Using Pivots To Explore Heterogeneous Collections: A Case Study in Musicology

Daniel A Smith, David Bretherton, mc schraefel, Richard Polfreman, Mark Everist, Jeanice Brooks, Joe Lambert

University of Southampton, Southampton, UK SO17 1BJ

Introduction

In order to provide a better e-research environment for musicologists, the musicSpace project has partnered with musicology's leading data publishers, aggregated and enriched their data, and developed a richly featured exploratory search interface to access the combined dataset. There have been several significant challenges to developing this service, and intensive collaboration between musicologists (the domain experts) and computer scientists (who developed the enabling technologies) was required. One challenge was the actual aggregation of the data itself, as this was supplied adhering to a wide variety of different schemas and vocabularies. Although the domain experts expended much time and effort in analysing commonalities in the data, as data sources of increasing complexity were added earlier decisions regarding the design of the aggregated schema, particularly decisions made with reference to simpler data sources, were often revisited to take account of unanticipated metadata types.

Additionally, in many domains a single source may be considered to be definitive for certain types of information. In musicology, this is essentially the case with the "works lists" of composers' musical compositions given in Grove Music Online (`http://www.oxfordmusiconline.com/public/book/omo_gmo`), and so for musicSpace, we have mapped all sources to the works lists from Grove for the purposes of exploration, specifically to exploit the accuracy of its metadata in respect to dates of publication, catalogue numbers, and so on. Therefore, rather than mapping all fields from Grove to a central model, it would be far quicker (in terms of development time) to create a system to "pull-in" data from other sources that are mapped directly to the Grove works lists.

Pivoting Across Data Sources

To this end, we have developed the concept of "pivots" across data sources, whereby a single mapping is defined between sources so that users can browse facets from either source at once, with the browser using the semantic mapping to enable on-demand cross-referencing across sources.

This is similar to the notion of the "visual pivot" explored by Robertson et al (2002), who described a system in which hierarchies are extracted from different databases, and pivot points are designated where point hierarchies intersected at common instances. For example, consider an organisation that holds a number of management databases, where a person can be contained in two different databases, a pivot point exists for each person that is in both databases. A three-dimensional visualisation is available to visually animate the transition from one hierarchy to another around the pivot point, both to show the user how the different hierarchies relate, and to enable the user to move from one related table to another to explore relationships across the data. These intersecting hierarchies are known as "Polyarchies," and have been formally compared to the representation model of facet as used in mSpace and the structure used in ZigZag (McGuffin and schraefel, 2004), where a populated taxonomy was presented to enable comparison of the differences in representation. By using logically associated intersecting pivots in data, it is possible to query a dataset's domain alone, and then pull in additional facets from other related sets through a semantic link.

In order for a user to browse from one faceted data set to another using pivots, a pivot-aware browser must be used. In our prototype we have used a modified version of the mSpace faceted browsing framework in order to demonstrate possible interaction methods for enabling pivoting. These interactions enable the mSpace to link automatically to other interfaces that contain pivots. These links to other interfaces are generated by a web service, and not by an individual interface. This means that they can be updated and changed dynamically without making changes to the interface. This allows links to new interfaces to be displayed in all relevant existing interfaces, and also the bi-directionality of links are preserved, therefore it will always be possible to go back to the source interface, allowing users to retrace their interaction choices. The ability to retrace their choices is important, as the recoverability of context has been shown to be beneficial to cognitive understanding during exploration (Wilson et al, 2009).

Musicological Case Study

In the musicSpace project, we mapped each source to a common schema, whose design was informed by the Music Ontology (Raimond et al, 2007). Specifically, from studying the available data in our combined dataset, we determined which facets could be created for our exploratory interface. During this process we encountered cases in which the common schema that we had decided upon, and which we thought would cover all eventualities of schematic representation required by our sources, needed to be altered to take account of additional metadata types found in new data sources which had not been anticipated. This necessitated the revision of the common schema, which in turn required the remapping of previously integrated sources. For example, we had specified that under a root classification of "People," the concepts of "Author" and "Composer." We then discovered that some data, specifically that supplied to us by Copac (`http://copac.ac.uk/`) in MODS-XML export format, did not differentiate between authors and composers, but instead referred to both as a "creator",[1] Therefore, in order to preserve the granularity of all data sources, we had to adjust our type hierarchy by inserting the concept of "Creator" on an additional level below "People", and above "Author" and "Composer" (see Figure 1).

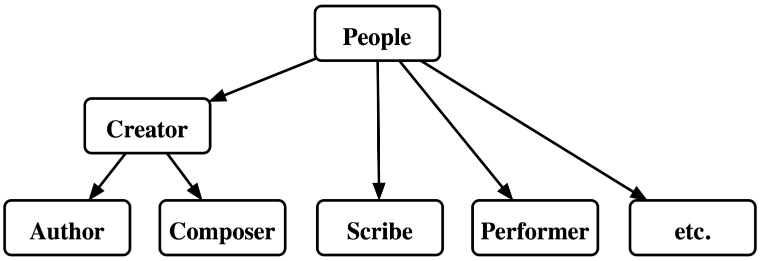
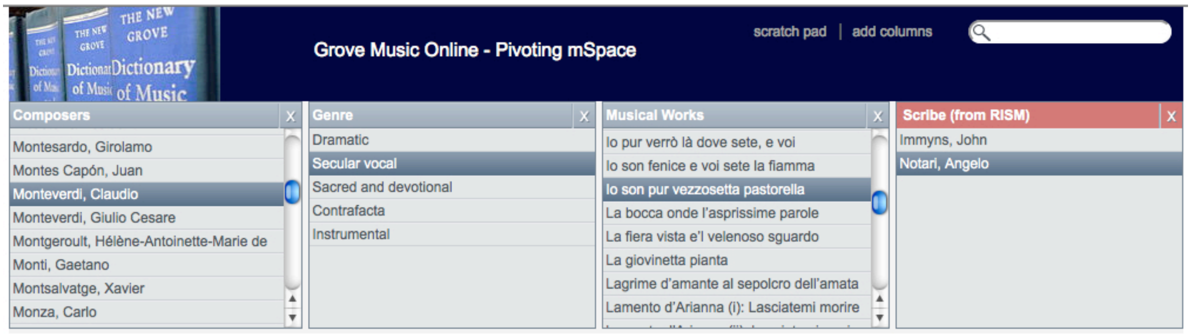


Figure 1. Illustration of the type hierarchy, with an additional level added for the "Creator" metadata.

This process would become more efficient if we could reduce the complexity of mapping to common schema, while also achieving a number of the benefits of doing so. To extend the Grove music example above, an example interaction is to search for songs in Grove, and stream them from Naxos Music Library (`http://www.naxosmusiclibrary.com/`). To do this currently, a user will search Grove by genre, retrieving a list of songs, and then cross reference that list of songs manually with recordings in Naxos, usually by querying Naxos's search interface for each song individually. By enabling a pivot on the songs in Grove with recordings in Naxos, a user can use a single pivot-enable mSpace to perform the cross-referencing for them automatically.

As a further example, consider occasions where a facet is only present in a single collection, and therefore, cannot be mapped to other sources. For instance, the Répertoire International des Sources Musicales (RISM) UK and Ireland (`http://www.rism.org.uk/`), provides metadata for each instance of an individual physical manuscript copy of a score, and thus has the facets of "Former Owner" and "Scribe" (all categories of "People"), which are not used by the other data sources that we are working with. In this case, using a pivoting approach enables this cross-reference entirely, as is illustrated in Figure 2 by the screenshot of our prototype.

[1] This type of schematic problem is by no means specific to Copac, and is in fact commonly encountered in scholarly collections that have various data legacy or curation issues.


THE NEW GROVE Dictionary of Music
Grove Music Online - Pivoting mSpace
scratch pad | add columns

Composers	Genre	Musical Works	Scribe (from RISM)
Montesardo, Girolamo	Dramatic	Io pur verrò là dove sete, e voi	Immyns, John
Montes Capón, Juan	Secular vocal	Io son fenice e voi sete la fiamma	Notari, Angelo
Monteverdi, Claudio	Sacred and devotional	Io son pur vezzosetta pastorella	
Monteverdi, Giulio Cesare	Contrafacta	La bocca onde l'asprissime parole	
Montgeroult, Hélène-Antoinette-Marie de	Instrumental	La fiera vista e'l velenoso sguardo	
Monti, Gaetano		La giovinetta pianta	
Montsalvatge, Xavier		Lagrime d'amante al sepolcro dell'amata	
Monza, Carlo		Lamento d'Arianna (i): Lasciatemi morire	

Figure 2. Screenshot of our pivoting mSpace prototype interface showing three facets from Grove Music Online, with one facet from RISM (UK and Ireland) "pulled-in" using a semantic pivot.

Conclusion

Our initial approach to aggregating numerous musicological data sources for the musicSpace project relied on the manual mapping of heterogeneous data schemas by domain experts. This approach, while effective, was also labour intensive, because the central schema potentially needed to be revised as each new data source was aggregated. This raised questions as to the scalability of our approach. In this paper abstract we have presented a technique for lowering the costs of combining heterogeneous sources that relies on the concept of Pivot-Awareness for Faceted Browsers, and we have described how this technique has been applied to mSpaces built around Grove Music Online and Naxos Music Library, and Grove Music Online and RISM (UK and Ireland). This approach will lower the cost barriers to aggregating data sources into faceted e-research systems, and will enable the full potential of existing metadata sources to be realised.

References

McGuffin, M.J., m. c. schraefel: A comparison of hyperstructures: zzstructures, mspaces, and polyarchies. In: Hypertext '04: Proceedings of the Fifteenth ACM Conference on Hypertext and hypermedia, New York, NY, USA, ACM Press (2004) 153–162.

Raimond Y., Abdallah S., Sandler M., and Giasson F: The Music Ontology. In: Proceedings of the International Conference on Music Information Retrieval (2007) 417–422.

Robertson, G., Cameron, K., Czerwinski, M., Robbins, D.: Polyarchy visualization: visualizing multiple intersecting hierarchies. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves, ACM New York, NY, USA (2002) 423–430.

Wilson, M.L., m.c. schraefel, White, R.W.: Evaluating advanced search interfaces using established information-seeking models. In: The Journal of the American Society for Information Science and Technology (July 2009)