

# **You Can Be Serious: *Broadcast Archives and Academic Discourse***

Richard Wright

BBC Research & Development

Re-Imagining the Archive, 12-14 Nov 2010

Broadcasting matters:

- what people listened to and watched in the 20<sup>th</sup> C
- large archives
- primary evidence for what people actually said, heard and saw
- engaging and compelling content

But audiovisual archive content is not:

- studied, cited or taken seriously
- considered a part of *scholarly communication*

**So: what can (or should) be done?**

**what people listened to and watched in the 20<sup>th</sup> C  
(and still do!)**

- 80% of young adults in the USA use broadcasting as a primary source of news
  - 30% use newspapers

Americans watch 38 hours of television per week

Britons are not far behind:

- 31 hours per week of television
- 22 hours of radio

[2007 data]

Richard Wright BBC R&D

FP7-ICT-2007-3 231161 PrestoPRIME

## large archives

total broadcasting output in 2000

- **100 million hours** of unique radio and TV content per year (globally)

Many European broadcast archives with over 1 million hours (BBC, INA, B&G ...) Sweden has 6 million!

Total *archived* audiovisual content: 100 million hours in Europe (based on survey totalling 30 million)

Total global estimate: **200 million hours**

- **primary evidence** for what people actually said, heard and saw
- actuality recording and footage [PDFs, next slide]
- speeches
  - Inaugural addresses (Lincoln?)
  - Nobel acceptance
  - BBC: the Reith Lectures
  - Addressing the nation: abdication speech, fireside chats, Churchill war speeches, Cuban missile crisis ... (Washington, Pericles?)

**–[www.bbc.co.uk/archive](http://www.bbc.co.uk/archive) = 56 curated collections:**

**Francis Bacon, Virginia Woolf, George Orwell at the BBC,  
Talking Hamlet, Coal Mining in Britain, Pop Music in  
Liverpool, Titanic Survivors, Birth of the NHS,  
Cambridge Spies, Burgess at the BBC, Suffragettes ...  
Plus 11 other BBC collections (Woman's Hour, Panorama,  
In Our Time , Learning Zone ...)**

**And this is a few hundred hours. INA has over 20 000  
hours of broadcast and related content – online.**

**-<http://www.bbc.co.uk/archive/collections.shtml>**

**-<http://www.bbc.co.uk/archive/writers/12240.shtml>**

## engaging and compelling content

Performance – Shakespeare with or without Olivier?

Interviews – Boy George on drug dependence

Vox pop – people in Berlin during airlift

Comedy – or just read the script? Smothers Brothers  
last programme, because of anti-war accusation

You are there – footage from Vietnam, the “TV war”

Reality Television – Meet the Borgias

And the news – time travel, with a navigator

Not considered a part of *scholarly communication*

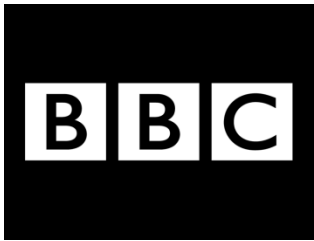
*Original Study: Accessing the Spoken Word*

Goldman et al, Jr on Digital Libraries, 2005

**Durability and accessibility** have given primacy to the written word for thousands of years. The exclusivity of text may now yield some authority to the spoken word ...

<http://papers.ldc.upenn.edu/ICDL2005/spoken-word.pdf>





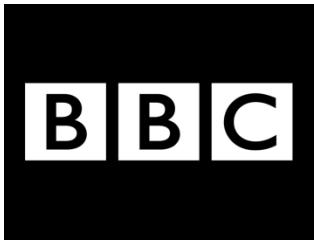
# Audiovisual archive content is not *serious*:



Not considered a part of *scholarly communication*

Example: Digging Into Data – 2009 'Challenge' in funding for data mining; NSF, NEH, JISC, Canada  
30 major online collections: NY Times, Internet Archive, Library of Congress, JSTOR ...

- seven digitised newspaper collections
- four collections with significant audiovisual content (cf Prelinger)
- NO broadcast content of any sort



# Audiovisual archive content is not *serious*:



Not considered a part of *scholarly communication*

*Example: citation indexes*

Paper by Prof Jaehnichen at IASA-AMIA, Nov 2010  
- archived audiovisual material not found in citation  
indexed journals: Thompson Reuters, ISI Web of  
Knowledge, Scopus, Springer Link  
[www.iasa2009.com/speakers/JAECHNICHEN.pdf](http://www.iasa2009.com/speakers/JAECHNICHEN.pdf)

Prof Schueller's comment: “no culture for going back to  
primary sources in ethnomusicology”

“**Durability and accessibility** have given primacy to the written word”

**Citation** needs both durability and accessibility

- Permanence: repositories with permanent URI's
- Access: all linked with a union catalogue
  - And tools to make a citation
  - And Infrastructure to render from citation

Then: an annotation *standard*

PrestoPRIME, W3G MAWG

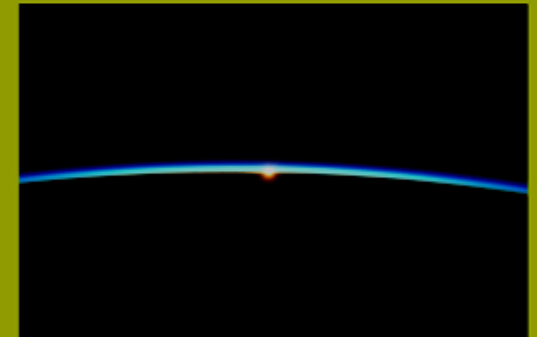
[www.prestoprime.org](http://www.prestoprime.org)

<http://www.w3.org/2008/WebVideo/Annotations/wiki/>

# Six requirements for serious access

- Granularity
- Navigation
- Reference and Citation
- Annotation
- Rights
- Access

## TRANSFORMING SCHOLARLY PUBLICATION THROUGH OPEN ACCESS: A BIBLIOGRAPHY



**Charles W. Bailey, Jr.**



OPEN ACCESS WEEK  
23 -27 March 2009



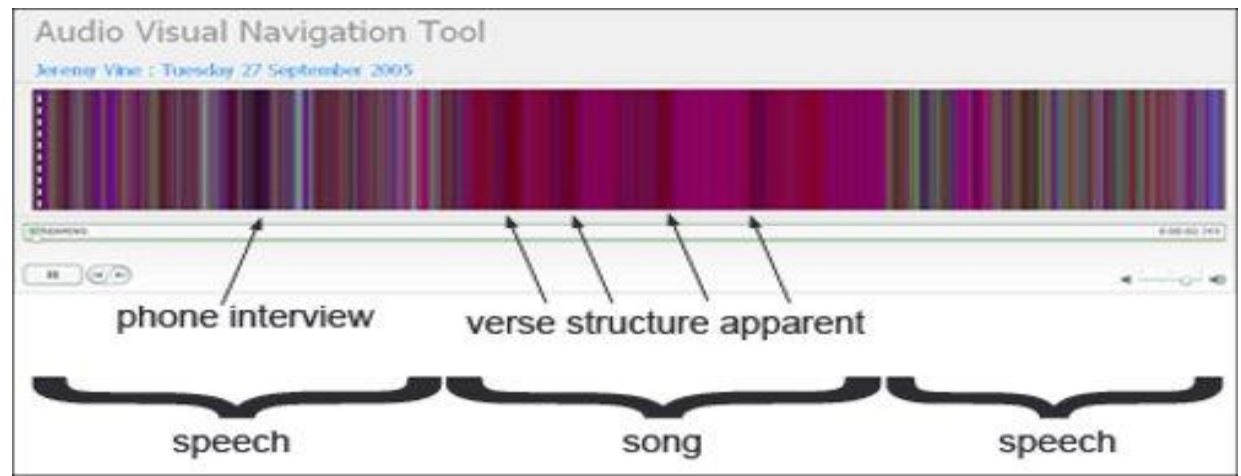
**M**Library

# Granularity - division into meaningful units

- Keyframes
- Other methods to represent video



- and audio:



# Navigation

- "Click and play" on visual representation of the meaningful units



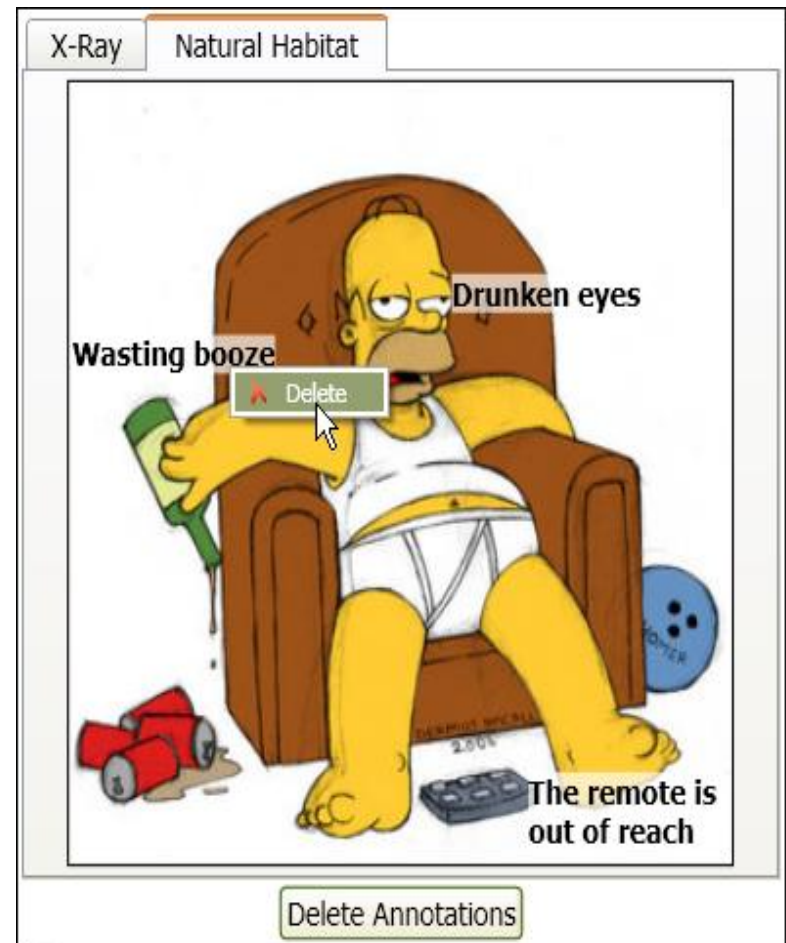
- The core requirement for scholarly discourse
  - along with a major change in attitude!
- Needs a permanent place for “things to be”
  - hence the need for stable audiovisual collections

*“Hamlet, for example, is comparable to Saxo Grammaticus' Gesta Danorum.[citation needed]  
King Lear is based on King Leir in Historia Regum Britanniae by Geoffrey of Monmouth, retold in 1587 by Raphael Holinshed.[citation needed] “  
wikipedia*



# Annotation

- the core requirement for social web = interactivity
- individual interacts with content
- individuals interact with other individuals





- 95% of cost of BBC trial of archive access
  - 20 hours per hour of content
  - Does not scale
- PrestoPRIME Ontology – supporting codification of owners, users and uses
  - Ten years ago we had matrices in a spreadsheet, now we have an ontology
- New laws needed (for publicly-funded content)
- USE “educational use” !!!

Digital objects can walk through walls (and violate most other laws of physics)

Unless access is denied  
– **interfering with a miracle!**



# Thank You

- PrestoPRIME [prestoprime.eu](http://prestoprime.eu)
- BBC R&D [bbc.co.uk/rd](http://bbc.co.uk/rd)
- Richard Wright [richard.wright@bbc.co.uk](mailto:richard.wright@bbc.co.uk)

# You Can Be Serious: Broadcast Archives and Academic Discourse

Richard Wright, Research Engineer, BBC Research and Development, London UK

Audiovisual technology, particularly film, radio and then television are defining features of 20th century life – and of our record of that life.

Broadcasting matters: about 80% of young adults in the USA use broadcasting as a primary source of news, and 30% use newspapers . Americans watch about 38 hours of television per week and Britons are not far behind with 31 hours per week of television, and 22 hours of radio . The total amount of broadcasting output was estimated in 2000 in a report on all forms of information and their storage requirements. The estimate was 100 million hours of unique radio and TV content per year (globally). Not all this material has entered archives, but a surprisingly large amount of material has. Current estimates (based mainly on Presto and TAPE results in Europe) indicate 100 million hours of archived content in Europe and 200 million hours worldwide .

The picture so far is of an activity that:

- provides an engaging, vivid record of the last century
- occupies large parts of the daily life of nearly everyone – not just Europeans and North Americans
- has been collected in huge archives.

These materials should be a meaningful resource: for academic discourse and for pursuit, by anyone, of information on how we live and have lived. However the use of these archives has historically been very limited, by three factors:

- technology: audio and video recordings are expensive to duplicate or publish; before the internet there was no effective means of access, and even in our internet age major problems remain:
  - very little content is online, though the paper will present notable exceptions
  - audiovisual content cannot easily be quoted or even cited, because that requires online content with persistent URIs and ubiquitous broadband access, as well as requiring an agreed, widely-implemented standard for referencing actual time positions within an audiovisual file (time-based references).
- rights: broadcast materials are subject to copyright; there is a ‘fair use’ exclusion for education, but that is (in the UK) for research, only.
- attitude: the one group that could use broadcast materials without copyright restrictions are academics, but the recent JISC, NEH, NSF, SSHRC Digging Into Data project did not list any broadcast corpora, in a collection of 30 corpora that were made available to this major international data-mining project.

The situation is changing. Archives are moving online, there is massive pressure for more access, and technology is supplying the needed tools: PrestoPRIME is working with the relevant W3C committee on time-based reference and annotation. Will this convince academics that broadcasting is respectable, and provides suitable material for serious discourse?

The paper will summarise the changes over the last decade in access to broadcast archives, review the current situation, and list the remaining obstacles – and recent and anticipated progress regarding those obstacles.

Richard Wright, Archive Research  
BBC Research & Development  
R&D South Lab  
BBC Centre House, 1st floor  
56 Wood Lane  
London W12 7SB  
Office: 030304 09653  
Mobile: +44 7802 469 684