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, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 NATURE AND PURPOSE OF THE SUSTAINABLE TOOL OR RESOURCE

ECOLM (1999-2002) was a project run by Tim Crawford which developed and populated a queryable database of lute tablature encodings with metadata 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 variable size, quality, and consistency (see section 2). These typically face challenges to sustainability similar to those of ECOLM (see section 3). It seems reasonable to try to address those challenges for such resources in general, at least insofar as they are useful to the same audience. Some of the maintainers of other resources have offered to contribute their data to such an effort.

2 AUDIENCE AND USERS, DISCIPLINES AND SUBJECTS

3 CHALLENGES FOR SUSTAINABILITY

4 FUTURE DIRECTIONS

We have identified three alternative directions for sustainable development.

4.1 "Enhanced ECOLM"

This approach retains the relational data schema of the existing ECOLM, which is detailed and fairly effective, although it has little in common with the other systems we have considered or with wider current practice. We then provide ETL-type (extract, transform, load) data loaders for other sources of interest.

Advantages of this approach include the ability to preserve existing code and to use original ECOLM records as a reference. The existing schema 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 representation.

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. 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 the system 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. 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 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"

Fundamental metadata representation is MARC, and we use Muscat to maintain it?

[1] http://doc.gold.ac.uk/isms/ecolm/database/

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