

Survey and Feedback on Digital Preservation of Moving Images and Sound

Introduction

An online survey¹ was created to elucidate from the community information in the types of projects being run and their feelings about what constitute the main issues and barriers to this activity. The questions allowed a free form answer but were constructed to allow us to gather numeric information as well.

Analysis of Results

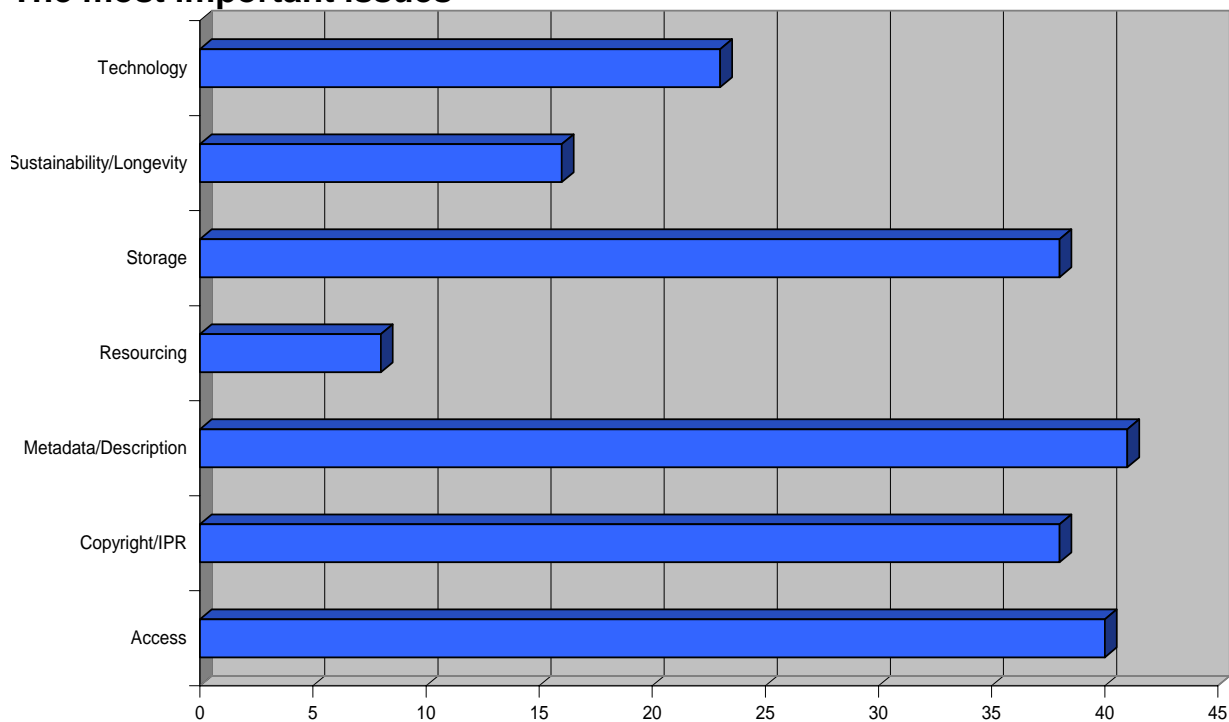
There were a total of 92 responses from the international community. Of these responses 35% were working on just digital moving images, 25% on digital audio, and 40% working with both media types. This suggests that the results will tend toward a moving image bias but will still be statistically significant.

The responses also tended to fall naturally into several clear categories for both key issues and barriers. These are:

- Access: relating to end-user access and the ability to maintain access and usability.
- Copyright/IPR: the rights to store, preserve and subsequently use content.
- Metadata/Description: the means of finding or describing the content.
- Resourcing: issues of cost, staffing and infrastructure.
- Storage: the means of storing digital content.
- Sustainability/Longevity: the ability to sustain the digital objects and services over time.
- Technology: the mechanisms and tools used for preserving digital content.

Respondents were free to include as many issues or barriers as they wished.

The most important issues



¹ <http://www.ahds.ac.uk/moving-images-survey.htm>

Metadata, access, copyright and storage are clearly of greater importance to the respondents than any of the other issues found. This suggests that these are issues which almost every project has to take account of and they are issues that are tricky to resolve.

Indicative comments that respondents made on issues are listed below in descending order of ranking:

Metadata (41 responses):

- Much metadata does not get to the heart of the information content, but concentrates only on technical metadata and somewhat banal and uninformative titles. The trouble is that providing really helpful metadata enabling resource discovery is very expensive.
- Metadata - technical and content details. Clarity about what is required, what is machine-readable, how the audio file and the metadata should be connected.
- The metadata standards are still being developed. METS / MODS / MIX / PBCore, etc. seem to be the most mature.

Access (40):

- Access is crucial I think, else what is the point... and good access is dependant upon decent search methods.
- Effective access mechanisms; digital rights management; appraisal, i.e. understanding which content to preserve to what level of quality
- The most important access issue is cataloguing of content - rather than metadata in its entirety - in the form of synopses, followed by copyright and the method of access.

Copyright/IPR (38):

- The key problem is copyright for materials which are strictly educational. The difficulty is finding a balance between the reasonable commercial interests of organizations and the long term educational and intellectual use of important materials. The issue is not one of commercial copyright, but of the recognition by commercial organizations and some museums, libraries and universities that inappropriate restrictions make work on, and use of, media materials very difficult. Yet it is increasingly evident these are crucial to any critical understanding of the twenty-first century
- Copyright is a big issue for 'orphaned works' - e.g. cassette recordings published by small no-longer-contactable publishers. We would like to be able to preserve access to this sort of material but it is at high risk of falling off the list because of the administrative cost of pursuing copyright clearance.
- Copyright is a big issue for performance work, as is consent. We have to ensure that all members of groups are aware of not just recording but of the storage of their image - what if someone became famous and someone sold the tape to 'before they were famous'?
- Unpublished field recordings involve performers and donors who sometimes feel strongly about widespread dissemination when the performers are not compensated, or in the case of possible misuse of music that was originally performed in a certain cultural context.

Storage (38):

- Digital storage media will only have a short life compared to film.
- Generally I would have said digital storage in terms of a safe repository plus the backups, simply because the places who do this in Australia are mostly

there by Federal government good will, even the ones put in place by legislation.

- Storage in multiple forms and formats, lack of standardization, and knowledge that whatever is produced is subject to repeated technology transfers (lack of a stable electronic medium).

Technology (23):

- The whole zoo-ful of file formats codecs standards etc, many of them often proprietary.
- Formats (would be good if somebody could standardise like TIFF for images). What is the standard format for Video? MPEG, QUICKTIME, WINDOWS MEDIA?
- Constantly changing technologies, that is broadcast industry-led.
- 1. preservation of the physical media so that it can be read
2. preservation of equipment to read that media
3. preservation of the knowledge of how to use and maintain that technology
4. ability to decode or decompress often proprietary digital bitstreams into contemporary formats with fidelity

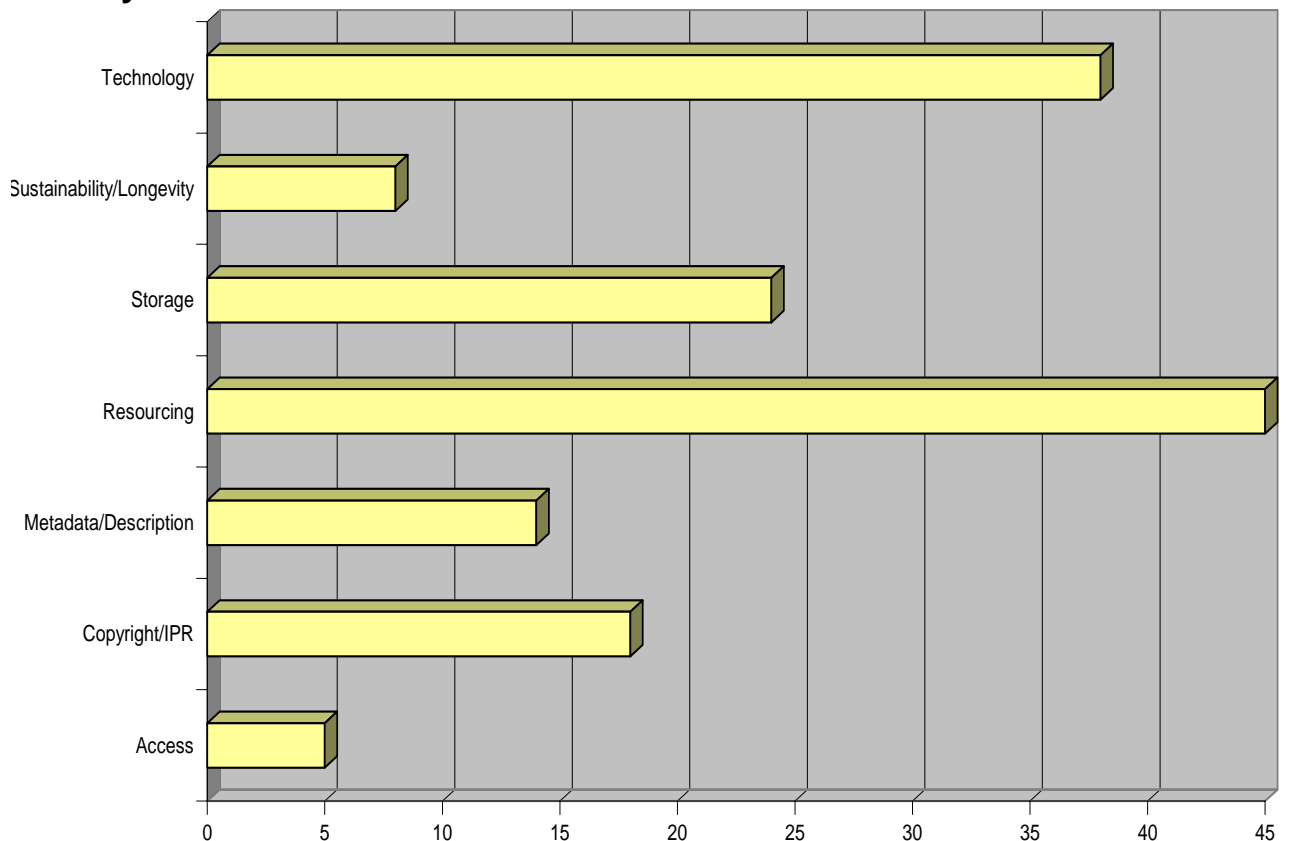
Sustainability/longevity (16):

- What happens if the internet ceases to be or changes beyond recognition understanding future forms of digital formats might help to flag up problems now.
- The most important issue may be compatibility. We will require agreed standards which will have industry support in the long term.

Resourcing (8):

- staff time for curatorial/documentation work

The key barriers



Resourcing and technology far outstrip any other responses, whilst storage, copyright and IPR make up the rearguard. This about face from the issues responses strongly suggests that these are barriers which are very common across most of the projects but are not deemed to be central to the activity itself. In other words resourcing is essential to enable the activity but is not seen as an integral issue of digital preservation is the way that metadata was perceived.

Indicative comments that respondents made on issues are listed below in descending order of ranking:

Resourcing (45 responses):

- Knowing what will last, with least staff intervention!
- The unknown cost and the lack of clear open standards which conform to an overall system. Technical skills in this area are evolving and so future skills required are both unknown and difficult to plan for.
- Technical know-how + cost of setting up and maintaining the equipment. Capture of preservation-worthy video is a very complex and time/money/staff intensive activity and requires so much specialized equipment & knowledge I can imagine it will be prohibitively complex/expensive for all but a few digital library programs. While outsourcing is a possibility, we've found that even the major vendors working in this area (transfer of magnetic media) don't work in a way that we'd consider appropriate for creating preservation masters.
- There is about fifteen years worth of digitizing time, at the current work flow rate.
- Convincing financial and political stakeholders that it is worth preserving.

Technology (38):

- The key barriers all derive from structure of the marketplace for digital media technology. Cultural Institutions must continually appropriate AV hardware and formats designed for consumer or professional use which naturally have a short product in order to meet their preservation goals. No manufacturer produces digital AV equipment with preservation in mind, and the complexity of contemporary digital technologies makes it impossible for an individual institution to design and build that equipment.
- The rapid change of formats and the cost of migration.

Storage (24):

- The cost of storage - access, metadata and copyright can be negotiated on, storage costs are the big one.
- Storage is still an issue for many institutions, obtaining uncompressed material, the need to provide alternate versions for access (web, downloadable in various formats) requires additional storage and processing.

Copyright/IPR (18):

- The key problem is copyright for materials which are strictly educational. The difficulty is finding a balance between the reasonable commercial interests of organizations and the long term educational and intellectual use of important materials.
- Clearing of copyright is the main barrier - this is time-consuming and intricate.

Metadata (14):

- The lack of metadata recorded by equipment, different levels of metadata being recorded by various equipment (same issue as digital cameras), obtaining appropriate descriptive metadata.

Sustainability/longevity (8):

- Digitisation is carried out constantly by many different staff in this institution - at the point of capture, decision about preservation and long-term usefulness are the lowest priority.
- As always the key barrier is a lack of sustainable resources.
- Uncertainty about future sustainability of digital repositories
- This material is intangible. We are essentially attempting to contain the horribly long list of ones and zeros in the right order in a format available to be read over a period of time which is untested. Seems a bit daunting.

Access (5):

- Archivists can find solutions to all of the issues noted in question 2 (except perhaps Copyright which requires government intervention), but the extent to which archives will be able to preserve and provide access to this type of material will be determined by the level of secure funding in the long term.

Conclusion

The responses seem to suggest that almost all projects have to address issues such as metadata, access, copyright and storage but that it is the levels of resource and the nature of the technology itself which are the major barriers to be overcome in achieving success. Moving images and sound is prone to very large file sizes and many formats and this was both a key issue and barrier for many projects – how to deal with storing all this content and then maintaining the usability when there are so many formats and standards to consider.