British Library Digitisation: access and copyright.

Ed King
Head, Newspaper Collections
British Library, UK[i]

Abstract

The British Library has in recent years undertaken a number of digitization initiatives. As the Library collections are very varied, containing many treasures, as well as a great mass of older printed texts, it has evolved a combined approach to its digitisation efforts in the last few years. High grade work such as the digitisation of the Gutenberg Bible, or the Shakespeare first folios have been realised - and all publicised under the "Turning the Pages" concept. At the same time, the Library sought funds for large scale digital conversion activities for three areas of its collections: Newspapers, Books and Sound Recordings. The paper will focus upon these three large projects. They are: Archival sound Recordings; 19th Century Newspapers; Out-of Copyright books. For the first two projects, funds have been awarded by the UK Joint Information Systems Committee (JISC). The Library has worked for the last few years with JISC to realise these two projects. To realise the third, the Library is in partnership with Microsoft.

The background to each of these projects will be described. The presentation of such a mass of material presents great opportunities to open up access. At the same time, the British Library must ensure diligence in relation to matters of UK copyright and to obtaining permissions of rights holders, wherever it is possible to do this. There will be an examination for each project of the copyright issues that had to be tackled, with examples of the constraints faced. The paper states that this remains work in progress, describing lessons learned; it may be impossible to trace all the potential rights holders within such large masses of material. However, The British Library has led in this area, and has a policy to show the process of due diligence.

The British Library is the national library of the United Kingdom and one of the world's greatest libraries.[ii] It receives a copy of every publication produced in the UK and Ireland The collection includes 150 million items, in most known languages. three million new items are incorporated every year. The Library houses manuscripts, maps, newspapers, magazines, prints and drawings, music scores, and patents. The Sound Archive keeps sound recordings from 19th-century cylinders to the latest CD, DVD and mini-disc recordings. The Library holds eight million stamps and other philatelic items. These Collections require over 625 km of shelves, and grow 12km every year. Whilst the Library has on-site space for over 1,200 readers, and over 16,000 people use the collections each day, the Library has sought in recent years to bring its collections to a wider audience via digisitation programmes.

BRITISH LIBRARY

British Library Digitization: aims

– maximize use of the collections by facilitating a greater volume of networked access, both in the reading room and remote, and by providing the enhanced functionality intrinsic to the digitized items

– reflect national and international priorities for wider access to, and enhanced use of, integrated collections of digitized educational, cultural or scientific materials which have a combined value greater than their component parts

–contribute to the conservation of original analogue materials for future users by substituting digital surrogates

–preserve the collection of analogue sound recordings

– generate income from those products with market appeal that can be exploited commercially by a partner, or the British Library itself, consistent with the aim of maximizing accessibility to the collection

**British Library digitization policy: within UK and international strategies**

–improving access to significant collections, both quantitatively and qualitatively, for all relevant communities with research and/or learning requirements for such reference material

–avoiding the unnecessary duplication of digitisation of particular materials

–avoiding duplication of cataloguing and indexing

–allocating responsibilities for digitisation of particular materials to designated libraries

–reducing long-term preservation costs by the use of large-scale data storage technologies in collaboration with partner institutions

Current examples of digitisation include:

- Collect Britain[iii]


- Treasures in Full [iv]
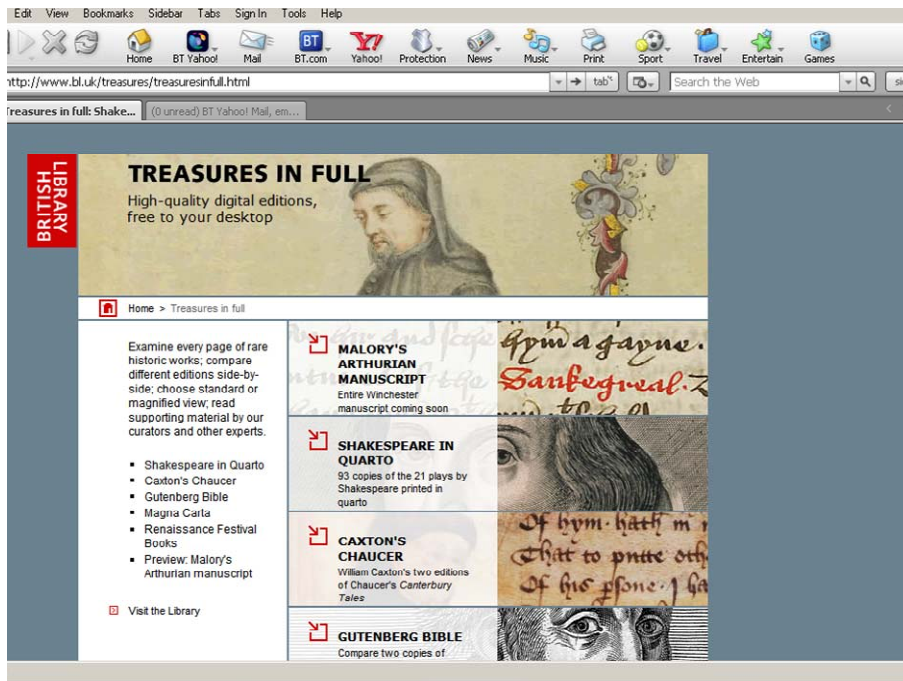
- Early Photographically Illustrated Books[v]
- Database of Bookbindings[vi]

In line with these objectives, the Library has sought funding and made partnerships to realise specific projects. In 2004, the Library was able to commence two large scale projects, with funds received from the UK Joint Information Systems Committee (JISC), whose mission is to provide world-class leadership in the innovative use of ICT to support education and research. These two projects were: **British Newspapers 1800-1900[vii]** and **Archival Sound Recordings**.[viii]

19th Century British Library Newspapers: access
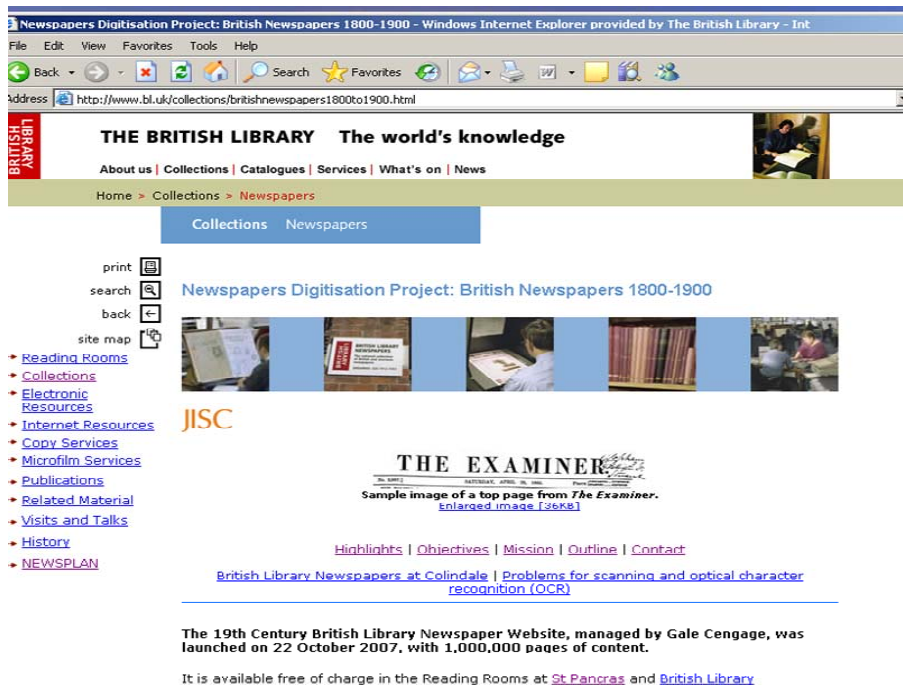
British Library Newspapers is the national archive collections in the United Kingdom of British and overseas newspapers. The collections consist of some 664,000 volumes and parcels of newspapers, together with 370,000 reels of microfilm, and together these collections occupy some forty-five kilometres of shelving. Since 2001, the British Library has been building upon its experience of digitising older newspapers.[ix] Activity in the Library accelerated with the start of the British Newspapers 1800-1900 Project in April 2004. This project digitised two million pages of nineteenth century newspapers, from 1800 to 1900. It was funded by the UK Joint Information Systems Committee (JISC), and formed part of its Digitisation Programme. A User Panel of experts was convened to steer the selection work, and a list of forty-eight newspapers published between 1800 and 1900 was agreed for digitisation. Preparation and production work was undertaken 2004-2007.
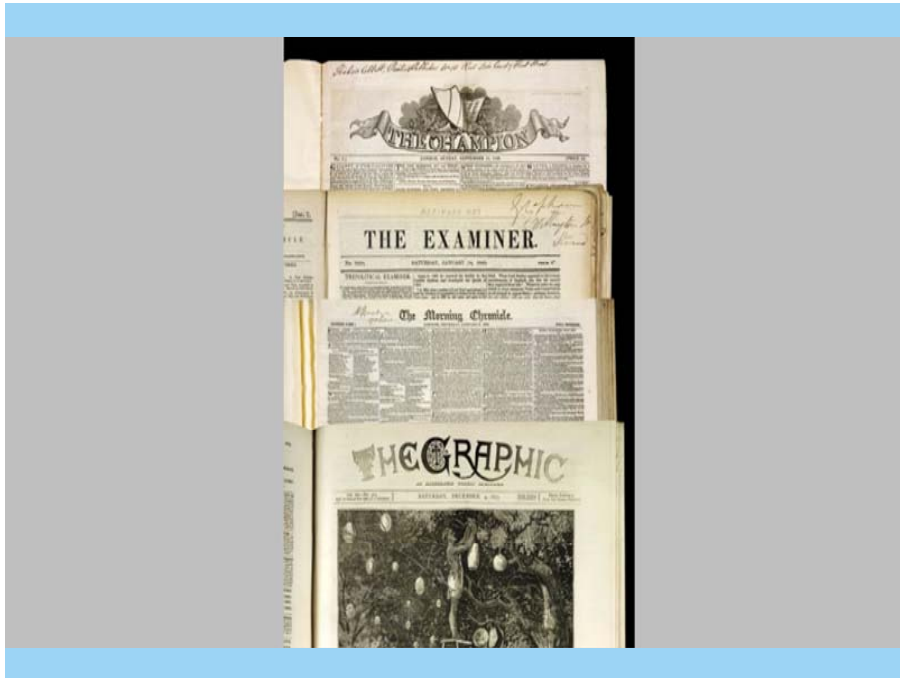
The forty-eight titles chosen represent a very large cross-section of 19th century press and publishing history. Three principles guided the work of the selection panel: firstly, that newspapers from all over the UK would be represented in the database; in practice, this meant selecting a significant regional or city title, from a large number of potential candidate titles. Secondly, the whole of the nineteenth century would be covered; and thirdly, that, once a newspaper title was selected, all of the issues available at the British Library would be digitised. To maximise content, only the last timed edition was digitised. No variant editions were included. Thirdly, once a newspaper was selected, all of its run of issue would be digitised. For example, *The Ipswich Journal* and *The Newcastle Courant* began publishing in the 18th century and continued into the 20th century. The project digitised the whole run of issues for 1800-1900. In addition to the regional newspapers selected, London national newspapers were also included, such as *The Morning Chronicle*, *Reynold's Newspaper*, *Lloyds Weekly*, with *The Graphic* and *The Pall Mall Gazette* starting publication later in the century. Also included were a number of newspapers relating to the Chartist movement in England, published during the 1830s and 1840s.

Liverpool Mercury.
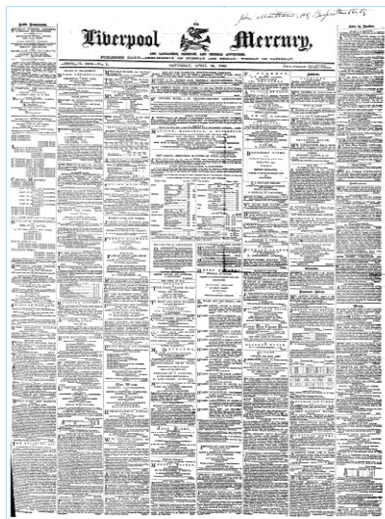
Articles on this page

* Advertisements & Notices

Articles found on this page
■ Select a title to find article on page
■ Mouse over article to see its title
■ Click any article to zoom and view

 You can imagine with a large number of stories appearing in each newspaper each day or each week, that this offers real difficulties when considering the matter of copyright permission to reproduce these texts.

19th Century British Library Newspapers: copyright

The launch to the UK Higher Education Sector took place at the British Library in October 2007. The BL formed a partnership with Gale to realise the web hosting and availability for the database. Gale is actively marketing the database under the title **19th Century British Library Newspapers**.[x] All of the two million pages are also be available as a free service to users in the British Library's Reading Rooms in London and Yorkshire. It is also free to UK Higher and Further Education organisations. These two million pages amounts to an enormous resource for research of all kinds, and their future availability to the UK and worldwide depends upon the matter of copyright being clarified for each newspaper. To date, over 200 UK Higher and Further Education Organisations in the UK have secured access to the database.

In the early stages of this project, the Library decided that newspapers published before 1900 were likely to be out of copyright. As part of ongoing dialogue between the Library and UK newspaper publishing companies, it was decided to write to all publishers, stating to them what the Library intended to do, and asking for comment. However, what seemed a straightforward step depended upon establishing who may be the rights owners. Many of the 19th century titles selected had ceased publication; others are still published; yet others were incorporated into other newspapers, and these in turn may also have been incorporated. An exercise was conducted to establish the name of the publishers.[xi] When the publishers were contacted in late 2004, there was but little response. The Library realised that publishers had not formed a view as to the value of the historic runs of newspapers that they owned, and that this probably accounted for the lack of response at this time.

During 2007 in planning for the launch of the digitised files of newspapers to the UK Higher Education Sector, the Library decided to be more specific about the copyright of newspapers published before 1900. After consultation with legal copyright experts, it was agreed that 1865 would be a cut-off date because papers published before 1865 are deemed to be out of copyright and papers published after this date could be in copyright

Much would depend upon establishing the date of death of the proprietors of each newspaper, in the period 1865-1900. It is known in the UK that pre-1912 the publisher's copyright ceases to exist seventy years after publication. If the proprietor of a newspaper published in 1900 died before 1936, the newspaper would be in the public domain in 2007; if the proprietor died in 1937, then the newspaper enters the public domain in 2008, etc. It was established that, of the forty-eight titles in the database, thirteen titles had ceased publication before 1865, and thirty-five titles published after 1865.

Initial checking of trademarks and to establish the names of owners was carried out.[xii]

BRITISH LIBRARY

Newspaper Copyright Checks: initial steps

1. checking whether each newspaper had been trademarked

2. contacting the UK Companies House in Cardiff to see what companies and proprietors were listed in 1900

3. conducting research into public records in the UK to determine proprietor names and the date of death of proprietors.

4. Looking at the online file of newspapers now available at the British Library to establish the name of the proprietor (or the publisher) of the newspaper at a given date.

Checking of names of proprietors/ publishers in press directories and other readily available references, both printed and online, was carried out.[xiii] It was also clear at the outset, that the Library could not possibly hope to carry out detailed checks on all articles in newspapers that had named authors. At this time, The Library simply has to wait to see if any claims may be made by the heirs of authors who wrote articles in these newspapers before 1900. At this distance in time, it may be impossible for either the Library or the heir to an author to prove what the nature of the contract was between a newspaper and the author at the time of its publication – it is unlikely that any contracts have survived. It is worth remembering that, as far as 19th Century British newspapers are concerned, only a minority of articles had author's names printed – it was not the practice of the day for these newspapers to print the names of the authors.
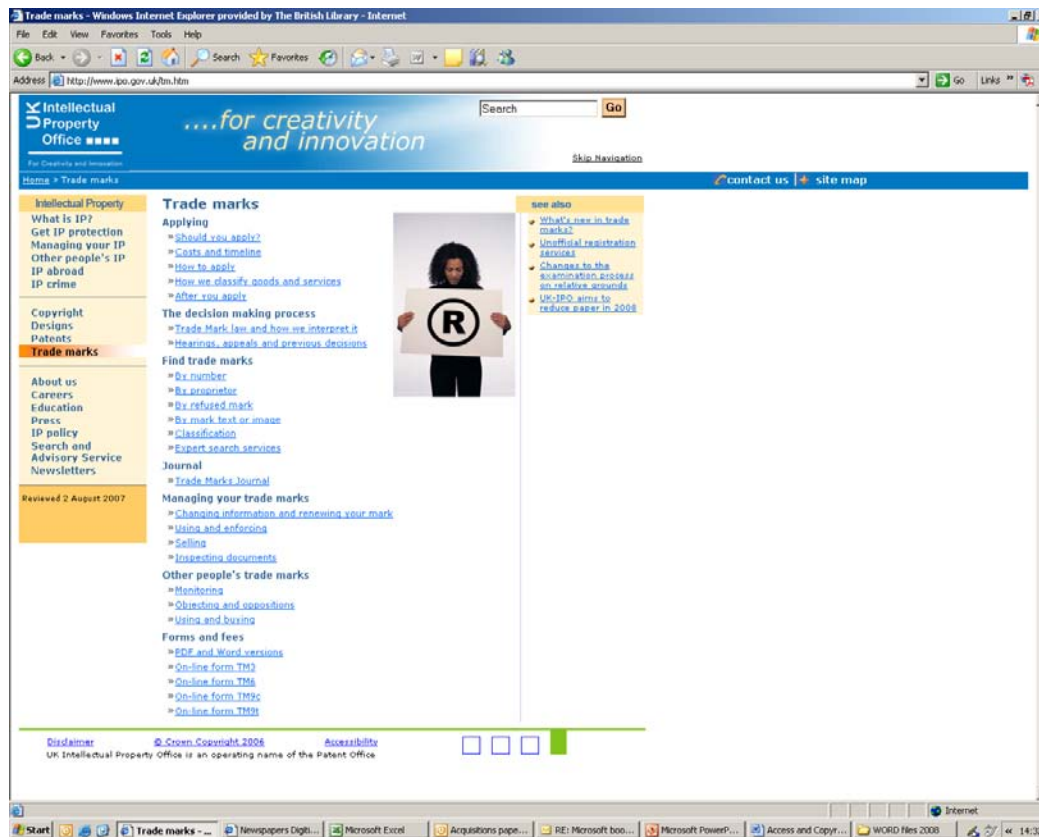
With regard to Trademarks, a search was conducted via the publicly available website for the UK Intellectual Property Office.[xiv] Of the thirty-five titles published after 1865, four of them had Trademarks. The Library is currently assessing whether any of these has relevance to the matter of making the individual title available online.

To carry out work at a detailed level, the Library commissioned a specialist company to do further investigations. It was necessary to consult records at the National Archives in Kew, London; to check local records offices in English regions. Some of the difficulties encountered during the work were:

Copyright Checks: problems

1. name of the proprietor could not be established at all

2. there were several names of proprietors, implying some form of co-ownership

3. names of company directors are cited, but without any evidence that one, or all, of these were the owners/ proprietors of the newspaper

4. the name of the proprietor is given as a company; the records relating to that company and its owners at a given time are not available.

5. the name of the proprietor is clear; however, it was not possible to establish the date of their death

This slide shows the sources that have been used to carry out the checking work for newspapers. As you can see, there are published works, histories, online files, website sources, printed newspaper history sources, and biographical sources, and other online resources.

**SOURCES USED FOR COPYRIGHT CHECKS**

Newspaper press directories (C Mitchell & co)

Willings press guide

The Encyclopedia of the British Press (edited by D Griffiths)

Companies House

The National Archives, Kew

The Story of the Pall Mall Gazette (John W R Scott)

Attacking the Devil (Chris Lloyd)

The Northern Lights (George Fraser)

BRITISH LIBRARY

SOURCES USED FOR COPYRIGHT CHECKS **(2)**

Dictionary of National Biography

www.newsletter.co.uk

www.newadvent.org

www.lancashire.gov.uk

www.hull.ac.uk

The Story Of the Western Mail

Sheffield Independent Book

Glasgow's Herald (Philips)

Welsh Biography online

As a result of the work carried out so far, it has been established that only one newspaper title is definitely out of copyright. For twenty-seven newspaper titles, the matter remains unclear: some dates in the run are out of copyright, with others being unclear, or the owner is not traceable. What the Library has done so far is to show as much diligence as reasonably possible to determine the matter of whether a particular newspaper is still in copyright. The process has been one of learning for us, for no other public organisation in the UK has yet sought to digitise so many titles, with so much content. The Library's efforts to determine dates of death of proprietors is not over, and the work of checking sources continues. The whole matter of permissions remains to be taken forward, in conjunction with newspaper publishers. The Library is also aware of work such as the European initiative i2010, which is looking at Orphan Works, and The Library is actively feeding into the debate regarding Orphan works, based on its experience.

Archival Sound recordings: access
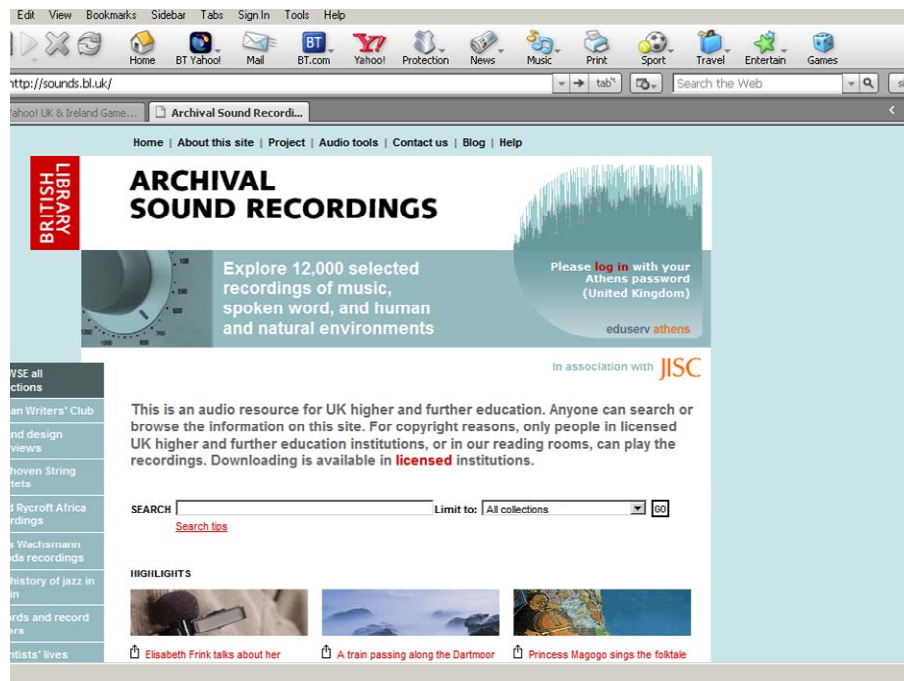
It was in 2004 that the **Archival Sound Recordings** (ASR) project began.[xv] The aim was to provide a major new online resource available free to everyone in Higher Education (HE) and Further Education (FE), providing easy access to thousands of hours of rare and historic sound recordings. The project was launched by the British Library in partnership with JISC (the Joint Information Systems Committee), with the intention will make available to students, researchers and academics some 12,000 unique materials from the dawn of recording history to the present.

The database was launched in 2006.[xvi] Archival Sound Recordings breaks new ground in the delivery of digitised sound recordings for use in education and research. It features a huge range of material, including classical and popular music, radio drama, oral history, and field and location recordings of traditional music.

Highlights of the fully searchable archive include:

- Unique and previously unpublished recordings of East African and South African music and cultural activities;
- The story of six decades of jazz in the UK, its varied styles, venues and characters, as told by musicians, promoters and label owners;
- A comprehensive archive of performances of Beethoven string quartets - unique in the way it reflects changing performance styles over the past 100 years;
- Insights into the lives and concerns of painters, photographers and sculptors through interviews with artists such as Elisabeth Frink, David Bailey, Fay Godwin, Eduardo Paolozzi and Anthony Caro;
- Radio material illustrating the richness and diversity of African writing and political culture during the 1960s and 70s.

- St Mary-le-Bow public debates

The £1m project has been made possible through JISC funding and is part of an overall £10m programme supporting the digitisation and online presentation of high-quality content including sound, moving pictures, newspapers, census data, journals and parliamentary papers for long-term use by the Further and Higher Education communities in the UK. The ASR service is accessible to any web user, but access to the audio content is limited to password-authenticated members of the UK HE and FE communities. The full service will also be available to users in the British Library's Reading Rooms in London and Yorkshire.

The website's interface was developed through extensive user testing to devise the best format for retrieving and playing back the recordings. Enhanced playback features will also allow academics and researchers to create play lists, and combine interdisciplinary material for their own projects, research and teaching resources. The digitisation work for ASR was carried out by Memnon Audio Archiving Services, which transferred recordings from a variety of analogue carriers to digital format, applying digital restoration techniques where appropriate. Hitherto, users had to come to the British Library's reading rooms, to access these collections. For the British Library to make available these recordings on its premises does not require a permission process. However, the act of copying and allowing downloads by academics and researchers in remote locations in the UK does require permissions, as these works are being made public, and therefore are "published".

Archival Sound recordings: copyright
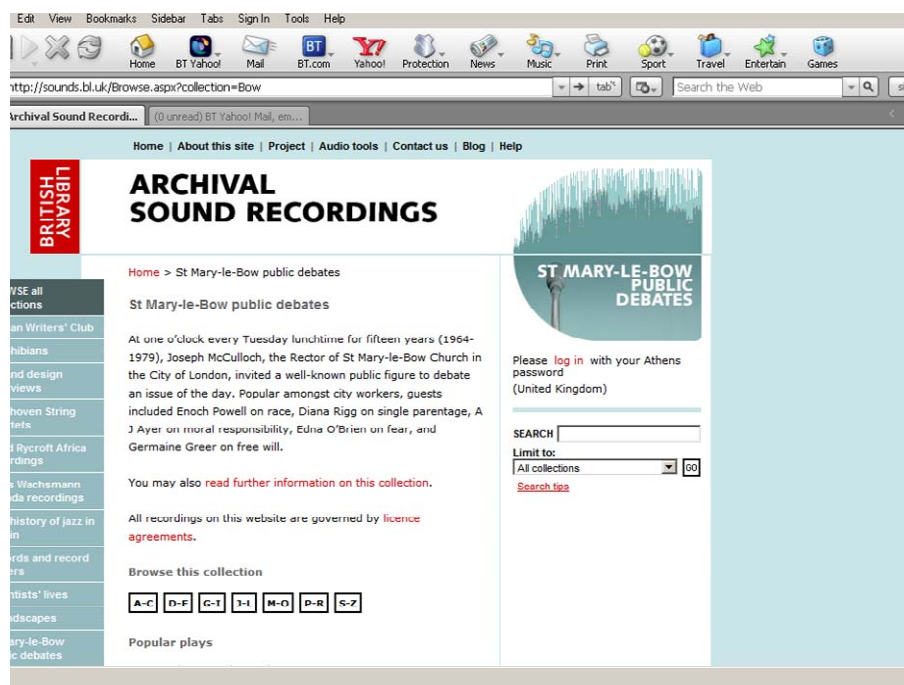
For example, you can search in the St Mary-Le-Bow public debates collection. If you search under the name of Diana Rigg, you receive a screen which shows her

discussing political freedom in 1976 . Here we see a brief description of what has happened to create this recording. To listen to the 32 minute recording, you then need to log in to hear it.

With regard to copyright, a couple of examples from within this project illustrate the problem that currently exists for material of this kind, which has high academic value, but low commercial value.[xvii] The St Mary-Le Bow collection is made up of recordings made between 1964 and 1979, created by the Rector of this London church, Rev. Joseph McCulloch. Well-know figures were invited to debate contemporary issues. Topics covered included poetry, race, free will, schools, education, architecture, and the church.

The project decided to eliminate all those recordings that included music and singing, to simplify the rights clearance process. It was also ascertained that no contracts were signed at the time of the recordings, so the rights situation was not clear. It was therefore decided to attempt to contact all the performers to seek their permission. The Library identified 299 rights holders whose permission was required. Some of these could be contacted and for some no contact could be made. Direct contact could be made to well known individuals, whose addresses were publicly available. Where addresses were unknown, contact had to be sought via the individual's former employers. In such instances, the UK law on Data Protection prevents employers from giving personal details to the Library, so the Library depended upon the employer to contact the rights holder. Some of the performers had died, and this raised the matter of seeking permission from the "estate" of the performer. For this collection alone, a total of 150 hours was spent by a freelance researcher, and 152 hours was spent by British Library staff on seeking permission, which resulted in eight permissions being received.

Another collection covered by this project was the African Writers' Club Collection. This has some 250 hours of recordings comprising dramas, political discussions and poetry readings. These were largely made by a company, the Transcription Centre, in the 1960s. These recordings equate to some 850 individual recordings. As this collection had not been previously catalogued, the project staff had to listen to all of the 250 hours, with information relating to performers, to literary readings, and to musical works, with all these being listed. Here, the matter of literary rights comes into focus: if a speaker is reading a text, rather than simply speaking, then permission needs to be sought from the original writer of the work in question. However, it proved difficult in practice to determine this, particularly in those cases where the recording is in African languages, such as Xhosa, Swahili, etc, and the requirement to effect a permission to publish on the web becomes very difficult, or simply impracticable to pursue owing to a lack of time and money.

Also in the collection of African material, there were excerpts of published and unpublished music recordings. Where names of the musicians could be ascertained, letters seeking permission to reproduce went sent out. Where this was not possible, a licence was taken out with the MCPS-PRS.[xviii] This organisation has links with music collecting societies throughout Africa, and the intention was to distribute some of the monies back to musicians in Africa. As a result of this work, the British Library was able to identify some 154 performers, but it is likely that the actual number of these would be far more than this, with 850 recordings being scoped as part of the project. A further round of work involved contact with the Writers Guild in London, who suggested additional contact be made with the South African Writers Guild, and they were also contacted without success.
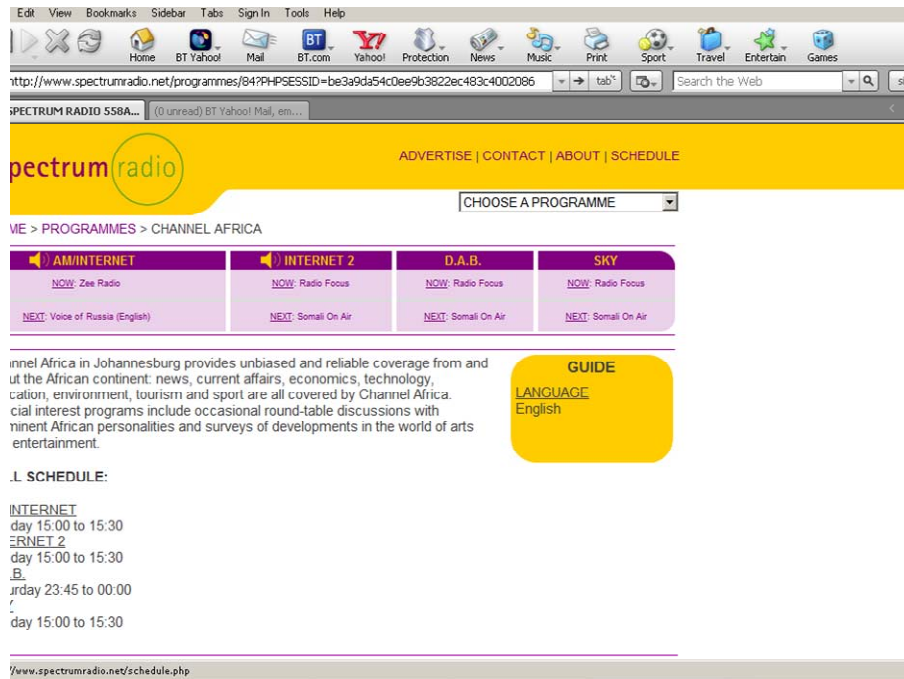
The Library also followed up a suggestion that an advertisement be placed through the Africa Centre on Spectrum Radio,[xix] which broadcasts nationally and internationally, reaching a large audience in Britain's African ethnic communities and in Africa. The Library secured permissions from the British Broadcasting Corporation (BBC), the Transcription centre, and the MCPS-PRS (via a blanket "African Music" permission). To date, The Library has been able to attain two permission from performers, one of whom worked for the Transcription Centre.
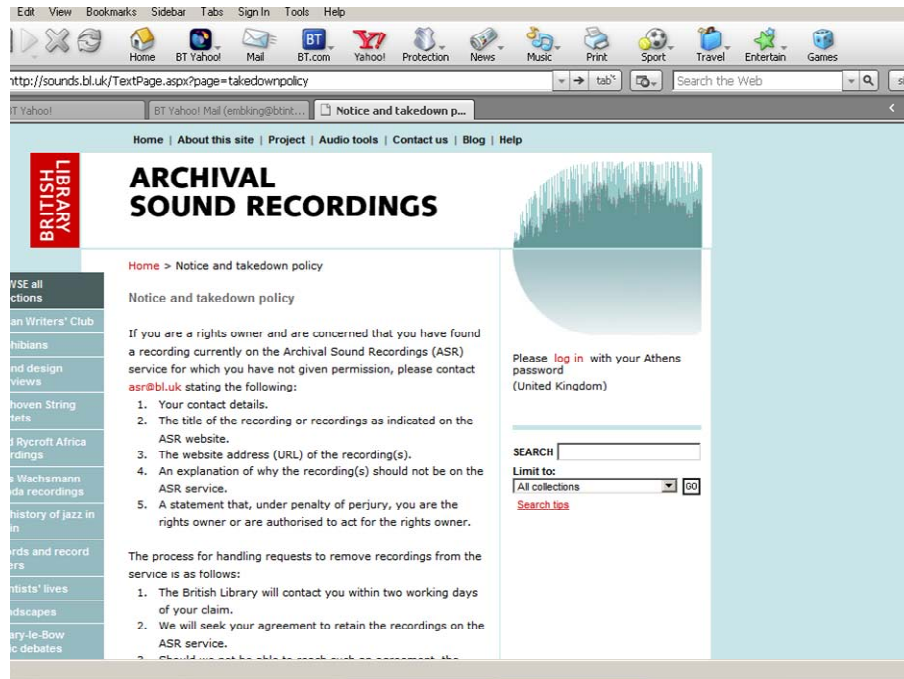
To continue to deal with any future claims that may be made, the Library has also implemented a Notice of Takedown Policy, to ensure that possible claimants may contact the Library.[xx]

As you can see, the issue of how to secure permissions from a hugely diverse set of source materials, for just two of the collections in the scope of this project, from the original performers, artists, and musicians of the time, has posed a great challenge for the Library.

The JISC has now allocated a second tranche of funding under its digitisation programme.[xxi] The ASR2 project will digitise a further 4,200 hours of recordings and approximately 10,000 images.[xxii] Subject to IPR clearance, the following collections will be digitised and made available on the ASR service: ICA talks, early spoken word recordings, early record catalogues, Jewish Holocaust survivors interviews, additional early classical recordings, Decca West Africa recordings, the Gerzon contemporary music archive and a substantial number of additional ethnographic recordings.

The current project is also developing an extensive online user community to extend the use the ASR resource for teaching, learning and research. Additional features such

as semantic tagging and resource sharing functions will be added to the site in time for project completion at the end of March 2009.

Microsoft / British Library Digitisation Project

BRITISH LIBRARY Microsoft Digitisation Project
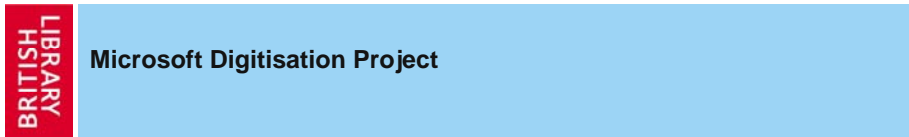
2

This project was agreed in 2005, with planning and pilots carried out throughout 2006 & 2007.[xxiii] Microsoft and the British Library have created a strategic partnership to digitise twenty-five million pages, around 100,000, of the Library's out-of-copyright printed book collections. Delivery of the search results will be through the Live Search Service;[xxiv] and the Library's Integrated Library System.
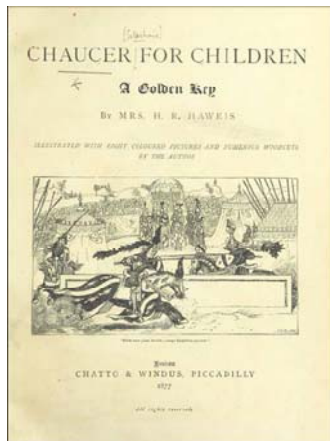
5

With books published before 1900, the likelihood is that they are out of print, and that many of the works are "orphan" works, i.e. they posses no clear holder of the copyright. However, as with the other areas of our collections where the Library is undertaking digitisation projects, it is prudent to consider what method could be deployed to acknowledge copyright.

Such work needed to balance the rights of the creator with the public good. Any solutions that the project partners could devise had to be time dependent; be scaleable; be work specific, and be flexible in application. In the UK, published works are deemed to be out of copyright seventy years after the death of the author. Books with authors who died before 1937 would be out of copyright by the year 2007.

After some months of planning and prior notification of the project in appropriate trade journals, the Library has embedded checks for authors into the pre-processing of the books through the production system. The author checking is done via retrieving the Library catalogue record from the Integrated Library System (ILS) of the British Library. Books published before 1900 are fed into a local record set. The 1900 date was set after joint Library / Microsoft discussions with two UK organisations: the Publishers Licensing Society (PLS) and the Authors Licensing Collecting Society (ALCS). This set is checked against data retrieved from the Dictionary of National Biography and the ALCS database. Where it is shown that the author may still be in copyright later than 1937, the record is not processed any further, and the book is not retrieved.
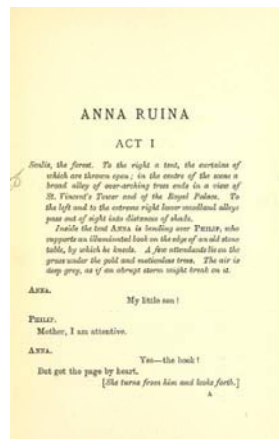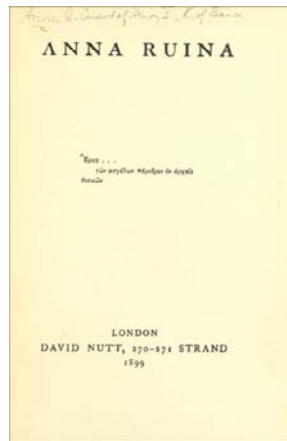
9

The remainder of the records are for books whose authors are deemed to be out of copyright, or are in-copyright but the copyright process employed by the Library has not identified a rights holder. Given the age of the material however, it is likely that only a small percentage of titles will be out-of-copyright.. Where is has not been possible to trace the author or the rights holder, these works are referred to as "orphans", because the author is not known and cannot be traced from the computer database. The Library currently estimates that a minimum of 40% of all copyright works ever produced (not just this project) fall into this category. The Library and Microsoft have agreed a "notice and take down" procedure detailed on their websites and will remove items if evidence from the rights holder is presented. The work of identification currently continues as the project builds production of titles being digitised.

Conclusion

Conclusion

1. For each project, a considerable amount of time has been taken to pursue the matter of obtaining permissions.

2. Necessary to keep driving these projects forward, to achieve the greater aim of making British Library collections more accessible.

3. Essential to plan into project proposals a generous amount of time to identify copyright holders, and to seek permissions.

4. Where copyright cannot be identified, it will be necessary to examine further if the wider UK national and international debate on orphan works will permit the Library to cite an international agreement or precedent

Work will continue upon the matter of copyright clearance and permissions for as long as the Library engages upon digitisation conversion work, via further specific projects. What has our experience shown in the last few years? As you have seen today, the working outcomes for each project have been very different, and the methods employed to determine copyright holders depend very much upon the published origin of the material: newspapers, sound recordings, printed books. The different points of creation and origin determine the research methods that have to be followed to identify copyright holders.

Our knowledge of this debate regarding un-copyrighted works is growing, but it may be several years before any significant measure of international agreement is reached on this topic.[xxv] Placing a number of works into the "orphan" category would also have the advantage of lessening the time taken to do all of the checking work, reducing the cost overhead to digitisation projects of this kind.

E M B King
12 may 2008

---

[i] I am grateful to my BL colleagues: Ben White, Licensing and Copyright Manager; the Project Managers Jane Shaw (Newspapers); Peter Findlay (Archival Sound Recordings) and Neil Fitzgerald (Microsoft Books), who have read and made comments upon this text.

[ii] See: http://www.bl.uk/aboutus/quickinfo/facts/index.html (Visited April 2008)

[iii] See: http://www.collectbritain.co.uk/ (visited April 2008)

[iv] See: http://www.bl.uk/treasures/treasuresinfull.html (visited April 2008)

[v] See: http://www.collectbritain.co.uk/collections/19thphotobooks/ (Visited April 2008)

[vi] See: http://www.bl.uk/catalogues/bookbindings/ (Visited April 2008)

[vii] See: http://www.bl.uk/collections/britishnewspapers1800to1900.html (Visited April 2008)

[viii] See: http://sounds.bl.uk/ (visited April 2008)

[ix] For the story of the British Library's steps towards large scale digitisation of newspapers, see: *Digitisation of Newspapers at the British Library*. **The Serials Librarian**, Vol. 49 (1/2) 2005, pp. 165-181.

[x] For further details, see: http://www.bl.uk/news/2007/pressrelease20071022a.html; and http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation/digitisation_bln.aspx; and http://gale.cengage.com/servlet/ItemDetailServlet?region=9&imprint=000&cf=n&titleCode=DASC12&type=4&id=242585
(visited April 2008)

[xi] With thanks to Gareth Burfoot of the the Library's Preparation Team, who carried out this checking work to establish previous company owners.

[xii] With thanks to Fiona McHenry of the Library's Preparation Team, who carried out this first set of checks.

[xiii] For example: *Mitchell's Press Directory* published annually, …; the *Oxford Dictionary of National Biography online.*

[xiv] See: http://www.ipo.gov.uk/tm/t-find/t-find-text/ (visited March 2008)

[xv] See: http://www.bl.uk/news/2006/pressrelease20060926.html (Visited April 2008)

[xvi] See: For more details see: **www.bl.uk/sounds**   (Visited April 2008)

[xvii] For an extended account of what is presented here, see: Benjamin White. *Rights Clearance of Sound recordings by the British Library. Copyright World*. Issue 167. February 2007, pp.24-26.

[xviii] See: http://www.mcps-prs-alliance.co.uk/Pages/default.aspx (visited April 2008)

[xix] See: http://www.spectrumradio.net/programmes/84?PHPSESSID=be3a9da54c0ee9b3822ec483c4002086
(visited April 2008)

[xx] See: http://sounds.bl.uk/TextPage.aspx?page=takedownpolicy (visited April 2008)

[xxi] See: http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation.aspx (Visited April 2008)

[xxii] See: http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation/asr2.aspx (Visited April 2008)

[xxiii] See: http://www.bl.uk/news/2005/pressrelease20051104.html (Visited April 2008)

[xxiv] See the article: Collins, Barry. *Digitising the British Library. How do you digitise 100,000 19th century books in two years?* In PC PRO, issue 160. February 2008, pp. 132-137.

[xxv] There are useful links from the Wikipedia article on Orphan Works: http://en.wikipedia.org/wiki/Orphan_works