

## MusicAustralia: Towards a National Music Information Infrastructure

A paper by Ms Robyn Holmes and Dr Marie-Louise Ayres,  
presented at the International Symposium for Music Information  
Retrieval (ISMIR) Conference in Barcelona, 10-14 October 2004.

### **Abstract**

MusicAustralia is a national music discovery service, developed by the National Library of Australia and ScreenSound Australia, National Film and Sound Archive. The service aims to provide seamless access to music and music information resources, in multiple formats, from custodians across all cultural sectors. This paper describes the development of the service, including its architecture, and content base. Service development to date has concentrated on metadata contribution and discovery strategies, together with development of the national digital music collection. In the future, digital content developed to populate the service could be subjected to Music Information Retrieval applications, to further enrich understanding of Australian music. The paper finishes by examining the challenges of achieving these advanced services in an environment where MIR research is relatively undeveloped.

### **Introduction**

This paper sets out the challenges, issues and methodologies involved in developing a cooperative national music information infrastructure. MusicAustralia has been developed through a partnership between the National Library of Australia<sup>1</sup> and ScreenSound Australia, the National Film and Sound Archive<sup>2</sup>, in cooperation with a range of institutions and specialist music projects across sectors<sup>3</sup>. We want any user, anywhere, to find and access Australian music – across all genres and styles, historical and contemporary, and whether online or

not – and to find and access information about Australian music and music-making in multiple forms, including online entries about people and organisations, or books, websites, manuscripts, ephemera or pictures. This free service operates on a federated model, with a centralised metadata repository for resource discovery coupled with distributed management and delivery of resources and information across multiple organisations, small and large<sup>4</sup>.

What makes MusicAustralia conspicuous is its approach to national cooperation: building a collaborative mechanism that enables institutions and the musical community to create, share, harvest, and aggregate the nation's musical data, resources and information in one virtual, seamless space – and as cheaply as possible! The success of such a national service is dependent not only on a sustainable central discovery service, but fundamentally upon the willingness and capacity of contributors to participate in such a federation. In addition to building the central infrastructure, the National Library has cultivated desire and commitment from a range of organisations across sectoral boundaries. And we have built capacity among custodians or creators of music resources and information – capacity to share metadata about resources and people; capacity to build and deliver digital collections to ensure technical and business sustainability; and, we hope, capacity to think ahead to the ways in which these collections might be exploited for new business – including 'musical' access by Music Information Retrieval (MIR) applications – in the future.

## Background

Building a demonstrator<sup>5</sup> – released at an international music librarianship conference in August 2002 [4] – was a key and effective strategy in generating excitement and commitment from our stakeholders. When the project began in late 2001, there were no digital printed music collections in Australia, and although a small number of national institutions had been digitally preserving their sound collections for some time, a business case to deliver these collections online had not been made.

At the National Library of Australia, the new business imperative to digitise and deliver the Library's printed music and extensive archival sound collections to support MusicAustralia had a major effect on development of the Library's Digital Collections Manager, an in-house application to support collection, storage, management and delivery of the Library's digital collections across all formats – pictures, printed music, manuscripts, maps, sound, books. Music specific needs provided solutions for other collection types where items consist of multiple parts and online navigation must be logical and as easy as possible for users.

The Library's experience has been replicated in other institutions; having developed digital systems for printed music, contributors have moved on to digitise and deliver archival holdings, newspapers and other complex formats. Managing and delivering complex sound collections – which may require simultaneous delivery of sound and transcript, or navigation to tiny musical fragments in an oral history interview – has proved even more challenging. However, the National Library and ScreenSound Australia are poised to make large portions of their sound collections available online from early 2005, thanks to further application development, and effective approaches to the rights and permissions issues arising from such collections.

The MusicAustralia demonstrator adopted a model used successfully to deliver the PictureAustralia service<sup>6</sup>, which provides access to more than 1 million online images delivered by nearly forty Australian libraries, archives and museums. The PictureAustralia model is extremely simple. Contributors map their local descriptive metadata to unqualified Dublin Core and make it available for harvesting by the PictureAustralia repository. Users discover images through PictureAustralia, but immediately navigate to the contributing organisation's digital delivery system to view the image. The PictureAustralia model was pioneering when it was adopted in 2001, and has been widely replicated by other services around the world. It is still an extremely effective and efficient way to provide users with access to many collections, and is relatively cheap both to run, and to contribute to.

We needed to test whether this model was suitable for a national music service. We particularly needed to test whether a simple metadata format such as Dublin Core was adequate for a service incorporating almost any imaginable material format: scores, born-digital scores, sound, multi-media objects, manuscripts, websites, pictures, text. We also needed to test whether this model would give users the kinds of 'seamless' navigation experiences we envisaged, e.g. viewing a score delivered by one institution while listening to a series of potentially related sound recordings delivered by one or more other contributors. At the back of our minds were the implications of a distributed national digital music collection for future MIR applications.

Our initial assumptions that the pilot would lead directly into a full production service of the same kind were overturned as we evaluated the lessons we had learned. We found that available descriptive metadata was very rich and complex – albeit in a number of different formats – and that reducing these complexities to simple Dublin Core metadata would do music and contributors a disservice. We found that – in contrast to the situation of pictures in Australia, and of music in many jurisdictions – the majority of the Australian notated music corpus was

already well described in Australia's national union catalogue, the National Bibliographic Database (NBD). Secondary sources were also well represented. The majority of the nation's published sound collection, however, was not described in the NBD, as its primary custodian, ScreenSound Australia, is an archive employing its own descriptive metadata structures in their custom-built stand-alone MAVIS<sup>7</sup> database.

Rights management was clearly a key inhibitor to the delivery of contemporary or older but in-copyright music. We therefore decided that working with 'aggregators' representing particular musical constituencies, and brokering rights on a licensing basis, would necessarily form a major part of the MusicAustralia strategy.

We also realised that we needed to put more effort than expected into assisting other contributors to digitise, store and deliver their music objects, or to transform preservation processes into full delivery systems. 'Going digital' required major rethinking and redevelopment for all concerned, and few contributors – even state libraries – had the kinds of IT resources available to a national library.

This realisation had a direct effect on our first phase ambitions. In particular, our hopes of enriching available metadata to better expose relationships between items and their parts, or between versions of works have not yet been realised. Similarly, although the centralised discovery and distributed content delivery model had great advantages as a 'discovery' strategy, we acknowledged that the diversity of storage, management and delivery systems was likely to be a major inhibitor to cross-service MIR applications.

The demonstrator did not include any work on the concept of providing information on people and organisations associated with Australian music, or our desires to display music resources in a context recognising creators and performers.

Following the pilot, MusicAustralia's joint project board – including senior staff from the National Library of Australia and ScreenSound Australia – made a number of decisions about the production service:

- MusicAustralia would collect descriptive metadata only into an XML repository;<sup>8</sup>
- for compelling business and cultural reasons, the national digital music collection would be a distributed collection;

- all resource metadata would be routed through the NBD, requiring new business solutions;
- a richer descriptive metadata format – the Library of Congress' *Metadata Object Description Schema* (MODS) [6] would be adopted as the service's preferred exchange format, to retain the richness of original records;
- MusicAustralia would provide access to all Australian music, and to all information about Australian music, in all formats, and whether online or not;
- MusicAustralia would 'piggyback' on new NBD services being developed, including services allowing users to 'get' physical format materials discovered through the service; and
- the National Library would develop a schema to encode information about people and organisations, *Metadata for Australian Parties Schema* (MAPS) [9] after evaluation of other available schemas. The locally developed schema will support conversion to and from *MARC21 Concise Format for Authority Data* [7], *Encoded Archival Description* (EAC) [3] and the Library of Congress' *Metadata for Authorities Description Schema* (MADS) [8] and other relevant schemas.

We also decided to work from 'easiest to hardest', that is, to develop the resource database first (core and well understood library business), followed by the people database. Value added services such as MIR and 'interactive' services such as collection of user annotations or individual contributions of data or objects are 'new' business, and will only be implemented as expertise and resources permit.

## Architecture

While the business solution to a scattered national music collection – a single discovery service – was driven by the inherent complexity of music as a symbolic and performed art form documented in multiple formats, the architecture we adopted for the service is not music specific. Indeed, we moved away from seeing music as having special needs needing separate information infrastructure, and towards thinking about how existing and developing national infrastructure could be used to ensure music service sustainability.

Our core decision was to build on the strengths of the Australian National Bibliographic Database, which has flourished for more than twenty years. More than one thousand libraries contribute to the service, which includes descriptions

of more than 14.2 million items, with 38 million holdings. More than half a million of these items are available online. More than 120 000 items are Australian music<sup>9</sup>.

This level of music representation was not just the result of serendipity. Even before we decided to place the NBD at the heart of our service, the National Library had worked hard to overcome problems with the way in which local music catalogue records appeared in the union catalogue, and to encourage key contributors to increase their cataloguing efforts. In fact, throughout 2002 and 2003, music was the fastest growing area covered by the NBD – evidence of MusicAustralia’s success in raising awareness of the need to enhance access to music information and to translate this into concrete actions.

Nevertheless, the NBD was essentially a service to which libraries contributed, and which libraries used on behalf of their clients. We therefore needed to rethink what the NBD could be, to find new ways in which other organisations – whether traditional collecting organisations or specialist online music services – could contribute to the NBD and to find new ways to exploit its rich data for end-users.

The time was ripe for such a rethink. The Library was formulating its future strategies for the NBD service, and in fact commenced a major redevelopment in the second half of 2003. New technologies and protocols such as the *Open Archives Initiative Protocol for Harvesting Metadata* [10] and eXtensible Stylesheet Language Transformation (XSLT) offered practical and affordable solutions to harvesting and converting data from many contributors and in many formats. Maturing XML data repository platforms also offered us opportunities to re-use existing data and to tailor presentation of the data to particular audiences.

MusicAustralia is also pioneering a new kind of federated service for Australia – a federated service for information about people and organisations. Libraries have long used ‘authority’ data – structured information to assist cataloguers to select appropriate name and other headings in bibliographic records – to enhance retrieval in online systems. However, sharing biographical and organisational history data across sectors has not previously occurred in Australia. Around the world, a number of projects to investigate such sharing are currently in place<sup>10</sup>. Many of these are funded to levels unimaginable in the Australian context, so we needed to find a cheap and affordable way to co-locate this disparate data, and to exploit it to provide people-centred context to the resource discovery aspect of MusicAustralia.

Our technical solution, therefore, consists of a number of modular – and therefore independently replaceable – components:

- a data harvester, capable of harvesting data in any XML format (DC, MODS, MARCXML, local schemas) by FTP, OAI or HTTP;
- a contributor profile specifying expected contribution behaviours, formats, conversion requirements and data destinations, e.g. the NBD, the MusicAustralia people database;
- a data conversion tool, capable of invoking specified XSLT conversion stylesheets according to contributor profile, e.g. converting ScreenSound Australia 's MODS versions of their original MAVIS records to MARCXML;
- the existing National Bibliographic Database utilities, to be replaced with new utilities over the next two years;
- a data extraction profile, capable of automatically extracting records from the NBD according to specified criteria, e.g. all music materials, and all materials about music;
- a maintenance suite, in an early stage of development; and
- XML data repositories.

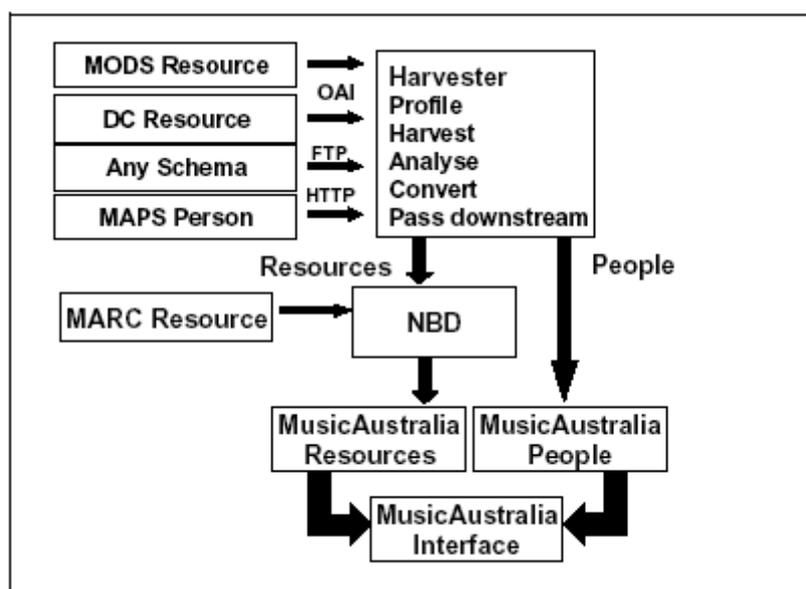


Figure 1. MusicAustralia Data Workflow

Records from the Australian Music Centre<sup>11</sup> and ScreenSound Australia formed the test-bed for this new infrastructure. Indeed, the achievement of the Australian Music Centre – with a total of a mere twelve staff – in redeveloping their business processes so that they could present their rich catalogue records in the MODS

schema and via an OAI repository for harvesting, conversion to MARC and inclusion in the NBD was recognised in the form of the Kinetica (NBD service) Innovation Award for 2004.

These data-sharing workflows can, of course, be used to expose Australian collections to an even wider audience. The National Library's entire digital music collection, for example, is now accessible through the US-based Sheet Music Consortium, and indeed metadata for the Library's entire digital collection is freely available for harvesting via OAI servers.<sup>12</sup>

The important thing to note here is that – with the exception of the MusicAustralia repositories – these new pieces of infrastructure are not specific to music, or indeed to any particular downstream repositories or data structures. The Library can re-use them for any number of purposes, and is indeed doing so. Examples include harvesting of Dublin Core metadata describing online government publications, and harvesting of records describing archival collections held by large and small institutions.

## **Production service**

The MusicAustralia production service will be released in late November 2004. A single interface will allow users to retrieve records from the resource and people repositories, using simple or advanced searches, or a browse interface. Users can sort and refine results, limit their searches to online items only, and save and email results within a single user session.

The resource database includes descriptive metadata for more than 120 000 music and music information items across all formats. Of these, approximately 12 000 items are available online, including more than 9 000 digitised scores, 2 000 sound recordings and more than 500 Australian music websites<sup>13</sup>.

The numbers of items findable through the service will increase enormously in 2005, when ScreenSound Australia will export its MAVIS records for harvesting and conversion to MARC for use in the NBD and MusicAustralia and will grow still further as other major non-library collecting organisations contribute their data to the service. The proportion of content available online will certainly increase, especially as institutions find appropriate business and rights management models to deliver their enormous digitally preserved sound collections, e.g. the National Library's own massive folklore archives.

Where content is available online, users will navigate to the contributing organisation's delivery system to view, listen, read or perhaps request a higher

quality copy of the item. Where content is not available online, an integrated 'Find and Get' service will enable users to request loans, or pay for document delivery in a single workflow, rather than having to contact individual contributing organisations.

The people database includes descriptive metadata for more than 13 000 people and organisations associated with MusicAustralia resources. More than half of these records include biographical or organisational history information, in addition to names, alternative names and dates. Many people records will be sourced from contemporary music services, such as Australian Music Online<sup>14</sup>, an audience development initiative of the Australia Council for the Arts, and illustrated with images from contributors' own collections.

Content highlights will be showcased in 'themes' – curated collections of digital objects around a particular musical theme, genre, period or person. Themes are designed both to give 'surfing' users an entry point if they do not otherwise know where to start, and to highlight content which might otherwise be 'overwhelmed' in a service in which most content is not available online, and in which online scores will significantly outnumber online sound in the short term.

## **Scheduled enhancements**

In the six months following release, new user services will be added to MusicAustralia. The Library's premier federated service – the NBD – is currently being redeveloped, and many of its new user services can be easily migrated to MusicAustralia, as the services will employ the same repository solutions and data structures. This includes extending document delivery options to e-commerce suppliers, such as music publishers and retailers. It will also include portal or 'My MusicAustralia' functionality, in which users can become registered users, able to retain saved searches and results, and able to specify alert services letting them know that new music content of interest has been added to the service.

Once these services are put into production, the basic MusicAustralia service will essentially be complete. The databases will be continually refreshed with new content, or new access to digital content, or new themes. The great advantage of the models we have selected is that the majority of MusicAustralia's growth will require little additional service specific effort. This is essential for long-term sustainability of such a service for music that is, after all, a very small and specialised part of larger institutions' business.

However there will still be a heavy emphasis on identifying potential content contributors – especially those outside the Library sector – and working with them to ensure that their content can be accessed through the service via metadata mapping, conversion and contribution. Likewise, we anticipate a continuing role in advising contributors developing digital repositories and delivery systems – especially in the university sector. We will continue to ‘intervene’ in early stages of new online music projects and services, for example JazzAustralia and the National Indigenous Recording Project, aiming to persuade developers and funders of the benefits of using standards based solutions, and structuring their data to support its inclusion in MusicAustralia and other federated services.

An important service for contributors will also be trialled in the first year of operation. Many contributors wish to provide access to in-copyright score and sound content, much of which is not available in the marketplace, and has little residual economic value. The National Library and ScreenSound are working towards brokering a content licence on behalf of contributors. This is new business for both institutions, and the pilot is an opportunity to establish whether this is a viable and sustainable model for enhancing access to the richness of our national music collections, especially new music.

## **Desired enhancements**

Other user services, however, are likely to remain on our ‘nice to have’ list unless recurrent funding is supplemented by external funds such as grants and sponsorship. While we would love to develop these services – and feel passionately about their value to the music and wider communities – our first imperative is to provide access to resources, and our first responsibility is to ensure that our service is sustainable in the long term.

MusicAustralia development has been fully funded from its developers’ recurrent budgets. All digital content – and the complex infrastructure which supports its capture and delivery – has similarly been developed from contributors’ recurrent budgets. There is no national digitisation strategy or funding in Australia, and very few avenues for raising either startup or maintenance funding for such projects and services. One consequence of this dearth of funding options is that music researchers working in universities have little or no access to digital music collections at their home institutions. This lack of access itself may be a factor in Australia’s relatively low MIR research activity, and indeed in the limited exposure to MIR’s possibilities among the general music research community.

MIR applications are, of course, high on this list of desired enhancements. But because we operate primarily within a collection, documentation, access and delivery framework, we are particularly interested in the ways MIR could be coupled with other metadata activities. For example, we wish to apply IFLA's *Functional Requirements for Bibliographic Records* [5] information model to the resource database, as we recognise and have documented [1] the great value its version modelling offers to music users. We also wish to build annotation services, supporting users to add their scholarly apparatus – or even non-scholarly views – on particular pieces of music.

We see the potential for MIR, FRBR and annotations to operate together and multiply benefits. Applying Musical Character Recognition (MCR) applications to printed music to produce machine-readable texts and generate MIDI files could, for example, indicate version differences between printed music items which human cataloguers do not always identify. Applications to create MIDI files from existing sound recordings could similarly reveal melodic relationships between sound recordings or indeed between scores and recordings which are otherwise hidden due to differences in descriptive metadata. We can envisage software applications that translate these version differences into appropriate metadata supporting FRBR displays. Similarly, we can imagine annotation applications which both allow users to contribute their knowledge about music resources or people, and which can trigger maintainers to establish relationships between music items, or between music items and people, or indeed between people and people.

Applications to support some of these services are already well advanced in the MIR and information research communities, and indeed in international demonstrator and production services. More importantly for a national service, key infrastructure capability to support these and other 'data-mining' applications is currently being built in Australian research universities. The National Library is a key partner in three large infrastructure projects recently funded<sup>15</sup> by the Australian Commonwealth Government. One is primarily concerned with developing institutional digital repositories, with the National Library developing a repository specifically for non-affiliated scholars and creators, and the federated discovery service covering all repositories. One aims to ensure that the myriad formats which will be deposited in such repositories will be preserved and accessible into the future. The third is developing systems to support authentication based access to a range of services – including repositories – across the higher education system.

Clearly, these three important infrastructure initiatives could be used effectively to support advanced music applications across collections, institutions and even

sectors. The work we have already done with universities contributing to MusicAustralia has placed music and its needs firmly on these agendas.

## Conclusions

The desire to apply musical solutions to information retrieval and interrogation of the resources in MusicAustralia has always been part of the long-term vision, and investigation of potential applications has informed the context in which MusicAustralia has been developed. However, for reasons outlined in this paper, business decisions have been taken to substantially drive and adopt generic repository solutions as a first priority.

This strategy has successfully positioned digital music services across formats, from preservation through to delivery, within the core sustainable business enterprise of the key partner institutions, the National Library of Australia and ScreenSound Australia. Music is only one part of the business of large cultural institutions, which have many competing demands on their resources and limited access to external or special project funding. However, their momentum to develop seamless access to the nation's documentary resources and information, delivered to anyone, anywhere, and for any purpose – and this has embraced music! – has shaped and guided some key tenets of the MusicAustralia service. Its audience is broad; its musical scope and coverage is comprehensive; discovering and getting musical objects and information with ease are primary goals; cooperation between institutions, organisations and individuals, across the nation and across sectors, is an essential; and its service model is loosely 'democratic', centrally sustained but with each custodial organisation responsible for all aspects of its content.

In the process, however, MusicAustralia has forged a national music information strategy and is creating an extraordinarily rich Australian music 'content bank' with enormous potential for exploitation from a variety of research perspectives.

Key questions, perhaps even tensions, underpin such exploitation: whose creative imagination and which discipline will bring such research inquiry to the service? Whose business is it and who will pay? Who will develop the tools, methodologies and applications – often seen as cutting edge, experimental, or risky – to apply systemically to a production level service that is interdependent with the capacity of its contributing organisations? How can the requisite research expertise be developed in a musical community when digital music collections and repositories in Australia are still in their early stages?

The opportunities for further MIR applications best lie in identifying the synergies between the needs of users, the needs and aspirations of researchers, and the capacity of the creative and research communities to engage with the service in partnerships that can bring external funding to their development. This project has the potential to move far beyond a national discovery service for Australian music to a powerful research environment capable of driving major innovation. Its strength lies in its cooperative national strategy, and it could support cooperative, interoperable and standards based MIR activities across the nation. *The question remains: who will drive this and how might the whole vision be achieved?*

## References

- [1] Ayres, Marie-Louise, "Case Studies in Implementing FRBR: AustLit and MusicAustralia", paper presented at Evolution or Revolution, The Impact of FRBR, Melbourne, Australia this and all subsequent cited online documents viewed 20 August 2004, <<http://www.nla.gov.au/lis/stndrds/grps/acoc/ayres2004.doc>>, 2 February 2004.
- [2] Dublin Core Metadata Initiative, *Dublin Core Library Application Profile*, <<http://dublincore.org/documents/library-application-profile/>>, 2002.
- [3] EAC Working Group, *Encoded Archival Context Beta*, <<http://www.iath.virginia.edu/eac/>>, 2004.
- [4] Holmes, Robyn, "MusicAustralia: A Digital Strategy for Music", paper presented at the International Association of Music Libraries (IAML) Annual Conference, Berkeley, California, USA, <<http://www.nla.gov.au/nla/staffpaper/2002/iaml8Aug02.html>>, 4 August 2002.
- [5] IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, UBCIM Publications, <<http://www.ifla.org/VII/s13/FRBR/FRBR.pdf>>, 2004.
- [6] Library of Congress Network and MARC Standards Office, *Metadata for Object Description Schema*, <<http://www.loc.gov/standards/mods/>>, 2003.
- [7] Library of Congress Network and MARC Standards Office, *MARC21 Concise Format for Authority Data*, <<http://www.loc.gov/marc/authority/ecadhome.html>>, 2003.
- [8] Library of Congress Network and MARC Standards Office, *Metadata for Authority Description Schema*, <<http://www.loc.gov/standards/mads/>>, 2003.

[9] National Library of Australia , *Metadata for Australian Parties Schema*, <<http://www.musicaustralia.org/schemas/maps/maps-v1.1.xsd>>, 2004.

[10] Open Archives Initiative, *Open Archives Initiative Protocol for Metadata Harvesting v2.0*, <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>, 2002.

## Footnotes

1. <<http://www.nla.gov.au>>
2. <<http://www.screensound.gov.au>>
3. Contributors include the Australian Music Centre (a representative body for Australian contemporary classical composers and performers), all Australian State Libraries, Australian university libraries, museums and archives, and several online services, including Australian Music Online and Australian Sound Design.
4. The National Library undertook extensive research into existing and developing international music projects.  
<[http://www.nla.gov.au/wgroups/projectma/Related\\_work/register.htm](http://www.nla.gov.au/wgroups/projectma/Related_work/register.htm)>
5. <<http://www.musicaustralia.org>> The demonstrator will be replaced by the production service in November 2004.
6. <<http://www.pictureaustralia.org>>
7. MAVIS was jointly developed by ScreenSound Australia and Wizard Information Systems, Australia , and is deployed in a number of international audio-visual archives, including the Norwegian National Library, the BundesArchiv, and the Library of Congress.  
<<http://www.wizardis.com.au/ie4/products/mavis/introducingmavis.html>>
8. The Library deploys the Teratext Database System from Inquisition Pty Ltd.  
<<http://www.teratext.com/index.html>>
9. As at 20 August 2004 . The documented Australian music corpus is, of course, relatively small, with 40 000 years of oral tradition pre-dating a mere 230 years of European settlement.
10. For example, the European Union's Linking and Exploring Authority Files (LEAF) project. <<http://www.crxnet.com/leaf/>>

11. <<http://www.amcoz.com.au/>>
12. The Sheet Music Consortium is a cooperative service led by UCLA, <<http://digital.library.ucla.edu/sheetmusic>>
13. Longevity of music websites, many of which include musical content, is being achieved through archiving these sites in the National Library's PANDORA web archive. <<http://pandora.nla.gov.au/index.html>>
14. <<http://www.amo.org.au>>
15. The Australian Partnership for Sustainable Repositories (APSR) led by the Australian National University , the Australian Research Resources Online to the World (ARROW) project led by Monash University , and the Meta Access Management System (MAMS), led by Macquarie University were awarded multi-year Australian Commonwealth Government funding in 2003.