Software Sustainability Challenge: ECOLM and Lute Tablature

Goldsmiths
UNIVERSITY OF LONDON

Chris Cannam Particular Programs Ltd
David Lewis, Tim Crawford Goldsmiths, University of London

Read more here!

What is ECOLM?

"Electronic Corpus of Lute Music", a series of projects (1999-2006 and 2012) to make a database of lute tablature encodings with metadata, for scholarly use, queried using a web interface. The resulting database was hosted at Goldsmiths, University of London, and is still running today, although nobody is formally responsible for maintaining it.

The ECOLM interface as it appears today.

ECOLM is a relatively small scholarly resource with around 2,000 encodings. We are also interested in the future of a whole field of other tablature resources of varying size and editorial standards. These include **lutemusic.org** by Sarge Gerbode, **mss.slweiss.de** by Peter Steur and the late Markus Lutz, **Lute Society publications** curated by John Robinson, and a set of **editions by Pierre Phalèse** curated by Jan Burgers. In general they are CC-licensed or have maintainers open to collaboration.

Audiences

These resources serve a spectrum of audiences from performer to musicologist. We are most concerned with sustainability for musicology and other academic purposes. We asked three exemplary users of online early-music resources—a **musicologist**, a **computational musicologist**, and a **lute teacher**—for their views about them so as to understand scholarly expectations.

They highlighted the importance of **trust and provenance**, particularly 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 (Digital Image Archive of Medieval Music), the Vihuela Database, the Josquin Research Project, and RISM (Répertoire International des Sources Musicales) which is a near-ubiquitous entry point for musicological queries.

The Challenge

The problems these resources face are highly typical: curation and maintenance by **individuals or small groups of enthusiasts**, in some cases of retirement age; maintenance in **limited periods of spare time**; 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**.

Lute tablature faces further difficulties because it is such a specialised field. Wider digital solutions for music distribution and study typically do not handle it and its resources face significant risk of becoming inaccessible.
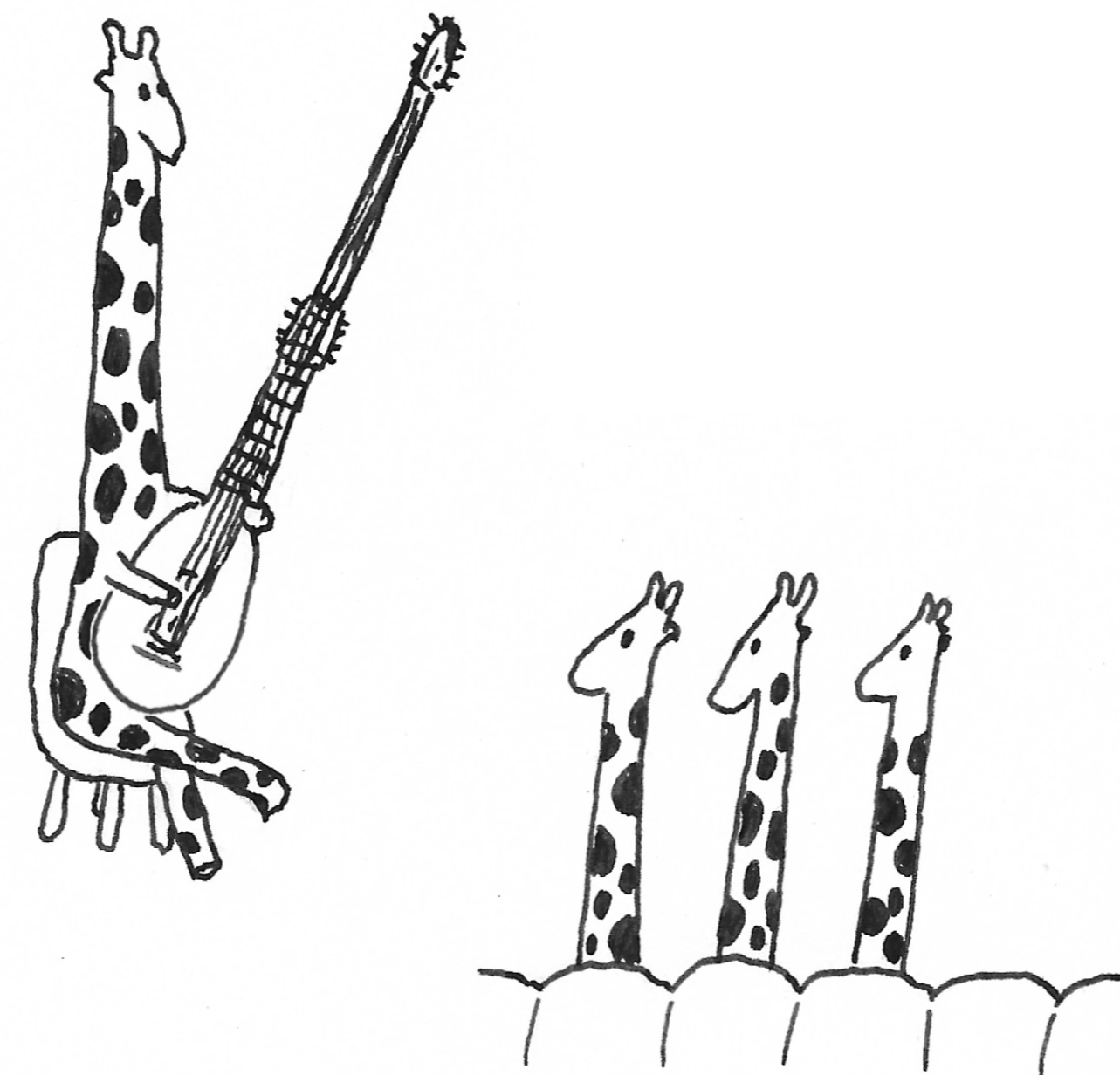
A giraffe playing an archlute.

A Sustainable Outcome

Requirements for a socially-sustainable outcome:
Sufficient immediate utility to justify the work and to ensure that users are invested in the outcome;

Sufficient scholarly quality to ensure that the outcome is genuinely helpful in expanding and improving the work that users are able to do;

Accessible and robust technical choices to enable other enthusiast users to maintain and contribute to the outcome without substantial difficulty or risk of causing damage.

Common Patterns

The three technical directions we have identified are common patterns for work of this kind:

1. Choose one of the existing technical solutions and adapt the others to it;

2. Adopt a higher-level ("linked data") approach in which existing resources are promoted to a common format;

3. Find an industry partner, adopt their tools, and contribute to their existing ecosystem.

Possible Future Directions

1 "Enhanced ECOLM"

Retain the relational data schema of ECOLM, the only one of the datasets to have a formal schema. Provide data loaders for other sources. Publish schema, data dumps, automation to rebuild or mirror the data, and the code of a query interface. Encourage others to add their own tools.

Advantages: Ability to preserve existing code and use ECOLM as a reference. Schema has some good design decisions. Relational data import is well understood. Can focus on UI and data conversion. Relatively low-risk.

Disadvantages: Schema has little in common with the other ad-hoc solutions, so all import and export would be custom. Also little in common with wider current practice. Does not solve problems relating to stable identifiers, versioning, or providing queryable APIs or data sources.

2 Graph-based

Take the fundamental representation to be a graph of triples like RDF. Convert all metadata to that for import, and from it for query. Identify external data (transcriptions, media) by graph IDs such as URIs.

Advantages: Widely-understood model that meets common expectations about compatibility and APIs. Can draw ontologies from existing systems. Reasonably amenable to versioning and to use of "idempotent" import flows with automated testing, for ongoing import of upstream changes. Could be more easily maintained or mirrored by third parties.

Disadvantages: Requires all existing data to be converted and discards existing UI. Graphs not generally used for manual data management, so little in common with existing ad-hoc schemas of enthusiasts. Stable identifier mappings still hard. Good testing required to avoid "silently missing data" on query. Does not address non-graph data such as media.

3 "RISM-aligned"

Focus on compatibility with formats and software used by RISM as the dominant "entry point" in this field. Though our sources are not all up to the standard indexed by RISM, we want to facilitate linkage 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 here would involve replacing this project entirely with an aspect of RISM.

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