfM'23, November 10, 2023, Milan, Italy
2023. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

representation. If the work fails, the result should be at minimum a more open publication of the existing ECOLM.

Disadvantages include that the schema has little in common with any of the ad-hoc solutions other maintainers have settled on, so all import and export would be custom. It also has little in common with wider current practice. The schema is perhaps already overspecified for its current use, yet does not address any problems relating to stable identifiers, versioning, or providing queryable APIs or data sources.

Although we could at least initially reuse the existing user interface, it is no longer considered a strength of ECOLM and would need some work to update to modern expectations.

4.2 Graph-based

In this approach, we take the fundamental representation to be a graph of triples in the model of RDF, and convert all metadata to that for import and from it for query. External data such as transcriptions and multimedia resources are identified by graph-relatable identifiers such as URIs.

Advantages include the use of a widely-understood and accepted model that meets common expectations about data compatibility and API provision. For schema we can draw ontologies from a number of existing systems including the widely-used RISM. The structure is reasonably amenable to versioning and to use of "idempotent" import flows with automated testing, offering the option of ongoing import of changes in upstream sources. In principle existing tools may be used for review, query, inferencing, and format conversion. The use of standard formats with automated tests could lead to a result far more easily maintained or mirrored by third parties.

The approach has difficulties as well. It discards the existing user interface work and requires even the existing ECOLM data to be converted. Although graph representations have wide application, they are not generally used for manual data management and therefore have as little in common with the ad-hoc schemas of enthusiast lute resources as with that of ECOLM. Significant work would be required to maintain stable identifier mappings from external sources. With a more flexible structure than ECOLM's relational database, care and good automated testing would be needed to avoid "silently missing data" problems on query. Finally a separate solution would be needed to the problem of identifying and retrieving non-graph data such as media resources.

Although in this approach we could no longer use the existing ECOLM user interface, that may be slightly mitigated by the ability to adapt other graph-driven UIs to the model.

4.3 "RISM-aligned"

In this approach, we concentrate on compatibility with formats and software used by RISM. The motivating principle is that RISM is the dominant "entry point" in this field and, although our sources are not all of the standard indexed by RISM, we want to facilitate linkage (now or in future) for those that are, and to be prepared if in the future RISM should grow to cover the whole area of directly represented lute tablature. The ideal future would involve replacing this project entirely with an aspect of RISM.

In this approach the metadata representation might be MARC[8] (Machine Readable Cataloging) as it is within RISM, and possibly the RISM Muscat[9] application might be used for management.

[8] https://www.loc.gov/marc/
[9] https://rism.info/community/muscat.html