DIGITALPRESERVATIONEUROPE

DIGITAL PRESERVATION EXCHANGE

EXCHANGE REPORT

Funded applicant

Martin Iordanidis, MA

North Rhine-Westphalian

Library Service Centre

Hochschulbibliothekszentrum

des Landes Nordrhein-Westfalen

Jülicher Straße 6

50674 Köln, Germany

iordanidis@hbz-nrw.de

martin.iordanidis@googlemail.com

Hosting institution

BBC Information & Archives

Broadcast Centre

London, United Kingdom

W12 7TP

27th of February 2009

Table Of Contents

Institutional Context

BBC Information & Archives (BBC I&A) has a network of research centres and a range of self-research tools that aim to support production areas with information, audio, TV recordings and visual images required to make their programmes. Holdings include the BBC TV and Radio programme archives; Information & Archives manages a commercial music collection covering all musical genres, a news and photograph archive and a sheet music collection. *BBC I&A* also look after the BBC's programme documentation (scripts, programme files etc). Their research teams offer a range of research services that researchers can either mediate for specialist research, or can advise on, or arrange for access to online self-research tools.

The *North Rhine-Westphalian Library Service Centre (hbz NRW)* is one of six library networks in Germany. Besides offering traditional library services such as German Library Statistics, the hbz Union Catalogue and the hbz Digital Library, the establishment of various digital repositories for the sciences and humanities has become an important area. Dealing with the prospects and risks of digital files, strategies for the long-term preservation of 'document-style' data have been developed and refined within the digital library community. Their application to audiovisual material raises many interesting questions with reference to digital preservation in the broadcasting world and the notion of 'integrity' of information objects in general.

Purpose of Exchange

In a 2005 interview with Digital Curation Centre, BBC preservation specialist Dr Richard Wright addresses a "potential gap in industrially-based repositories [being in] contact with leading academic thinking, which leaves [broadcasters] at the mercy of vendors and the commercial market." [1]

Engaged in digital preservation of data originating from the scholarly domain at hbz NRW, the specific requirements of audiovisual data appear to remain widely unconsidered in the digital library world. While broadcaster's technical approaches to large-scale storage solutions are similar domains such as digital libraries, the broadcaster's real-life demands of digital library technology are yet largely undefined.

Potential alignments between IT-based broadcasting production processes and digital library technologies – mainly those concerning descriptive and technical metadata – were the core motivation for the exchange. Imminent issues regarding the preservation of digital files with audiovisual content are also important in both domains. During the exchange, two other incentives played an important

---

[1] Digital Curation Centre Resource Centre (2005). "Interview with Richard Wright, Information and Archives, BBC". http://www.dcc.ac.uk/resource/interviews/richard-wright

role. Firstly, the 2009 launching of the PrestoPRIME[2] project was to be introduced to BBC staff by Dr Wright in order to raise consciousness for digital preservation in-house. Secondly, the applicant wished to sharpen his objectives for a PhD thesis[3] covering audiovisual preservation issues on the level of metadata. In this respect, attending BBC Archives was also to lay a foundation for further research. Although the doctorate is not directly linked to the PLANETS project [4], its proximity to University Of Cologne possibly allows for integration with deliverables of the PLANETS project.

Duration and Scope of Exchange

The exchange took place between January 5th and Jan 16th, 2009. The activities were centred around the *BBC Information & Archives* Department, but also incorporated visits to the *BBC Film Archive* at Windmill Road, *The British Film Institute* at Berkhamsted and the *BBC Research and Development Dept.* at Kingswood Warren.

The one-day visits initially served to enlighten me to 'traditional' approaches to film preservation. Both the BBC and associated institutions have gained decades of practical experience in preservation of audiovisual materials. Even these traditional ways of audiovisual preservation involve a great deal of digital storage. However, digital preservation in the sense of storing and maintaining *files* is opening up new grounds.

Since the exchange took place during the launch of the PrestoPRIME project, a secondary benefit was the chance to have an update on BBC activities that are - or will become - relevant for digital preservation issues.

BBC Information & Archives Department

Cataloguing Workflows

Within a general introduction, the range of activities of the *BBC I&A* Cataloguing Department were outlined by Guy Strickland (Media Management Development Leader). Throughout the following days, the major tasks of the *BBC I&A* units were presented by several staff from the points of view of

- TV and Radio Cataloguing
- Music Library
- In-house Document Management

Lindsey Sellors (Media Manager Online Information) clarified the characteristic workflows that support requests for content both from inside BBC and - to a lesser extent - external production companies. In

---

[2] http://wiki.prestospace.org/pmwiki.php?n=Main.PrestoPRIME
[3] The doctoral thesis is overseen by Prof Manfred Thaller, University Of Cologne.
[4] http://www.planets-project.eu

her weblog on Information Management[5] Lindsey regularly publishes valuable resources on video archiving and metadata within a digital context.

Robert Cousins and Lucy Wales (Media Managers TV/Radio Cataloguing) gave insight into the two approaches to TV/radio cataloguing.

In the **non-viewing** Post Production Paperwork approach, the main data source for cataloguing consists of transmission lists delivered by production teams. Using the software packages P4A[6] for TV transmission and Proteus[7] for radio reporting, summaries of programmes are transferred to the INFAX database.

Going by the **viewing** approach, programmes are screened and described using subject keywords. Instead of cataloguing shot by shot, pictures of the same type or showing similar objects are grouped and related to one subject using time codes. There is an enforced delay of 6 weeks between transmission of the programmes and the cataloguing process. All procedures are pursuant to an in-house Subject Cataloguing Manual that determines the handling of the following programme content:

- Network Programmes
- News Bulletins
- Regional Programmes
- Foreign Language Ethnic Programmes
- Open University Programmes
- Weather
- Repeats

Both the manual and printed examples of transmission forms were made available to the guest visitor. An ever-recurring impression was that *BBC I&A* maintains a *working archive* centred around the practical demands of programme makers both within BBC and associated TV production companies. *BBC I&A*'s core assignment is to offer decision support in assessing the reuse value of existing audiovisual content for ongoing TV productions.

Physical Intake and Acquisition

Intake in this context means the acquisition of physical and digital assets into a managed archival system. Assets are defined as discrete audiovisual media content together with its associated metadata.

---

[5] http://www.bloglines.com/blog/girlinthearchive
[6] P4A is a BBC-internal database holding basic descriptive metadata, transmission and production information.
[7] http://www.bbc.co.uk/bbctrust/assets/files/pdf/appeals/esc_bulletins/2008/brand_ross_moyles.pdf

The media content itself is stored on a number of various physical storage devices and related to the catalogue database via a unique identifier stored as a barcode. Since the early 2000s magnetic tapes following the open standard LTO (Linear Tape- Open[8]) have proven to be the preferred format for archival purposes. However, other storage devices such as DigiBeta and a small number of tapeless production systems are used as well within editorial workflows.

Similar to traditional preservation activities mentioned above, 'digital' is not to be used synonymously with 'digital files', as the major part of audiovisual material is not stored on hard disc but on discrete physical media.

It is also worth noting that the *BBC I&A* cataloguing department only receives tape copies of the *finished* programmes after transmission. The raw footage (*"rushes"*) recorded during the production process remains with the editorial staff. With this historically-grown workflow, insights into production processes are practically unfeasible.

Structure of the INFAX Catalogue

Regardless of its age, the INFAX catalogue database is still a vital technology for BBC Information & Archives. According to an extensive interview with Sue Turner, long-time Operational Manager within *BBC I&A* , the core of the system dates back to the 1960s. However, several generations of software technology and cataloguing standards have refined the INFAX databases and its content to this day.

The INFAX core is based on a LonClass[9] classifier scheme for keywords that describe the subject of shots within the catalogue items. Additional domain-specific fields qualify this scheme for its use in a broadcasting context. Lonclass is broadly based on Universal Decimal Classification[10], which is in turn a European derivative of the Dewey Decimal Classification[11]. The cataloguing of names relies largely on the Anglo-American Cataloguing Rules (AACR)[12]. Further BBC-internal classification schemes extend the INFAX database.

With its heterogeneous structure that has evolved throughout more than four decades, the INFAX database has become a stable and reliable, yet labour-intensive and expensive search instrument. BBC staff have had to get used to a 'language of its own' in order to use INFAX and widely agree on the 'clumsy nature' of the INFAX database.

As a first step to counter this, BBC management recommended the merging of catalogue data into the

---

[8] http://en.wikipedia.org/wiki/Linear_Tape-Open
[9] http://www.bbcmotiongallery.com/Customer/Help.aspx
[10] http://en.wikipedia.org/wiki/Universal_Decimal_Classification
[11] http://en.wikipedia.org/wiki/Dewey_Decimal_Classification
[12] http://www.aacr2.org

internationally recognised ISAD(G)[13] archival standard in 2008, where possible. Originally used for 'documents in a conventional sense', the standard extended its coverage for audiovisual cataloguing.
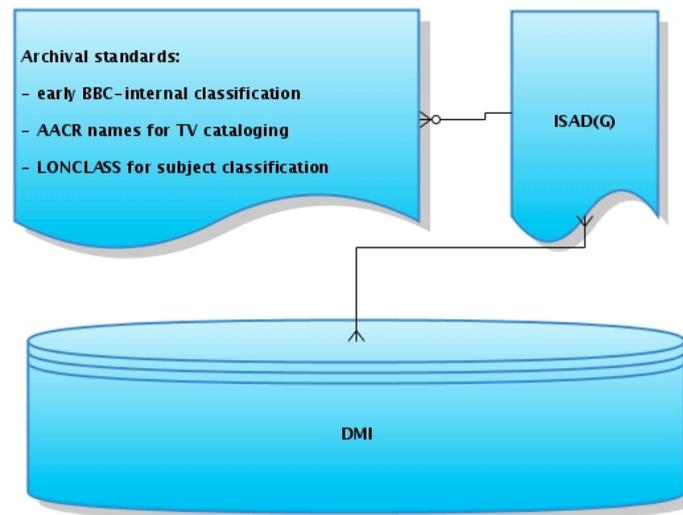

Fig. 1 - Harmonisation of INFAX catalogue data

Despite its trade-offs, the INFAX catalogue does serve a clearly-defined purpose; to this day it is piloting the decision whether production staff will order an archived tape through the BBC Information & Archives or not.

Once the order decision has been made, the physical media are identified by their INFAX barcode and sent to the recipients via dispatch rider. The BBC Film Archive is located outside central London – a fact that can cause considerable delays whilst facing time-critical production deadlines.

Having arrived, the ordered material could still turn out to be irrelevant to the production. Among many motives, it is traditionally-grown yet ineffective processes like these that have led BBC to a fundamental restructuring of its technology framework - known as the *Digital Media Initiative.*

Digital Media Initiative (DMI)

What is the DMI?

The Digital Media Initiative (DMI) aims to radically transform the way BBC produces content, making it quicker, easier and cheaper to make programmes. By creating an ever-evolving digital toolbox, programme-making, archiving and metadata management will be brought together in one single system. The DMI tools will partially automate routine tasks whilst making the required resources available at

---

[13] http://www.ica.org/sites/default/files/isad_g_2e.pdf

every BBC desktop computer. BBC's move towards a fully digital, tapeless production workflow is guided by its principal technology partner Siemens IT Solutions and Services (SIS)[14]. The DMI project continues the collaboration between BBC and SIS which started as a 10 year-outsourcing agreement in 2004.

Fig. 2 - DMI and BBC services

Data Migration and Preservation Issues

At the time of the exchange, BBC was in the midst of merging several legacy systems into a prime database using Oracle. In a presentation held by Peter Knight (INFAX Data Migration Manager), the following main legacy databases were highlighted:

P4A "Pre/Post Production Paperwork"

- *A standardized paper-based workflow documenting transmission information*

INFAX

- *Catalogue inventory ranging from the 1960s to the present day*

Sports Library Database

- *Based on an early INFAX spin-off database*

Music Reporting System

- *Data covering legal requirements to avoid damaging legal disputes*
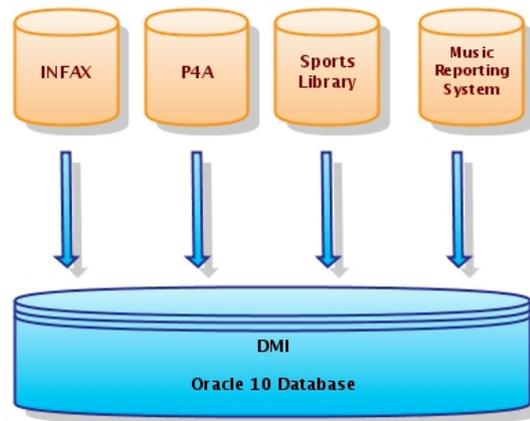
---

[14] quoted from the "BBC gateway" intranet.

Fig. 3: Data Migration within the Digital Media Initiative (DMI)

Within the migration process, data will move into the DMI infrastructure without any further refinements. At this stage, some degree of information loss is expected and taken into account by the BBC department in charge. An example of the information loss is generalisation of data caused by catenating separate legacy metadata fields into one field in the DMI prime system. All four legacy databases are known to have different attributes and strengths, however there is no active preservation action taken to keep the legacy data at this point.

The launch of the DMI system is expected for August 2009. Retroactive preservation action could be taken on the basis of the 'clean data' after the DMI has launched. Conceivably, the legacy databases could be preserved in text/csv-format stored on DVDs as a physical backup. The deployed legacy database technologies include SQL Server, Informix, Oracle and Access databases plus a small number of MySQL databases.

Defining Significant Properties

Above the complex structure of the legacy databases exists an even-more complex structure of metadata. Possibly being the most valuable starting point for preserving significant properties of the audiovisual assets, a "Consolidated Metadata spreadsheet" has been developed by a working group within *BBC I&A*. The metadata spreadsheet was introduced within a meeting by Margaret Katny, Archive Consultant at *BBC I&A*.

Documenting more than 2000 metadata fields together with their functions, the spreadsheet achieves a lot more than giving an overview of INFAX catalogue data structures. One paragraph is especially dedicated to the 'OAIS Functional Model' equivalents of the metadata and could be arguably translated into a list of significant properties on the level of metadata. This section was both designed and introduced by Elena Psarra, Media Manager Records Management, who has a specialist background in digital preservation requirements from the digital library point of view.

The allocation seen below could possibly set a foundation for an OAIS-oriented preservation architecture in the future, as it maps metadata being considered relevant to the OAIS-compliant modules.

| MetaData Name | Definition | OAIS Functional Model | | | | | | | Data Validation Types | Elements Required by iPlayer, Open Archive, APS and PIPS |
		Pre-Ingest	Ingest	Archival Storage	Data Management	Administration	Preservation Planning	Access		
Actual end date	As-run end date		x	x				x	Format - Date(s)	
Actual end time	As-run end time		x	x				x	Format - Time	
Actual publication non-linear metadata	any details we might need in future re actual publication by the Web, iPlayer, iTV, etc (e.g. geographical location, no of downloads,streams, etc)		x	x				x	Free Text	
Actual start date	As-run start date		x	x				x	Format - Date(s)	
Actual start time	As-run start time		x	x				x	Format - Date(s)	
Additional Notes Field	Any additional notes that may need to be taken into account eg. Any delivery requirements.	x	x					x	Free Text	
Airing Order No	If a series has a number of episodes, it will have an airing order number that may change at transmission.	x							Format - Numbers	
All Rights Release Form (Contributors)	Contributors				x				Selection - yes/no	

Fig. 4: Consolidated metadata spreadsheet (OAIS Functional model excerpt)

At this stage, the OAIS Functional Model segment is of secondary importance to immediate DMI activities. In context with digital preservation, however, it could gain top priority and should be reviewed by a larger working group within *BBC I&A*.

Preservation At Work: The D3 Project

In 2006, around 380,000 tapes in the now-obsolete D3 video tape format were stored in the BBC Archives. Some of the programs on the tapes were dating back to 1967 which had been transferred to D3 from earlier tape formats. The trigger event for transferring the assets from D3 into a file-based format was due to Panasonic ceasing support for D3 video players in 2006. The BBC D3 project was launched just in time. The preservation project was created to preserve the content of the D3 tapes by creating a file-based video archive, *"putting an end to the previous practice of transferring video from one obsolete video tape format to another video tape format whenever old video tape formats became difficult to support"* [15].

Out of the whole array of obsolete eD3 tapes, 100.000 items were selected by retention policies. The D3 project has started ingesting the assets in November 2007 and will require approximately five to six years in all. The BBC D3 team expects a total of 5-6 petabyte of digital files as the project's outcome, which equals roughly 1 petabyte (1024 TB/1048576 GB) per year. Checking routines include tests for historical errors, migration errors and Photosensitivity Errors (PSE) which are known to be catalysts for Photosensitive epilepsy [16]. The D3 project staff is located at the BBC Archives Windmill Road, West London. Thanks to James Insell (Preservation Specialist) and David Jordan (Project Manager & Technologist) the activities at this BBC location were introduced in-depth.

Within a highly industrialised workflow, the D3 tapes are converted into MXF (Material Exchange Format) wrapper files, a digital file format that has gained major acceptance in the professional video production and broadcasting domain [17]. The resulting digital files are stored on LTO-3 data tapes for archival purposes. Furthermore, the bit-streams are converted into a H.264-compressed [18] "browsing copies" which are likely to become convenient formats within the "Digital Media Initiative" context. In terms of metadata, the static fields of the INFAX catalogue are retrieved using the barcode identifiers as reference.

The resulting MXF files contain uncompressed video together with basic descriptive and technical metadata. MXF allows for the storage of technical metadata *within* the file, provided that the technical metadata creation is prepended in the workflow.

---

[15] Cunningham, de Nier (2007). "File-based Production: Making It Work In Practice". BBC Research White Paper WHP155
[16] http://en.wikipedia.org/wiki/Photosensitive_epilepsy#Television_programs
[17] http://ftp.irt.de/IRT/mxf/information/specification/index.php
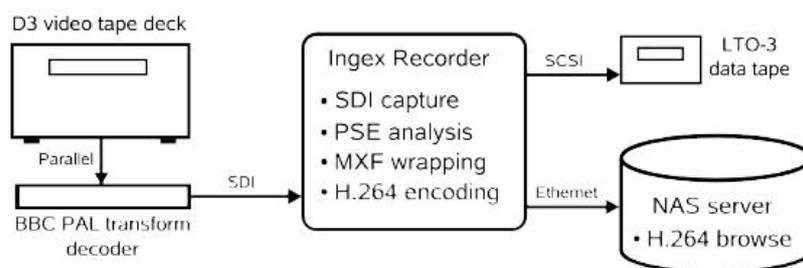[18] http://www.itu.int/rec/T-REC-H.264

Fig. 5: D3 project architecture

The D3 project can be considered a best-practice example of audiovisual preservation that offers parallels to sophisticated retrodigitisation projects in the digital library world. It is not only implementing a large-scale migration process with verified scalability but also preparing the outcomes for future application scenarios. The reliance on a powerful and standardized exchange format proves that the D3 project is - beyond its technological exigence - a heightening of preservation awareness. However, the paradigm shift towards digital files implies new requirements for preservation in the long term.

Conclusions

BBC Archives is, like many memory institutions, at the cutting edge of preservation. With decades of expertise in the practical migration of audiovisual materials on analog storage devices, their experience is likely to cross-fertilize future digital preservation processes. Clearly defined tasks of the *BBC I&A* department have resulted in transparent selection, retention and disposal policies which could be reused in digital preservation scenarios. The need for active preservation management is an essential part of BBC's corporate culture. A "Consolidated Metadata Spreadsheet" offers a promising starting point for defining significant properties of the moving image. It could undergo further analysis in order to build the backbone for *significant* preservation metadata. The D3 project and the Digital Media Initiative (DMI) denote a paradigm change towards digital files. It enforces the engagement in digital file preservation, which the BBC satisfies by partaking in the PrestoPRIME project.

Benefit And Future Research

The greatest benefit of the exchange was the hands-on insight into functional requirements of audiovisual content in the broadcasting domain. It became clear that metadata documenting audiovisual content has to be by far more explanatory as well as multidimensional compared to document-related metadata – for instance, the notion of *time* is still largely unconsidered in preservation metadata.

Having learned about the "real-world" metadata technologies within the TV/broadcasting domain,

structural similarities with wrapper formats, complex object encodings and exchange formats in the digital library world will be revealed more easily. At the same time, assessing the actual practical value of academic approaches to multimedia description became feasible.

Being occupied with repository management tasks on a preservation level at hbz, expectations from "off-the-shelf"-solutions for digital preservation have increased even more. It also became obvious that thorough understanding of the envisaged domain is mandatory to enduring preservation. Even though there seems to be little reference between digital library repositories and broadcast archives today, future generation preservation systems are likely to incorporate features from both sides.

On an institutional level, the exchange has created a potential link between Planets and PrestoPRIME as well as a link between hbz and the BBC Archive. All parties are held together by a strong common interest in preservation planning and development.

The applicant's future research activities will be closely related to the PrestoPRIME project. PrestoPRIME is a 2009 launching EU project that addresses the long-term preservation of audiovisual files. It is a follow-up to the PretoSPACE project[19] which set out to help institutions face the technical, organisational, resource and legal challenges in taking on the migration to digital formats. Out of the eight work packages defined for PrestoPRIME, BBC is responsible for two:

> WP2) basic approach to audiovisual digital preservation
> WP7) training and dissemination activities

As part of Work Package 2, Task 2 encompasses the production of preservation metadata. Commencing with the definition of significant properties in moving images, constraints for the completeness of an audiovisual Archival Information Package (AIP) will be framed. Considerations will include a comparison between the Material Exchange Format (MXF) structure and container formats known to digital libraries such as METS [20] and the OAI-ORE[21]. Furthermore, modes of interaction between digital library repositories and broadcasting archives as well as the integration of domain-specific metadata standards will be examined.

---

[19] http://www.prestospace.org
[20] http://www.loc.gov/standards/mets
[21] http://www.openarchives.org/ore

Acknowledgements