Abstract

In this interim report we seek to provide insight into the issues we are addressing and the direction that the work of the group is taking.

1 Introduction

Our diverse cultures rely increasingly on audio and video resources. We need to chart a steady course for the preservation of these resources and to determine the most effective ways to access the rich content embedded there. For example, though our nations possess enormous collections of spoken-word materials, much of these collections will decay or remain inaccessible to the public unless we act to chart a preservation and access path. Our aim is to forge agreement on topics vital to preservation and access so that as technology changes, we will be able to rely on our collections to understand and preserve these vital components of our cultural heritage. We also need to focus research support on areas of preservation and access that we believe will yield the greatest benefits across many intersecting disciplines.

Much of our collective aural heritage from the 20th century, and most TV and radio broadcasts, remains in analog audio format. We confront significant issues in:

- identifying standards appropriate for digital preservation;
- creating open-source, scalable architectures for storage and retrieval of audio;
- developing tools for audio structuring, browsing and search;
- addressing language differences so that non-native speakers may access foreign-language spoken-word resources with relative ease;
- confronting intellectual property concerns.

There have been a variety of projects, in Europe and the United States, concerned with collecting, indexing and searching audio collections. These include: digital archiving projects such as the Northwestern University digital audio archive for U.S. Supreme Court oral arguments and the EU-supported Euromedia project; spoken document retrieval projects such as the THISL and OLIVE projects in Europe and the Informedia project in the US; and evaluation programmes such as the TREC Spoken Document Retrieval track.

Building on this platform, it is now possible to discuss an agenda for future research, standardization and development of best practice in this field. To support this goal the European Network of Excellence for Digital Libraries (DELOS, supported by the European Union Information Society Technologies programme) and the US National Science Foundation (Digital Libraries programme) have collaborated in establishing a working group whose remit covers the creation and dissemination of spoken-word digital audio archives. The field is at an intersection of several areas, and this is reflected by the diversity of the working group members who come from computer science, political science, libraries and archives, broadcasting, engineering, language technology and history. This report is the output of the working group, resulting from two meetings, the first at Northwestern University, Evanston IL, USA (June 2002), the second at TNO, Delft, The Netherlands (September 2002).
2 Use Cases

In this section, we illustrate the broad range of research issues raised by spoken word collections using six use cases. The organizing principle for these use cases is content similarity—in other words, we group potential uses of spoken word collections by the characteristics of their content. We seek to suggest the scope of each use case with illustrative examples that highlight specific user groups and particular types of information needs. We augment this with a brief overview present practice for one or more spoken word collections that are within the scope of the use case. Finally, we draw on these multiple sources of insight to identify two issues that will be important in the next stage of our work: (1) types of contextual metadata that might be important to users of these collections, and (2) research issues that, if addressed, could have a substantial beneficial effect on access to these collections.

It is not our intention in this section to provide a comprehensive set of use cases—our goal is rather to be illustrative. Furthermore, we do not seek here to identify all of the important research issues—that is the goal of the next phase of our work. Rather, our goal here is to identify research issues that arise from the use cases we introduce, in the hope ideas we develop here will help to shape our future discussions.

2.1 News Broadcasts

News broadcasts cover the material broadcast on television and radio to convey information about current events. In broadcasting there is a range of material relating to news, which can be categorised on a dimension of timeliness:

- Bulletins
- Scheduled news
- Daily, weekly, monthly and annual reviews
- Special-subject reviews

In the last ten years there has been an increasing amount of so-called 'continuous news', which is a version of the scheduled news that is updated on something like an hourly basis.

**Bulletins** break into the normal broadcast schedule, to issue brief, urgent reports on items of major significance. The definition of 'major' varies according to the nature of the broadcaster, but in general a bulletin is expected to be of compelling interest to the bulk of the viewers and listeners.

BBC example: death of the Queen Mother

**Scheduled news** usually has a daily focus, with updates on a timescale of hours.

BBC examples: TV news at 1, 6 and 10; Radio 'World at One'

**News reviews** collect the major stories of the period under review, and usually offer greater depth on one or more items than is offered by scheduled news.

BBC Example: Newsnight

**Subject reviews** or current-affairs programmes may cover exactly the same topics as covered in the other types of news, but would stick to a single issue, or linked set of issues.

BBC Example: Panorama on foot-and-mouth outbreak in the UK.

A late-night news review like BBC’s Tonight is a combination of scheduled news and one or more in-depth subject reviews.

2.1.1 Content characteristics

The whole point of news broadcasts is to convey information. Therefore the speech is generally scripted, and clearly spoken—often read from a transcript. However it is only the news presenter’s speech that can be so described. There is also an enormous range of material (actuality) that can be live interviews or
recordings at an incident or event. The non-studio material can be noisy, unscripted and often emotional. In general material used for the news is meant to be of relatively high intelligibility, but if the pictures are interesting or if the situation is compelling irrespective of intelligibility, then the audio component of news broadcasts can in fact be of low intelligibility.

There have been studies in automatic speech recognition of news material, and word recognition rates of 60% to 70% are reported [7, 8, 9]. These rates are inadequate for the production of a sensible transcript, but have been accepted (by the BBC and others) as sufficient for useful information retrieval. The recognition provides a corpus of text which allows users (of a free-text search approach) to find material of interest.

Archiving  There is a well-established tradition within broadcasting for archiving of programme output. This tradition certainly includes news output, as news material is the most used part of broadcast archives\(^1\). The material is used for background information, and also is used in making new programming. Archive material is especially important to news review programmes, but a surprising amount of archive material is in the scheduled news: 30% of the 'actuality' comes from the archive for BBC news, and other broadcasters (with a less extensive global reporting network) usually have higher figures.

Cataloguing  Because of this high level of re-use, news material is quickly and thoroughly documented. Most major broadcasters will catalogue (index) news material within 24 hours. Every item (story) will be indexed, for (at least) names, places and subject matter. The BBC uses an extensive numerical system (for classification), in addition to descriptive text (for cataloguing). Most broadcasters use key words or a thesaurus for classification, and retrieval based on key words is usually augmented by free-text retrieval (against the cataloguing text).

Documentation  Automation has affected news production just as it has affected all media. Major newsrooms now usually run from some form of news production system\(^2\). Such a system is a sort of process-controller plus database, designed to capture and track the elements needed to make a news programme, manage the workflow, and produce at the end a running-order linked to all the elements (audio and video clips, graphics, text to be read by the presenter) of the scheduled programme. The use of news production systems means that, in principle, there is documentation of the running order, and transcripts for the news readers material, for all scheduled news programmes\(^3\).

Rights  The running order documentation provides a form of provenance information, because it should identify the source of each element in the programme. This information is important for rights and payment issues, as a considerable part of major broadcast news is supplied by news agencies such as AP, PA and Reuters or by other broadcasters. Often a contribution from an outside source is only purchased for presentation on one day, or similar limited licence (a temporal, spatial or procedural limit on the item’s use). When agency material is archived and rebroadcast, further payment (to the source) may be required. Both the news production areas and the archives are very sensitive to this issue, and try to maintain adequate documentation and procedures to prevent unlicensed reuse of agency material.

Pre-broadcast news  Before television, the only way to both see and hear about current events was at the cinema, with filmed news (newsreels). It was standard practice (until about 1960 in North American and Western Europe) to show film news programmes as part of a cinema presentation. There are archives of this material (eg Pathe\(^4\)). It would not have the documentation available with current news production, but would share most of the other characteristics of a current weekly-review type news programme. Newsreel material dates from about 1910 until about 1970.

---

\(^1\)At the BBC, News represents approximately 30% of the archive holdings, and gets 40\% documentation on archive usage).

\(^2\)The BBC use ENPS, a product developed by Associated Press: http://www.enps.com/

\(^3\)The same method can be used for news reviews, but such programmes (especially those reviewing a week or longer) may be made by conventional programme-making approaches and not use a formal news production system.

\(^4\)http://www.britishpathe.com/
Radio news broadcasts date from the early 1920’s. The BBC began news broadcasting in 1922. There are archive recordings from this era, because audio was recorded onto gramophone discs (the BBC continued to make such discs for nearly 50 years, before switching entirely to tape).

2.1.2 Physical characteristics

[notes; needs final text]

Physical status and condition of broadcast archives

Radio  PRESTO data; 70% on 1\textsuperscript{4} in tape; 15% digital (CD, DAT); tiny amount online

TV  multiple analogue formats, including film (often with separate sound track that are now the principal source of vinegar syndrome); more like 30% on digital videotape (various formats); some material ‘web ready’ in MPEG I or MPEG II; some material on CD (MPEG I) or DVD (MPEG II) or datatape (MPEG II; RAI, INA, CNN...)

Online  Growing practice of simulcast news, esp radio. BBC has integrated news and has web as ‘third medium’, and provides specially-edited material specifically for the web.

It is common practice for websites to have an online ‘archive’ of previously published material. In the BBC, the broadcast news as presented on the website has been continuously archived since 1997, when the site was launched. The BBC online archive is accessed by a free-text search engine, and includes any associated audio or video+audio.

The archiving web audio-video content is still under development, partly because of the technical complexities. The archiving of websites is now of international significance, as a legal and a research issue. Various nations have passed ‘library of deposit’ legislation governing websites, but the legislation is, to some extent, in advance of the ability of national libraries to implement solutions particularly with regard to archiving of streamed content.

BBC uses RealAudio but prefers to move to a non-proprietary, licence-free format. MP3 a contender, along with MPEG IV, ... (EBU document has quality vs bandwidth ratings for various compressed audio formats).

Amount of online broadcast news:

2.1.3 Representative user groups and tasks

• Internal users / rapid access for re-use (20 minutes)
• Education / area studies
• External sales of stock audio / repurposing
• Financial analysts / alerting
• Librarians / question answering

2.1.4 Examples of current practice

We may consider the daily output of the BBC, and associated documentation.

TV News  The BBC now has a 24-hr news channel (News 24). This does not mean 24 hours per day of unique material, as there is a considerable amount of repetition. In addition, there are two hours per day of scheduled broadcasts, plus a morning programme with brief news updates, and a one-hour news review at night (Newsnight). All these are UK-wide output, and in addition there are regional news programmes and national programmes (covering England, Wales, Scotland and Northern Ireland). In total, the BBC broadcasts about 10 hours per day of scheduled TV news (in addition to News 24).
Radio News  Radio also has main scheduled news three times a day, with regional and national material, and with a 24-hr rolling news station. It has a major news programme for 2 hours every morning (Today), plus a review of parliament (when in session). In addition the World Service provides international broadcasting, including scheduled news programmes, in 44 languages. For many of the World Service programmes, the core content of the programmes is at least similar, if not identical.

ENPS4 is use for news production in TV and radio, including in the regions and within the World Service. Today in the BBC, ENPS has more than 12,000 users and is accessible on about 6,500 devices in 300 different locations.

Although the BBC has 13 regional TV and 39 regional radio operations, the archiving for regional material (and for the World Service) has not been comprehensive. The introduction of server-based playout systems will make the logistical side (the media-handling side) of archiving much easier. The main BBC archive has the UK-wide TV and radio output, going back to 1922 for radio and 1936 for television. Approximately 250,000 hours of TV and radio news material has been archived. The level and quality of the cataloguing has varied over the years.

Online News  At present (8 August 2002), BBC Online News provides two video+audio (BBC One, BBC News 24) and two audio-only 'latest news' summaries (world and national), plus specific reports on 14 topics (11 are video + audio), plus simulcasts of the three main daily scheduled news programs on TV and radio. The website gives a short paragraph on each topic, but NOT a transcript.

Example  Scan delays 'put patients at risk'
Almost half a million people are forced to wait months for important scans at NHS hospitals across England and Wales, a report reveals.

However most of the topics with audio or video+audio also have full articles on the website: (see Appendix 1).

ENPS  A sample screen is shown in figure 1. The running order is on the left side, and a portion of a script is in the upper-right window. In this implementation, there is also a built-in audio player. The key issue regarding documentation is to get running order and script information into the archive, properly associated with the recording of the actual news programme (as broadcast). In the BBC, this level of metadata integration has not yet been achieved.

The highest level of fully-automated integration of news production and archiving is probably at CNN, where full production information is collected as an 'auto-archive' process, with minimum manual classification and cataloguing.

INFAX  INFAX is the BBC broadcast archive catalogue, used for all programme, not just news. Figure 2 shows an example of two reports from a Newsnight programme.

2.1.5  Contextualizing information
Major broadcasters archive their news programmes, and the archives usually provide formal documentation (classification and cataloguing). This documentation provides professional-level contextual information. In addition, the subject matter is usually the prominent issues of the day, though of course minor or filler or odd and entertaining items are often included in news programmes. For the major items, newspapers and other contemporary current-events publications provide additional context.

2.1.6  Research issues
- Well researched baseline for comparison with other use cases
- Multilingual content might require cross-language search and use
- Exploiting editorial process (capture ancillary materials or reverse engineer)
Tone and/or style may be important in some applications

NewsXML may provide a framework that we can build on

The TDT collection is a useful research collection (English/Chinese/Arabic)

Commercial indexing services provide training data for topic classification

Topic granularity may be application dependent

Multisource summarization (cross-source agreement, development over time)

Item-level access including content identification and tracking

User commentary

Reference / quotation mechanism

Timecode

2.1.7 Appendix: Example of report on BBC Online which has a full text story in addition to a Real Audio video+audio version

Example: Thursday, 8 August, 2002, 07:03 GMT 08:03 UK

Scan delays 'put patients at risk'

Almost half a million people are being forced to wait for important scans at NHS hospitals. A report by the independent watchdog the Audit Commission shows huge delays in radiology services in hospitals across England and Wales. It blames shortages in equipment and staff, and inefficient use of machines. And it warns the delays are causing a “bottleneck” in the NHS and are preventing many patients from getting the medical care they need. There may be patients with diseases who if they were diagnosed earlier could have a better chance of survival

Doctors believe the long waits could be putting the lives of some patients at risk. About 500,000 are waiting for some form of radiology, which may mean a straightforward X-ray within hours. But the report also reveals many patients in need of MRI and CT scans - which can detect cancer and heart defects - are waiting months to be seen. It found that 110,000 patients were waiting for MRI scans in March 2001. While the average wait was five months, many patients were waiting almost nine months to have the test. Another
Kirsty WARK intros item as Nato claims it has more evidence of Serb atrocities. Tonight programme shows new pictures of murders from inside Kosovo. The village Klina was attacked on 21/5/98 by Serbs. Matt FREI reports.

> sp s vt, amateur video, graphics, stills

INTERVIEWS: Xhevdet BUJUPI (Refugee) re saw a dead woman and baby who'd been mutilated (3m15s-4m07s). Fatime BAHRAINE (Refugee) re we were sent here, my daughters were killed (5m06s-44s); they threatened to run us over w tanks (5m55s-6m08s). Clint WILLIAMSON (Legal Officer, War Crimes Tribunal, Yugoslavia) re everyone is trying to react to the situation, we'll stay long enough to gather the evidence needed. (9m55s-10m23s)

VOX POPS: Refugees - we were kicked out of our houses (4m43s-5m06s); smuggled video out by hiding it on my mother (8m22s-9m09s)

INDEPENDENT ACTY: var s of Albanian refugees; incl babies and old people; inside large room of building in Tirana (1m40s-2m25s) (4m07s-17s) (5m44s-55s) (6m40s-57) (7m 26s-50s) (10m23s-40s).

AMATEUR VIDEO: var s of bombing attack on Klina 21/5/98; incl civilian dead bodies on ground (2m25s-3m15s); ls convoy of trucks w soldiers in leaving Klina; burning buildings in bg (6m08s-40s); wailing villagers Klina; 25/5/98 (4m17s-43s); var s of 6 dead bodies wrapped in blankets, w relatives mourning over them (7m50s-8m22s) (9m09s-55s).
128,000 were waiting an average of two months for a general ultrasound. But at some hospitals, patients wait as long as three months. There were similar long waits for colon tests, gynaecological ultrasounds and CT scans.

Generally, doctors prioritise those patients who are suspected of having life-threatening diseases like cancer or brain tumours and these are seen more quickly.

Risk to patients

Those affected by these long delays include those with hip and other orthopaedic problems and those with neurological symptoms like severe headaches. But doctors acknowledged that some patients with serious conditions may be affected.

Professor Helen Carty, a senior member of the Royal College of Radiologists, told BBC News Online: “There may be patients with diseases who if they were diagnosed earlier could have a better chance of survival.” Prof Carty, who works as a consultant radiologist at Liverpool’s Alder Hey Children’s Hospital, said the report was a fair summary of what is going on.

Peter Wilkinson of the Audit Commission told BBC Radio 4’s Today programme the delays extended the pain for those in a less serious position. The report revealed 41 equipment used to carry out scans on the NHS is out-of-date. It also acknowledged a huge shortage of staff was contributing to many of the delays.

Some radiology services are creating a bottleneck in the system and delaying patient care. There is an estimated national shortage of 5,000 radiographers while almost one in seven radiologist posts are vacant. The commission called on the Department of Health to draw up a national strategy to tackle these staff shortages.

However, the report also found some hospitals use their equipment much more effectively than others. For example, some hospitals carry out 4,000 MRI scans each year while others complete just 2,000. The commission urged NHS trusts to review the way they used the equipment to ensure they were being as productive as possible.

The controller of the Audit Commission, Sir Andrew Foster, said the findings highlighted the need for hospitals to learn from each other. “There are concerns that some radiology services are creating a bottleneck in the system and delaying patient care,” he said.

Investment

The Department of Health said the situation had improved since the Audit Commission’s figures were compiled. A spokeswoman said: “Over 100m has already been provided for new equipment. In the 16 months since the Audit Commission data was collected there are 80 new CT scanners and 21 new MRI scanners in use in the NHS.” She added that a “major new initiative” to expand the radiology workforce is underway with extra training places and posts.

Mike Richards, the government’s ‘cancer czar’, told Today the issues were being addressed as a “high priority”.

2.2 Lectures

2.2.1 Content characteristics

The category “Lecture” encompasses a number of speech acts, all which share some important similarities. The genre itself is best defined across two distinct dimensions: 1) language, and 2) speech. Most of the content in question is planned, often written, and delivered over a consistent acoustic channel. The language and speech content of lectures has a relatively narrow scope of characteristic features, while the intellectual content varies depending on the topic and target audience.

**Language Content: Syntax** The syntax of lectures is predominantly declarative. It is also likely to have a large number of relative clauses and coordination structures. The sentences are usually complete, with full subject-predicate agreement, complex modifier structures, and little syntactic ambiguity. This is as important to human perception as it is to any automated NLP (Natural Language Processing) parsing strategies. The relative variety of syntactic structures used in a lecture is rather limited and likely to be in close proximity to the standard written variety of a given language. Thus, for example, we should expect a lecture delivered in Norway, to be written and presented in Bokmal or Nynorsk ((the written standards), rather than any of the more common spoken varieties. By the same token, we might expect a planned speech
addressed to a North American audience to contain syntactic structures characteristic of upper-middle-class white speech, rather than those from the repertoire of, say, working-class Hispanics. There are, of course, exceptions, such as sermons and certain types of political speeches where non-standard syntax is used for language solidarity and identity purposes.

**Language Content: Phonology**  Defining the phonology and phonetics of lectures is important, particularly for the purposes of speech recognition, which is becoming the preferred method out transcribing large volumes of recorded speech. Lectures are most typically characterized by carefully articulated segmental (individual sounds) and suprasegmental (melody, rhythm, intonation) features. Individual syllable nuclei often reach their articulatory targets, which decreases the level of potential ambiguity and minimizes the strain of human and machine comprehension. There is relatively little reduced syllable deletion (contractions such as “shoulda” for “should have”) and final consonant deletion (e.g., “he asset me” for “he asked me”). At the phrase level (in English), there is usually one primary phrases stress with an accompanying falling intonation contour. It is quite common for articulation and phonation to be slightly exaggerated and free from casual speech features such as creaky voice.

**Language Content: Lexicon and Semantics**  Due to the formal nature of lectures, words are chosen primarily from the formal, standard dialect lexicon. Ambiguity is skillfully avoided. There is relatively little use of metaphoric and figurative expressions. Slang is virtually non-existent, unless it is used for special effect. Political correctness often dictates the choice of particular words and expressions. Talkers are careful to avoid loaded words such as “retarded” in favor of politically correct terms such as “person with learning difficulties.” The most controversial semantic fields include race, sex, disability, and religion. Many planned lectures are written to avoid repeating words, which, in turn, results in thorough use of synonyms and antonyms. Any well-constructed speech recognition language model will have to take that into account.

**Language Content: Pragmatics**  In terms of pragmatics, planned speeches are very monochromatic. They are, most typically, monologues (with the exception of Q&A) and Grecian maxims are seldom violated. There is an underlying sense of full cooperation between the speaker and the audience. In is in the speaker’s best interest to adjust their pragmatics to avoid unwanted ambiguity, emotionality, and communication break-down. Some of the more outspoken presenters (such as Martin Luther King, Jr.) have been known to use a variety of pragmatic strategies to raise the audience’s active comprehension of the speech. Techniques such as repetition and call and response are frequently found in Southern as well as African American sermons and political speeches.

**Speech Characteristics: Acoustic Phonetics**  Lectures are usually characterized by rich and consistent speech-relevant acoustic information in the signal. The microphone is usually placed at a constant distance from the speaker’s mouth, which results in consistent intensity levels. This is of benefit both to human and machine recognition. It also minimizes the amount of necessary digital signal processing (DSP) to prepare audio files for delivery. The dynamic range of such recordings is usually good enough to make aggressive use of a compressor/limiter and/or gate/expander unnecessary. Frequency formants (frequency range from 300 to 10000 Hz) are most typically strong, as is signal-to-noise ration (SNR).

**Speech Characteristics: Channel Characteristics**  Most lecture recordings have a wide dynamic range, broad frequency response, and high SNR. There are, of course, exceptions, but they are most commonly due to recording techniques used rather than the nature of lectures per se. The consistency of the acoustic channel is crucial to successful machine recognition. Speeches are, therefore, among the first types of speech data to be fed into large vocabulary recognizers.

### 2.2.2 Representative user groups and tasks

- Education / distance education (MIT Open Knowledge Initiative?)
2.2.3 Examples of current practice

HistoricalVoices.org has had a fairly long experience dealing with lecture/speeches recordings.

**Presidential speeches** HistoricalVoices.org has a vast collection of presidential speeches ranging from William Taft to Ronald Reagan. MATRIX (http://MATRIX.msu.edu) has used a custom audio digitization methodology and a suite of scriptable applications to convert analog recordings into a 96,000Hz/24-bit PCM data stream. We have been storing digital audio master files in the Broadcast Wave format on CDROM (2 copies), while the downsampled (22,050Hz/16-bit) production copies have been saved on spinning disk. The master file has no added DSP, while the production copy, most typically, undergoes a custom DSP process consisting of sample rate and bit-depth conversion, 2:1 compression with a soft-knee limiter, and volume normalization across the entire batch. It has also been common to created short, edited clips suited to particular educational or scholarly needs. MATRIX categorizes such clips as “derivatives” and applies aforementioned DSP to them as well.

**Audio Metadata** MATRIX has adopted the METS standard for all of its metadata purposes. While it is important to provide detailed descriptive metadata, MATRIX pays close attention to the administrative section of METS. Most notably, we describe audio (and video) material in ways that ascertain platform, machine, and software-independent playback of digital audio files. In theory, one can easily build an audio playback application to play our files, based purely on their technical metadata description. We believe that this crucial for the purposes of long-term preservation of audio material.

**Transcripts and speech recognition** We strongly believe that having transcripts is an integral part of storing and serving audio (speech) material. A time-aligned multimedia corpus puts a different perspective on primary sources. It allows a computer-assisted, quantitative type of modern scholarship. It is also a very useful pedagogical tool, as well as a good way to make aural resources available for the hearing-impaired. To date, MATRIX has used hand-transcription as well as machine transcription. Despite its obvious advantages, hand transcription is impractical for large bodies of audio material. MATRIX has collaborated with The Language Processing Lab at the University of Colorado to develop a large vocabulary recognizer to produce so-called “dirty transcripts”, which are accurate enough to provide rich searchable material, but not quite accurate enough for public display.

2.2.4 Contextualizing information

2.2.5 Research issues

**Search and metadata** One of our current R&D issues is improving search quality and efficiency by using comprehensive metadata and machine-generated transcripts. Our experience so far (e.g., http://www.historicalvoices.org/flint) has shown that searching through transcripts dramatically improves search accuracy. Time-alignment (by means of XML) allows users to view scrolling transcripts as well as give them the ability to stream only the portion of the clip they’re interested in at any given time. This user-centric approach to streaming audio is one that MATRIX has a very strong interest in pursuing.

- TED collection (Eurospeak 93) is available from ELDA/LDC
- Rhetorical styles offer possibilities for structuring access
2.3 Meetings

Meetings are at the heart of virtually all organizations. They may range from an informal group of 3 or 4 people, to a large, formal gathering in an auditorium. The participants may be face-to-face or at remote locations interacting using teleconferencing software.

Automatic recording and processing of meetings is closely linked to the development of smart rooms, equipped with multimodal sensors — participants may be equipped with close-talking or lapel microphones, together with a microphone array in the room.

2.3.1 Content characteristics

Meetings have several participants (we assume 3 or more, to distinguish from a dialog) and are characterized by verbal interaction between participants (as distinct from lectures, which primarily consist of a single speaker). The interactive nature of meetings is demonstrated by the large amount of overlap between speakers: in the ICSI meetings corpus [?], over 30% of some meetings consisted of 2 or more overlapping speakers.

The speech styles of meeting participants tend to be relatively spontaneous, or unplanned, although this is clearly dependent on the nature of the meeting. Some meetings may have significant components of planned speech (eg talks with visual aids, prepared reports), but the interactive parts of the most formal meetings feature phenomena associated with spontaneous speech, such as filled pauses, incomplete utterances, false starts and backchannels. This is in contrast to a lecture, in which interactive question and answer sessions may be formal (much less overlap) and form a small part of the overall lecture.

A typical meeting is characterized by challenging acoustic conditions, with multiple simultaneous speakers, room reverberation effects, and unpredictable background noise. Existing meeting room recordings have attempted to lessen the problem through the use of wireless close-talking microphones [?] or lapel microphones [?], usually with additional far field microphones. However, physically attaching microphones to meeting participants is less desirable (from an end-user view) than the use of less obtrusive far-field microphones, and future work in the area may use microphone arrays [?] or binaural recordings. Even when close-talking microphones are used, to obtain hopefully unambiguous ground truth, there are still problems of cross-talk (a speaker’s close-talking microphone picking up another speaker), background noise, and speakers without a microphone (eg someone entering mid-meeting). These problems are, of course, worse with lapel microphones. In addition to these audio challenges, meetings pose speech recognition modelling problems that arise from multiple simultaneous speech (separation and localization), spontaneity (an ongoing challenge for speech recognition) and reverb - a significant problem in some environments not designed for recording. Other speech recognition issues that arise from the recorded meetings include the use of specific vocabulary and language models, arising from the particular domain or topic.

A spoken message is richer than its text transcription. The prosody of the message—its timing, intonation and energy— can disambiguate a text transcription (eg question or statement), as well as adding new information (eg the speaker’s emotional state). These effects are important in meetings, as they are often used for things such as turn-taking, interruption, agreement and frustration. Indeed, in many meetings—particularly informal ones—much of the information is spread across non-lexical channel. For example, an analysis of how a decision or compromise was reached may be rather incomplete if only the text transcript is available (although a human reader may be able to use information from the lexical patterns to infer some of the non-lexical components).

Finally, meetings are ubiquitous and there is overlap with the other five uses cases we have considered. Examples include: panel discussions on news broadcasts; extended question and answer or discussion sessions at the end of a lecture; government committee meetings; meetings that form a collective oral narrative (eg the ICSI corpus includes a meeting of several eminent computing pioneers from SRI International, which could be classified as oral history); telephone conference calls.

2.3.2 Representative user groups and tasks

Meetings occur in many different contexts, so there is a wide variety of potential user groups. However, many of these user groups may share the same underlying tasks. A generic application in this area the
recording of a set of meetings, generation of a textual transcript, probably with added markup relating to
tings such as speaker identity, prosody and environmental events (e.g., a door slam). Given the provision
of a set of recorded meetings and their rich transcripts, required tasks are likely to include structuring,
browsing and querying.

Currently, many examples of information access from spoken audio rely on a textual metaphor, in
which text-processing techniques (e.g., text retrieval, coreference resolution, or named entity extraction) are
applied to the word-level transcript of the audio, typically generated by a speech recognizer. Current
state-of-the-art spoken document retrieval systems [?] provide a good example of this, and offer a good
performance on domains such as TV and radio news broadcasts (compared with text retrieval from the
equivalent transcripts). If additional non-lexical information is considered, then not only can it add richness
to information access tasks arising from text processing (e.g., using prosodic information in the ranking
of relevant clips of a meeting returned from a query), but it may also help to define new tasks such as
developing an “emotion map” of a meeting, analyzing relationships between participants, and tracking the
evolution of an agreement.

Public access to government Opening up of democratic procedures by not only broadcasting meetings
of committees, etc., but through the provision of meeting browsers and flexible querying of a
archive of meetings, to enable citizens to track issues of interest and the progress through legislature, and also to find
related issues.

Business / shareholder relations Recordings of meetings of public companies made available to share-
holders.

Reconstruction of meeting acoustic scenes Recordings of meetings (perhaps using binaural micro-
phones, or using signal processing on multiple far-field microphones) can make it possible for a future
listener to obtain a similar acoustic experience to a participant in the meeting. This offers the intriguing
possibility of placing the listener in: the acoustic space of the British House of Commons during heated
debate; the place of the US president during a crisis meeting, receiving advice from all quarters; the role of
a celebrity being interviewed at a large press conference.

Meeting facilitation training An archive of representative meetings, together with appropriate browsing
and playback software, with marking of key segments, could form the basis of a training programme
for meeting facilitators. This might be particularly useful well combined the audio scene reconstruction
referred to above.

Ethnography Disciplines such as ethnography, linguistics and cultural history could be transformed by
the potentially huge match of empirical data that could be generated by the routine recording of formal
and informal meetings. With appropriate processing and analysis tools, it would be possible to empirically
investigate questions such as how people interact, how decisions are reached, and so on in a day-to-day
setting.

2.3.3 Examples of current practice

Most examples of current practice are research projects:

- ICSI/UW/Columbia meeting recorder project http://www.icsi.berkeley.edu/
- CMU ISL meetings project http
- EU IST M4 (Multimodal Meeting Manager) project http://www.dcs.shef.ac.uk/spandh/projects/m4
  Aims to develop a smart meeting room, equipped with multimodal sensors, for the recording of
  meetings with 3–12 participants, together with facilities to structure, browse and query the recorded
  meetings,
2.3.4 Contextualizing information

- Purpose of the meeting
- One-off or part of a series? Regular or irregular?
- Level of formality
- Aiming to accomplish a particular task?
- Speaker details: age, education, native speaker of meeting language, dialect information, ...
- Within an organization or multi-organization?
- Meeting agenda
- Minutes of previous meeting
- Discussion papers, reports, etc. supporting parts of the meeting

2.3.5 Research issues

Audio

- Cross talk detection
- Speaker overlaps
- Speech recognition from far field mics (including microphone array and binaural approaches)
- Speaker localization
- Recognition of informal, spontaneous speech
- Dealing with room reverberation

Information access

- Decision points and tracking the evolution of a decision or agreement
- Roles of participants
- Interaction maps
- Retrieval of meeting segments relevant to a topic or query
- Topic detection and tracking
- Summary of a particular meeting
- Summary of previous discussions about a topic
- Summary of an individual's contributions to a topic
- Emotion maps
Privacy and rights issues

- Who owns the content?
- Does meeting recording pose privacy issues for members of an organization (e.g., quantitative analysis of employees’ behaviour in meetings used by employers)
- How much is the behaviour of meeting participants constrained by the knowledge that they are being recorded (cf. whether Usenet posters’ behaviour has been modified by Google groups...)

2.4 Government

‘Government’ encompasses activities of legislatures, parliaments, city and town councils; administrative hearings; civil and criminal legal proceedings; local school board activities and other events. In general, citizens have a right to know what public officials do and to observe them in the course of their work especially when official decisions force tough choices. In many examples, especially involving elected officials, broadcasting of government activities has become routine. Real time broadcasting remains accessible only a small portion of the public, however. The challenging research issues focus on archiving, searching and retrieving such content.

2.4.1 Content characteristics

This category comes closest to ‘government at work.’ It is characterized by spontaneous and prepared speech. While the content may be planned and constrained (for example, discussion of a new sewer system considered by the Birmingham city council), the presentation is rarely fixed by a prepared text. Frequently, there are multiple participants at a single event, locations vary enormously in their acoustical features and audio capture can vary from none to primitive to professional. Speech is by turns though there can be interruptions and overlaps. Proceedings follow accepted structure with time limits more or less followed by participants. Today, a growing number of government units rely on audio and/or video recording either to replace the transcript of proceedings or to provide an oral transcript. Some American state government units have 50-year-old audio archives. Transcripts remain the most useful way to narrow the search for audio materials.

The syntax of public government proceedings covers a wide range from formal presentations (for example, a speech by a legislator to accompany the introduction of legislation in the U.S. Congress) to on-the-spot debates with quick question-and-answer exchange (for example, Prime Minister’s Question Time in the British House of Commons). Oftentimes, speakers take turns though participation by others (for example, audience participation in the form of cheering or booing) is not out of the question.

This category of speech covers a wide range of syntactical structures. Speech will reflect expected institutional norms for elected officials. But other participants will not be constrained in their speech patterns to elevate their grammar or pronunciation unless prompted to do so.

Phonology also covers a wide range depending on the type of government proceeding. Court trials may call for testimony from non-native speakers or require translators to articulate from non-native to native languages. Again, designated government officials usually reflect a more uniform and comprehensible form of speech.

2.4.2 Representative user groups and tasks

Open government is a public good. Televised proceedings of national legislatures in US and EU assure a measure of wide public access. However, observing government in action typically conflicts with work, family, education, and leisure.

Average citizens may wish to access the decisions and reasons of their government at all levels. The school board, for example, may have debated a new policy regarding student expulsions. The local news may not provide detailed reporting on the issue or the issue may be reported episodically but not immediately. There is no easy way for citizens on their own to track down and re-create the events from that meeting although the meeting may have been recorded (audio or video) and transcribed. High courts in the
US and EU reach decisions that affect the lives of the citizens they serve. Some courts announce their decisions orally in proceedings open to the public as well as in writing. When judges in their official capacity speak to citizens, there should be a way for any citizen to ‘hear’ what has been said.

Elected officials have a special interest in the evidence or quasi-evidence of their participation in official activities. A copy of an official’s remarks is a demonstration that he is doing something. (In the United States, a member of Congress may request the introduction of remarks to the legislative record even when he was not there to deliver them.) It might prove more compelling still for officials to post their activities so that constituents could ‘experience’ them.

Bureaucrats, lawyers and judges have an interest in remarks from legislative debates. They provide evidence of ‘legislative intent’, which is a way to examine the meaning of laws, whose studied ambiguity in drafting must give way to the creation of bright lines for decision-makers.

Legal professionals (judges and lawyers) and the public are typical consumers in legal settings. Open justice systems provide for citizens’ physical presence to observe proceedings. Increasingly, court systems have allowed cameras and microphones in courtrooms permitting unlimited public access without the requirement of public attendance. Audio and/or video recordings of depositions in civil cases provide more efficient review of testimony during the discovery phase of civil litigation. These depositions become part of the public record.

The creation of a text record of court proceedings has been a long tradition stretching back hundreds of years, with court reporters serving as official stenographers for courtroom events. When created in real-time, stenography notes serve as the official record of events.

Today, American appellate courts create versions of oral proceedings for the purpose of their internal review. Some courts provide access to these audio materials. The highest courts state and federal tape and transcribe oral arguments. The transcripts may be used as a reference for judges and their staffs in the course of resolving the issues that come before them.

Educators are spoken-word government proceedings. At an elementary level, this may simply be a way to provide examples for civics instruction. At advanced levels, the ability to retrieve and deploy government decisions as they were contemplated and announced provides additional nuance. The choice of language, emphasis, use of metaphor and delivery bear on the study of rhetoric. In addition, legal education in the United States has been dominated by the study of the decisions of appellate courts. Though transcripts of public appellate proceedings are available, that evidence is rarely put to use for instruction. Specialized courses in trial or appellate advocacy could probably profit from easy access and distribution of court proceedings.

Historians seek access to government-in-action to study and document individuals and historical periods. Reliance on transcriptions of events is at one remove from the events themselves. Narration of the past relies in part on accurate representation of the past.

2.4.3 Examples of current practice

- British Columbia produces transcripts of its House debates substantially verbatim. Digital recordings are transcribed and edited. Only the paper version is used as the official record. (www.legis.gov.bc.ca/hansard/)

- FedNet (www.fednet.net) is the leading provider of Washington based, real-time and archival multimedia information. Headquartered on Capitol Hill, FedNet provides clients the ability to capture, webcast, archive, search, retrieve and bundle events and information related to the client’s special interests. FedNet provides news production services and real-time, live webcasts of:
  - Congressional Hearings
  - Floor Debates
  - White House briefings
  - Regulatory hearings
  - Press conferences

- CapitolHearings.org - a new public service from C-SPAN - lets you to listen to Senate hearings LIVE online. CapitolHearings.org distributes gavel-to-gavel audio coverage of U.S. Senate Committee
hearings. The Senate produces the audio feeds, and C-SPAN encodes and streams these hearings via (www.capitolhearings.org)

- Scottish Parliament has extensive audio and video archive. (www.scottishparliamentlive.com/archive_historic.asp)
- Europe by Satellite offers some live feeds. No evidence of archiving. (europa.eu.int/comm/ebsservices_en.html)
- British Parliament is available in webcasting and broadcasting modes (www.parliamentlive.tv)
- Audio live streaming and archiving in Missouri on a fee basis. (www.legislature.com)
- Administrative proceedings recorded in some states like TX which provide streaming and downloading. Some archiving. (www.tdi.state.tx.us/commish/audio2.html)
- India’s legislative bodies Rajya Sabha or Council of States and Lok Sabha or House of the People offer live audio feeds on an experimental basis. Note the disclaimer: Disclaimer:- The live audio transmission cannot be recorded, reproduced or quoted in anyway. The authoritative record of the proceeding of the Rajya Sabha is the 'Official Report' published under the authority of Chairman Rajya Sabha. Printed copies can be obtained on payment from the Sales and Archives Section of the Rajya Sabha Secretariat. (alfa.nic.in/)
- Oregon retains audio tapes of its legislative proceedings. (arcweb.sos.state.or.us/legislative/legislativelegal.html) including committee proceedings back to 1957
- The State of Florida offers real-time streaming of oral arguments from its highest court and maintains an archive of past arguments, litigants’ briefs and closed-captioning transcripts. (www.wfsu.org/gavel2gavel/archives.html)

2.4.4 Contextualizing information

Perhaps the most obvious contextual information concerns the identity of the speakers who rarely employ self-introductions. For government officials, it would help to know some additional details (office held, last elected/appointed, general descriptors (age, sex, gender)). The subject of the event also provide valuable context to the listener/viewer. Is this live-streamed event a question regarding the trash contract with a local hauler or the plumbing needs for the local recreation center? Extended listening probably clarifies, but it would be helpful to know something of the context at the start.

Many public government events are recorded either by the government units themselves or by private companies. In the past, the recording was accomplished stenographically. The result was a transcription of the events with the audio materials only secondary to the 'official transcript of proceedings.' These transcripts attempt to capture what has been said. In some cases, government officials are permitted to revise these transcripts to improve accuracy and to avoid possible embarrassment.

2.4.5 Research issues

Thousands of hours of government-in-action audio and video are now webcast whenever legislatures, parliaments, administrative boards and courts are in session. Some institutions have substantial analog archives but fewer still have well-thought-through digital archiving plans. With so many rivers of digital content converging into a swelling sea of information, the greatest challenge is to develop fast and efficient search strategies relying on participants, subject matter and content.

Substantive content searching based on speech-to-text transformation seems within the realm of possibilities. Capturing ambient sound characteristics and including them in the metadata may further assist search strategies. For example, knowing that the event in question is public, not private, could be captured from the background characteristics of the sound source. Another novel direction involves the search for emotive content or speaker affect.

Perhaps the greatest research challenge is the management of oceans of sound. Visualization of spoken-word collections might prove a novel direction. In this case, would life come to imitate art?
2.5 Oral Narrative

Oral narrative encompasses a broad range of activities that share a common theme of extemporaneous storytelling. Some disciplines make extensive use of recorded oral narratives, including oral history, personal history, and folklore. In other cases, oral narratives might be one small piece of a larger puzzle. For example, linguists creating spoken language corpora may include examples of oral narratives along with other types of interactions, and survey researchers may sometimes invite the creation of oral narrative in response to a few of the questions that they ask. Humans are by nature storytellers, so oral narrative has played an important role in the preservation and celebration of our cultural heritage since the dawn of human civilization. Improvements in our ability to collect, preserve and access these narratives are therefore likely to have important consequences, both for research use of these materials, and for the direct benefits that will likely accrue to the broader societies in which we live.

2.5.1 Content characteristics

An important characteristic of the oral narratives that we have chosen to focus on is that they are typically intended to be retrospective. One consequence of this is that other descriptions of the same (or similar) events may be available. This may provide technical benefits for some components of a speech retrieval system (e.g., suitable data for training speech recognition language models may be available). Another important implication is that users of such collections might benefit from the additional context that related information from other sources might provide. For example, a narrative describing a mountain climbing expedition might be rendered more understandable if photographs from the expedition could be displayed in a manner that is synchronized with the narrative.

A second characteristic of oral narratives is that they necessarily reflect people’s perceptions of the world, and that perceptions may differ. Different people may perceive the same event differently, and the same person may recall an event differently at different times. Support for comparative analysis (e.g., “more like this” searches) will therefore be of particular consequence to many users of oral narratives.

A third characteristic of oral narratives is that their production is a creative act. People typically tailor their delivery to suit their audience, so we expect that it will often be important that the user understand the purpose for which the story was told. Some people are gifted than others as storytellers, and in some cases the searcher may value the affective reaction that a story creates in a listener fully as much as they value its informational content (e.g., a documentary film maker may seek to establish an empathic reaction). Recommender systems, which allow people to share their assessments (or evidence of their assessments), offer one way to support search based on affective reaction, and research on this topic may produce others.

Perhaps the most striking feature of oral narrative is its diversity. There are presently well over 5 billion potential producers of oral narrative on the planet, and recording technology that is both inexpensive and easily portable has been available since the introduction of the cassette tape recorder forty years ago. Together, these factors have two important consequences: (1) a lot of oral narratives have been collected, and (2) nobody knows where they all are! A few libraries, archives and other organizations have assembled relatively large well-organized collections, but the vast majority of oral narratives exist in shoeboxes, in filing cabinets, and, increasingly, on hard drives that are not actively managed and for which no effective discovery mechanism presently exists—the world’s first Web crawler for spoken word collections has yet to be invented.

As storytelling, oral narratives have a great deal of useful structure. Temporal ordering is typically dominant, with the story most often being told in the order that events occurred. There are, of course, digressions, but extended digressions are themselves normally presented in order. Narratives presented in an interview format additionally offer a dialog structure that can be leveraged to help searchers identify distinct parts of the narrative (e.g., if the interviewer asked questions from a script that is accessible to the searcher). But diversity again poses a challenge—the great diversity of oral narratives results in a great diversity of structures. So there will likely be only a few broadly applicable ways of using the structure

---

5Certainly there are cases in which people will wish to describe their present condition and perceptions, but we would not consider that to be storytelling. There will also be cases in which stories are told that are entirely fabricated (e.g., a comedy act, or a contest in which the participants seek to tell the most outrageous “tall tale”). But we choose to think of oral narrative as focusing on perceptions of the real world, and therefore exclude that sort of narrative from this use case.
inherent in oral narratives, but perhaps a possibility of creating some collection-specific techniques that may have a substantial impact in a narrower range of conditions.

Developing technology to improve access to oral narratives also raises important ethical issues that have implications for the technology. Protecting the privacy of the speakers, or of people named by the speakers, may be of paramount importance in some situations (e.g., if immediate public distribution might put the life of the speaker in jeopardy). In other cases, access to the information content of a narrative may need to be restricted for a prescribed period for policy reasons (e.g., an interview in which future plans for monetary policy are addressed). The definition and enforcement of access rights are therefore likely to be critical factors for some types of oral narratives. Rights management will likely be a particular challenge for existing collections. Once appropriate access policy frameworks are developed, appropriate access rights can be obtained and specified at the time that materials are created. But for existing materials, the answers to some of these questions may remain unclear. For example, if an archive acquired rights to distribute certain materials to researchers in 1980, do they have the right to distribute that material on the Internet in 2002? These are questions that can likely only be addressed by evolving the legal framework of rights management to accommodate the capabilities of new technologies, a process in which the law and the development of technology must naturally iterate.

For most types of oral narrative, the spoken word is the authoritative version. Oral history is often an exception—in oral history, the spoken word is indeed the original version, but when an edited transcript exists, that transcript is typically considered to be the authoritative version. Authority is a social construct, and the practice of vesting more authority in a transcript than in what was actually said emerged for two reasons. First, transcription facilitates access, but finding the spoken version of a transcribed statement is not an easy process using shuttle search (fast forward, play and rewind) on a typical tape player. Citing an authoritative transcript therefore facilitates the use of oral history materials. It is now common practice to offer the person being interviewed an opportunity to review the transcript once it is ready, and to alter the content in ways that better express what they wished to convey. Indeed, sometimes this process is a condition for granting an interview—in such cases it is not uncommon for the tapes to be intentionally destroyed once the transcript is completed. This, then, is the second reason—when an interviewee will edit a transcript, the original tape is treated as a draft rather than as a fixed version of the interview. As the technology for access to original recordings improves, it will be interesting to see how professional practice in this regard evolves.

Spoken word materials such as political speeches, recordings of meetings, or recordings of telephone conversations are likely to be used by researchers for purposes that their creators did not anticipate. Some types of oral narratives are, however, themselves a product of research. Oral history, for example, often results from biographical research. Folklore collections and linguistic corpora are other examples of oral narratives that are typically created to serve a specific need that is anticipated at the time of their creation. An ability to anticipate future use offers an opportunity to better support future users. Appropriate standards for creator-assigned metadata will likely be of particular value for this purpose.

2.5.2 Representative user groups and tasks

Oral narratives can be used by many types of users, similar users may use the materials for different purposes, and the same user may make use of more than one type of oral narratives. In this section we present scenarios that are designed offer some flavor for the breadth of users and uses that we foresee.

Museum Curator / Simulated Conversations When oral history is used in a museum setting, a kiosk format is often adopted. For example, at the Kennedy Space Center, visitors select from a list of frequently asked questions (e.g., “What was it like to walk on the moon?”), and an excerpt that addresses the selected topic from a videotaped interview is then presented. Control is also sometimes provided through spatial interaction. In the U.S. Holocaust Memorial Museum, for example, visitors hear recordings of different people as they move through a room. Either way, the curator first selects the topics and chooses the excerpts, and the visitor then chooses from among the options presented. As oral narrative collections grow larger, it will become increasingly important to provide museum curators with support for identifying segments that may be of particular interest to visitors. Ultimately, it may become possible to give visitors even more control, allowing them to pose previously unanticipated questions for which answers might exist
somewhere in the collection. With advances in speech recognition technology, this might ultimately be implemented in the form of a simulated conversation, with visitors speaking their question and the system then selecting an appropriate “reply” from a collection of oral narratives. Present simulated conversation systems of the type used in computer-assisted language learning rely on extensive (and often domain-specific) language engineering, but research on domain-independent question answering techniques that leverage the availability of large collections may ultimately yield robust broad-coverage systems that are suitable for use in museum settings.

Cultural Anthropologist / Folklore  Understanding folklore can be an important component of an ethno-graphic study, and oral narratives provide a particularly useful way of gaining insight into that folklore. Because folklore is the product of a culture rather than an individual, cultural anthropologists make extensive use of comparative analysis. In addition to the “more like this” searches mentioned above, this type of use might be facilitated through inter-linear display of transcripts or more advanced ways of simultaneously visualizing the content of multiple narratives. Folklore collections will likely pose particular challenges for systems based on automatic speech recognition, since dialects, accents and speaking styles may differ markedly from those for which standard system have been designed.

Social Historian / Teaching  The spoken word offers a wonderful way to bring history alive in the classroom. Political history might be brought alive in many ways—through speeches, recorded meetings, or tapes of telephone conversations, for example. But bringing social history into a classroom requires that we bring in the voices of members of that society. The informational content of a narrative will be important, of course, but teachers will need more than topic-based searching. Students at different levels will benefit from different modes of expression (e.g., vocabulary complexity). The emotional state of the speaker may be important an important factor, either because it may make the message more compelling or because an absence of emotion may make it easier to grasp the meaning of what is being said. Speakers with intersecting lifelines might help to illustrate either common experiences or individual differences.

Genealogist / Personal History  In many archives, genealogists make more extensive use of the collection than any other user group. This use has generally focused on paper records that focus on facts (e.g., marriage licenses) rather than narratives. Oral narratives offer genealogists the ability to link facts that they discover with the stories behind those facts, thus painting a far richer picture of the portion of our heritage for which recordings are available than would otherwise be possible. This process, which combines aspects of genealogy, biography, and narrative, is sometimes called “personal history” (c.f., http://www.personalhistorians.org). Genealogists creating personal histories would benefit from search capabilities tailored to personal names, locations, and certain types of events (e.g., birth, marriage, and death). Visualizations of individual narratives based on life experiences (in the order lived, rather than the order told) might be particularly useful. And it would likely also be helpful to support the inclusion of specific portions of an oral narrative into the products of genealogical research. For example, hypertext links might be provided from a family tree to portions of an oral narrative that describe specific events.

Sociolinguist / Field Research  Language is one of the most exceptional creations of the human mind, and undoubtedly one of the most complex. There are over 6,500 languages and dialects spoken in the world today, this diversity offers important clues to both the nature of language and to the history of human development. Features common to a set of languages can provide evidence of past contact between cultures, thus helping us to better understand migration and trade before the development of written records [1]. Regularities seen over large sets of otherwise unrelated languages can offer insight into brain function and physiological constraints on the production and perception of language. Much of this research depends on access to speech in less widely spoken languages because it is there that unique patterns are most easily observed. Those languages are increasingly endangered, however, with an aging speaker population and increasingly strong incentives for younger people to learn more widely spoken languages. Sociolinguists therefore conduct field work in an effort to capture and analyze as much linguistic evidence through spoken expression as is possible before it is lost permanently. Much of this expression will naturally be in the form
of oral narrative, since narrative is a natural form of expression that is easily captured, and because the content of the narratives can provide evidence that complements the linguistic analysis. Sociolinguists would benefit from search capabilities based on language and dialect, allowing them to find oral narratives that have already been collected and to focus new collection efforts on cases that are not yet well documented. They might also benefit from tools that can identify specific characteristics of spoken expression, which could enhance the speed, and therefore the scope, of their work.

2.5.3 Examples of current practice

In this section, we offer some examples that show how oral narrative are being created, managed, and used in existing projects.

Oral History. The largest oral history project to digitize their entire collection is a set of testimonies by survivors, rescuers and liberators of the Nazi Holocaust that has been created by the Survivors of the Shoah Visual History Foundation [2]. The collection of 52,000 testimonies runs 116,000 hours—more than one person could listen to in a lifetime. About half the collection has been digitized to date; when complete the archive will require 180 TB of storage (in MPEG-1). Disk arrays of this size are still quite expensive, so a tape robot is presently used to deliver digital service copies. Archival masters for each interview are stored in a separate location on videotape. Manually transcribing a collection of this size is impractical (with an estimated cost of over $100 million USD), but no automatic indexing method presently exists that can accommodate the 32 languages, broad range of accents, and frequent occurrence of disfluent speech in this collection. The Foundation has therefore adopted a real-time indexing strategy in which professional cataloguers view a recorded testimony and create controlled vocabulary metadata that can later be searched. Even this effort will be costly, requiring a team of 30 cataloguers working for 4 years at a cost of $12 million USD.

Video Diaries. With the advent of inexpensive camcorders, many people have begun to document their personal experiences using audio and video. As digital storage and transmission becomes more widely used, it will become possible for people to share the oral narratives that they create within online communities. One early effort that offers some insight into how such a community might work is the British Broadcasting Corporation’s Video Nation program (http://www.bbc.co.uk/videonation/). Starting in 1993, individuals were given the equipment and training needed to craft and record their own stories on videotape. The resulting tapes were then professionally edited for broadcast on television as a short subject, typically about 5 minutes long. Although the program is no longer broadcast, the existing stories remain accessible online (at http://www.bbc.co.uk/videonation/) and new stories are still being created. The collection presently contains 260 stories, which are manually indexed by category (e.g., death, leisure, and values), geographic region, and creator.

2.5.4 Contextualizing information

An important part of information use is the process of sensemaking, and making sense of an oral narrative often requires some information about the context in which that narrative exists. In this section, we summarize the types of contextualizing information that have been mentioned in the context of this use case. Our goal here is to provide an easily consulted list that will help readers to understand the types of metadata that users of oral narrative collections might find to be valuable. Inclusion of an item on this list is meant to suggest that the information would be useful for some type of oral narrative, not that it is would be universally applicable.

- The research community for which the narrative was created.
- Purpose for which the narrative was obtained
- How speakers were selected
- What speakers were asked to talk about (e.g., the interview protocol)
- Postprocessing (e.g., correction of transcripts)
• Indexing (e.g., Encoded Archival Description finding aids that have been created)
• Links to related information from other sources

2.5.5 Research issues

Our goal in this report is to craft a research agenda. In this section, we collect the research issues identified above in the form of a list for easy reference. Readers should refer to the discussion above if necessary to understand the context of items on the list.

• Discovery
  – Of small collections (e.g., holdings of community libraries)

• Search
  – Based on broad topics (subject-based searching)
  – Based on persons, locations and specific events
  – Based on the nature of the expression (e.g., emotion, vocabulary complexity)
  – Based on the effect on the emotional state of the listener
  – Based on similarity to narratives that have already been found
  – Based on language, dialect, and linguistic features
  – Exploiting the structure of the narrative
  – Standards for creator-assigned metadata
  – Processing variable accents and speaking styles

• Use
  – Definition and enforcement of access rights
  – Visualization of individual narratives (e.g., as-experienced timelines)
  – Support for contrastive analysis (e.g., simultaneous visualization)
  – Support for simulated conversations
  – Inclusion of portions of an oral narrative into other works
  – Influence of audio availability on authority of transcripts

2.6 Telephone Conversations

2.6.1 Content characteristics

• Spontaneous informal speech ?Johnson tapes? 
• 2 speakers but could be a few, usually only 2 at one time (teleconferences are meetings by telephone)
• varied acoustic conditions (local, longdistance, cellular, handset, bandwidth limited (4kHz), 8kHz sampling background noise street/music/conversations)
• language switching? could happen when speakers are multilingual switching within one sentence, also call help centers
• ?anything special for historical recordings?
2.6.2 Representative user groups and tasks

- historians/ govt history
- authors/biographies
- education
- companies (help lines, services) - short term preservation only?
- administrative services (administrative papers, taxes)
- emergency services (police, medical, fire)
- security / surveillance

2.6.3 Examples of current practice

- company services (banks/loans/stocks gather info and enter manually, some with telephone touch-pad), recording of conversation for quality control and disputes
- information services (travel services- train/plane/hotels, entertainment- sports/theater/movies) - not recorded except for quality control?
- security: (authorized or not?) data saved on dat tapes? some metatdata known (what telephone numbers, suspected speakers) ? don’t really know what is done ?
- Johnson archives?? anything to say here??

2.6.4 Contextualizing information

- security - recorded in context of ongoing investigation or suspected criminal acts
- call centers: purpose of services offered, use of recordings - improve quality, improve/expand services, disputes known vs unknown vs new clients novice vs frequent callers
- historical- known political figures, record for posterity particular current events?

2.6.5 Research issues

- for companies
- call centers and emergency services - incentives to record calls (automated data mining can lead to better procedures and operating practices)
- dealing with emotional speech, stressed and noisy conditions, non-native speakers
- incentives for public figures to archive calls (dark archives)
- discovery of collections
3 Issues

3.1 Content Preservation, Content Packaging, and Standards Development

This section of the report presents an overview of the topics named in the section title: content preservation, content packaging, and standards. A final sub-section adds a few notes on four miscellaneous related topics. In part, the overview provides a sketch of completed work or work in progress by universities, libraries, and other research organizations in Europe, North America, Australia, and elsewhere. The Spoken-Word Digital Audio Collections Working Group is reporting on this broad field of endeavour in order to identify unmet needs that bear on spoken language content, and to place its proposed research agenda in context. In the area covered by this section, as will be noted below, the Working Group recommends possible activities regarding (1) file formats for spoken-language recordings, (2) the development of standard(s) regarding the markup of transcriptions or other specialized renderings of speech, (3) the development of standard(s) regarding the synchronization of sound and transcription, and (4) the applicability of data grids to dispersed spoken language data sets.

3.1.1 Content Preservation from a Digital Library Perspective

The problem of preserving content in digital form has received widespread attention in the library and archive community in recent years, starting with dramatic statements of the problem in the early 1990s, moving to solution proposals in the early 2000s. The solution statements call for of actions on multiple fronts—technical, policy and political, organizational—and some of these actions are beginning to come into view today.

The first problem statements employed familiar examples, e.g., old 5-inch diskettes that cannot be played in the new 3-inch drives in new computers, with added remarks about reading WordStar documents in a Microsoft Word environment. One of the most widely read and carefully developed expressions of these ideas was Jeff Rothenberg’s article “Ensuring the Longevity of Digital Documents” (Scientific American, January 1995).

In the Library community, the solution statements began by introducing a pair of terms that were at first seen as distinct and even opposite: migration and system emulation. The former was articulated in the 1996 report by Don Waters and John Garrett, Preserving Digital Information: Final Report and Recommendations (http://www.rlg.org/ArchTF/index.html). In that report and in discussions since, two meanings for migration have emerged. The first, which many call refreshment, refers to the movement of an unchanged bitstream from one media or device to another, more or less what happens when you move your files unchanged from your old computer to its new replacement. This is essential to the proper workings of any computer system—especially big ones—but less fraught with issues than the other type of migration. The second meaning, which some call format migration, refers to the movement of content or data from one bitstream structure (approximately “file format”) to another, more or less what happens when you take your WordStar document and convert it to Word. As the example suggests, this type of migration is usually contemplated because of the obsolescence of the existing bitstream structure or, more accurately, the obsolescence of the software tools needed to render that bitstream into a form comprehensible by humans.

Formats established by such standards groups as ISO (International Organization for Standardization), NISO (National Information Standards Organization), W3C (World Wide Web Consortium), and I3A (International Imaging Industry Association) are very likely to be migratable. Some widely deployed, openly documented industry formats industry are also likely to be migratable, e.g., the TIFF (Tagged Image File Format) format, developed by Aldus, Inc., now Adobe, Inc. TIFF is an example of an “open format,” marked by minimal proprietariness. There exists what might be called an “openness spectrum,” with degrees of protection applied by industry developers. For example, the PostScript page description language (also from Adobe, Inc.) could be described as “proprietary and open.” Very comprehensive and fully public documentation is available for some proprietary content formats, e.g., TIFF.

Other factors also increase the feasibility of migration. For example, machine readable texts will be easier to migrate when the character set is known, especially when the character sets follow ISO/NISO standards, with the UNICODE international character set seemingly the best choice for long term content.
preservation. But experts in language correctly warn that the encoding of important features of language requires more than just a character set.

Migratability is increased when thorough documentation for a given format is public and widely available, and will be further increased when the facts of the creator’s use of the format (e.g., facts about the customization of the creating software application) have also been documented. Although never a sufficient condition, migratability may be more plausible when a format is widely deployed and implemented, or when it can be opened in software applications from a broad range of companies. Migration will succeed when the significant or essential features of content—see below—are maintained when a given unit or subunit of that content is migrated from an obsolescent format to its replacement.

The shortcomings of format migration contributed to the advocacy of system emulation by Jeff Rothenberg in his report Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation (http://www.clir.org/pubs/abstract/pub77.html), and in similar recommendations by others. Rothenberg pointed out that format migrations change the look and feel of documents. Anyone who has migrated heavily formatted documents, e.g., word processing files containing such elements as tables or automatically numbered outlines, has experienced the damage inflicted by conversion (i.e., migration) programs and the consequent need to edit the new version to get back what was lost. This led Rothenberg to argue in favor of leaving the document unchanged and finding ways to emulate the systems that rendered it, e.g., and emulation of WordStar running on an emulation of Windows 3.1 running on an emulation of an Intel 286 chip.

Such true emulations may exist only for a few system types and may be costly to maintain; in some cases organizations will develop “pretty good” emulators that render content with minimal changes in appearance or functionality. The concept of what might be called levels of preservation—which apply equally to migration and emulation—is the topic of Paul Wheatley’s very helpful paper titled “Migration - a CAMiLEON discussion paper” (www.ariadne.ac.uk/issue29/camileon; also www.personal.leeds.ac.uk/~is-sprw/camileon/migration.htm; see Section 2 for Wheatley’s discussion of various levels of preservation). In a recent study, the matter of maintaining look and feel has been questioned. An activity under the direction of Margaret Hedstrom at the University of Michigan took a computer game and compared user reactions to migrated and system-emulated versions; see “Emulation vs. Migration: Do Users Care?” by Margaret Hedstrom and Clifford Lampe (http://www.rlg.org/preserv/diginews/diginews5-6.html#feature1). In both cases, there were changes in look and feel (in the words of the article, “change happens”) and it was a toss-up which approach provided a more satisfactory outcome for fans of the game.

Several digital library commentators argue that the preservation of content requires a judicious mix of format migration and system emulation, sometimes referred to as normalization. A normalization scenario might work like this:

- A library acquires content in a less preservable format.
- The content is normalized (i.e., migrated) to a format that is well understood by that library.
- The library creates a hardware and/or software system that can render the normalized content.
- The library commits to re-creating the rendering system for the long term, i.e., committing to emulate the “original” system designed by the library to render the normalized content.

Variations on the normalization approach have been developed or discussed in the digital library community. One example is the Persistent Archive design being tested within what is called the Data- intensive Computing Thrust at the supercomputer center at the University of California, San Diego, also known as the National Partnership for Advanced Computational Infrastructure (www.npaci.edu/Thrusts/DI/index.html). The persistent archive approach migrates content into what are called persistent objects for long term management; one version of the system is being tested at the National Archives and Records Administration (documents available from www.sdsc.edu/NARA/Publications.html). A second example is outlined in the article “A Project on Preservation of Digital Data” by the IBM computer scientist Raymond Lorie (RLG DigiNews, June 15, 2001; http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2). Lorie proposes that digital library organizations create Universal Virtual Computers to their own specifications and then reformat content for that computer environment. The organizations are thereby well positioned
to update the virtual computer over time as needed. Some comparisons may be made between this approach and ideas associated with the Java virtual computer promoted by Sun Systems and others. Third, the National Library of Medicine (NLM) preserves a number of medical e-journals by reformatting them from their native markup language (typically based in a particular XML DTD [Extensible Markup Language Document Type Definition] or schema) into an NLM-developed XML schema, thus simplifying management for the long term.

3.1.2 The Digital Repository and Content Packages

The preceding section outlines part of the digital library community’s current thinking about keeping content alive over the long term. The focus on bitstreams or files, however, leaves unsaid the considerable matter of who and how content management might be accomplished. One answer is packed into the term digital repository. Many writers, including the authors of the report Attributes of a Trusted Digital Repository (Research Libraries Group [RLG] and the Online Computer Library Center [OCLC], draft August 2001), use the term to name an organization in the fullest sense, i.e., both the staff and the technical systems employed to manage digital content. In the words of the report: “A reliable digital repository is one whose mission is to provide long-term access to managed digital resources; that accepts responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users; that designs its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it; that establishes methodologies for system evaluation that meet community expectations of trustworthiness; that can be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly; and whose policies, practices, and performance can be audited and measured.”

Regarding the technical systems, many in the digital library community have embraced the Reference Model for Open Archival Information Systems (OAIS). The model is introduced on the National Air and Space Administration website in these words (1999): “ISO has undertaken a new effort to develop standards in support of the long term preservation of digital information obtained from observations of the terrestrial and space environments. ISO has requested that the Consultative Committee for Space Data Systems Panel 2 coordinate the development of those standards.” (http://ssdoo.gsfc.nasa.gov/nost/isoas/) The document proper is at http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf

At this writing it would be fair to say that the digital library community’s embrace of the OAIS reference model has been somewhat theoretical; there is much discussion and some activity to implement portions of it. Several features have been especially appealing to this community, two of which are evident in the following simplified diagram. One is the modular nature of the architecture, which is in turn suggestive of what might be called a “content life cycle.” “Production” and “ingestion” are shown at the left, indicating that content is first shaped by its makers and/or reshaped by the repository organization upon arrival, to make it fit for long term archiving. Content is submitted by producers as Submission Information Packages (SIPs) for ingestion; then reshaped by the repository into Archival Information Packages (AIPs) for storage and preservation. And at the right in the diagram, content is once again reshaped by the repository into Dissemination Information Packages (DIPs) when delivery is called for, to make it fit for presentation to end-users.

The second feature of interest—in keeping with this modular view—is the notion that digital content can be seen as a “package” (for library preservation specialists this reminds us of how a microfilm of a book keeps all the pages together) and that the package may different when submitted to the repository systems, when managed by the repository for the long term, and when delivered to a customer. As this concept suggests, well-designed and standardized content packages also support the exchange of objects between repositories and their presentation to researchers.

How does this map to our notions of normalization, migration, and or system emulation? The discussions in the community often associate normalization with ingestion, i.e., when content is readied for archiving, although there may also be ex post facto normalizations as part of management of content in archival storage. Migration presumably will be an action that takes place as content is managed in archival storage. System emulations as a requirement will be part of archival storage management as well, and their availability and application will occur in the access module, when content is presented to users.
A third feature of the OAIS reference model that has been of interest to the library and archiving community concerns the four types of metadata that must be associated with content.

The two metadata categories at the right—packaging and descriptive information—are relatively self-explanatory. Packaging information is a kind of shipping manifest that binds the package together into a single identifiable unit. Descriptive information is the old librarian’s friend, information that can be indexed to support resource discovery. Content information names the content-carrying bitstreams themselves and the information needed to render and/or interpret the package. Preservation description information carries the metadata that is needed to preserve the content information.

Clearly these latter categories are crucial and they have been analyzed by working groups sponsored by the Research Libraries Group and the Online Cataloging Library Center; see the listing at http://www.rlg.org/longterm/index.html. The listing includes pointers to Preservation Metadata for Digital Objects: A Review of the State of the Art (PDF, January 2001; http://www.oclc.org/research/pmwg/presmeta_wp.pdf), Part I: A Recommendation for Content Information (PDF, October 2001; http://www.oclc.org/research/pmwg/contentinformation.pdf), and Part II: A Recommendation for Preservation Description Information (PDF, April 2002; www.oclc.org/research/pmwg/pres_desc_info.pdf).

The issues that bear on about shaping packages and retaining appropriate types of content information and preservation description information in the realm of spoken language data are nicely brought to life in Steven Bird’s and Gary Simons’s article “Seven Dimensions of Portability for Language Documentation and Description” (http://arxiv.org/abs/cs.CL/0204020). This article indicates the many challenges needed to produce (on the one hand) the needed metadata and (on the other hand) the supporting systems for emulation, migration, or normalization of language data in an OAIS package.

### 3.1.3 Packages, Repositories, and Standards

The preceding discussion highlights number of zones in which standardization is welcome. In some of these, true standards or open industry standards have been established or are proposed. In others, standards development is needed. The following overview is broad; within this context will be found certain zones in which the spoken language community can make a special contribution.

For repositories, the report Attributes of a Trusted Digital Repository (Research Libraries Group [RLG] and the Online Computer Library Center [OCLC], draft August 2001, pp. 12-14), outlines the attributes of policies, practices, and performance that can be audited and measured:

- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technology suitability
- System security
- Procedural accountability

The report discusses various models and approaches for certifying complex technical and administrative entities, e.g., the ISO 9000 process, noting that no such models or approaches have been established for digital-content repositories (pp. 14-17). The report also calls attention to the unanswered need to establish auspices and authoritative organizations for repository certification. The Working Group believes that
repository system design and certification are matters beyond the scope of a spoken language research effort.

Content packages may be teased apart into their component parts; separate standards may apply at each of the levels:

- Files or bitstreams that represent the content qua content, e.g., a WAVE file that reproduces sound, an XML file that contains a textual transcription, etc.

- Associated or supporting files, e.g., a DTD (document type definition) for marked-up text (e.g., for the XML file above), or pointers to the location of associated or supporting files

- Metadata, which may be further subdivided, as in the approach taken by the Metadata Encoding and Transmission Standard (METS):
  - Descriptive or intellectual metadata, e.g., a bibliographic record (like what is in a library catalog). The names of the languages represented in a given content item are customarily provided in this type of metadata.
  - Metadata about rights, restrictions, and/or other information that can be used to support access management systems.
  - Technical metadata, i.e., information specific to the files in a package, e.g., for audio, file format, sampling frequency, etc. In some cases, additional metadata about language, e.g., details of markup or encoding, may be provided in this type of metadata. Another form of technical metadata provides technical information about the “source” for the bitstream at hand, i.e., when an item has been reformatted from an analog source or migrated from one digital format to another.
  - Digital provenance metadata, i.e., information about the events/steps/processes that occurred in reformatting or migrating entities.
  - Structural metadata that records the relationship between the part of a package, thus permitting navigation, e.g., “this file is the first segment in a sequence,” “this file is the second,” etc.
  - Encapsulation schemes, e.g., the use of UNIX “tar” files to bind a package as it is moved within or between systems.

The spoken language community can contribute to standards development regarding these elements, taken from the preceding model:

- Files
  - Files that reproduce sound. Is it clear that PCM sampling structures are suitable for “master” files? Are certain sampling rates or bit depths (word lengths) preferred for spoken language recordings? Shall MP3 or other formats be recommended for the dissemination of content (“service” files)? Are there special features or structures pertaining to spoken language that are omitted from existing standards, e.g., for WAVE, Broadcast WAVE (BWF), or MP3 file formats?

- Associated or supporting files. This would include files that contain various types of transcripts and the “annotation graphs” employed by workers in the spoken language processing and speech recognition communities. To what degree have the practitioners in these fields defined the attributes of interest and developed standardized approaches to, say, text markup? Will elements of these communities’ approach be of value to members of other communities, e.g., oral historians, folklorists, and cultural anthropologists? Under what auspices might a standard-setting process proceed? To what degree will standards in this area address the concerns expressed in Steven Bird’s and Gary Simons’s article, cited above?

- Metadata
• In part, metadata issues are addressed in the preceding bullet regarding associated or supporting files. This metadata, however, may be embedded within files. One question will be: if a file is part of a coherent package with other elements, which parts of the embedded metadata ought to be “brought to the surface” in the object?

• Recommendations pertaining to the identification of languages and/or the markup and encoding of language representations in text.

• Does analysis carried out by the spoken language community lead to a recommendation when choosing between standards that overlap or perform similar functions, e.g., METS as compared to MPEG-21 for packaging, SMIL as compared to MPEG-7 and METS for segmented or synchronized content elements?

• Are there special classes of metadata pertaining to content of interest to the spoken language community not addressed by any other standards? Is there a process that will describe these classes and take the actions that may be needed to establish a standard? To what degree will practices in this area address the concerns expressed in Steven Bird’s and Gary Simons’s article, cited above?

3.1.4 Other Topics

It is worth enumerating a number of topics related to the management and preservation of digital content. Work in these areas will benefit spoken language researchers but need not distract them from focusing on their unique specialization.

Authentication of documents  The RLG-OCLC report Attributes of a Trusted Digital Repository (draft version, August 2001, p. 25) states that a trusted repository enables the content it holds “to be disseminated as authenticated copies of the original or traceable to the original.” Content that is made accessible by a trusted institution will generally be perceived as authentic or trustworthy, even in the absence of such technologies as digital signatures to authenticate the content. The InterPARES Project Authenticity Task Force Final Report observes, “The Task Force found that most contemporary records systems are a hybrid of electronic and paper records; that few explicit measures are employed to ensure the authenticity of electronic records, and that authenticity is generally assured through procedural means.” (Authenticity Task Force Final Report. International Research on Permanent Authentic Records in Electronic Systems [InterPARES], http://www.interpares.org/documents/atf_draft_final_report.pdf, p. v.) Some organizations, like libraries, receive materials from others and then provide access to them, and therefore may ask the provider to authenticate the documents when they are first submitted, “in order to produce authentic copies after they have been transferred to the preserver’s custody.” (Authenticity Task Force Final Report. International Research on Permanent Authentic Records in Electronic Systems [InterPARES], http://www.interpares.org/documents/atf_draft_final_report.pdf, pp. 1-2.)

Authentication of users  The legal status of many of the most valuable spoken language resources, e.g., oral history recordings, may not permit their broad and open dissemination. The provision of limited access to designated researchers therefore will require a user-authentication system. Some of the tools, e.g., public key infrastructures, that may serve to authenticate content may also support user authentication.

File integrity management  A feature of any well run repository is automated systems to monitor the integrity of files, e.g., the generation and comparison of checksums. Note, however, that such tools serve only to determine that a given bitstream has not changed, and will not address the need to determine that no significant information has been lost when a bitstream is migrated to a new format.

Data grids  Data grids pertain to the provision of shared access to dispersed scientific data. The web site for Globus (http://www.globus.org/datagrid/) leads off with this paragraph: “Distributed scientific and engineering applications often require access to large amounts of data (terabytes or petabytes). Future applications envisioned by our team also require widely distributed access to data. (For example, access in many places by many people, virtual collaborative environments, etc.). The Globus
Project’s data grid effort attempts to identify, prototype, and evaluate the key technologies required to support data grids for scientific and engineering collaborations.” Globus’s principle investigators are located at the Argonne National Laboratory and USC/Information Sciences Institute.

Another science-based data grid activity is Griphyn (http://www.griphyn.org/). Their site reports: “The GriPhyN (Grid Physics Network) project brings together an outstanding team of information technology (IT) researchers and experimental physicists to provide the IT advances required to enable Petabyte-scale data intensive science in the 21st century. Driving the project are unprecedented requirements for geographically dispersed extraction of complex scientific information from very large collections of measured data. . . . Our team is composed of seven IT research groups and members of four NSF-funded frontier physics experiments. We believe that only an integrated research effort will provide the coordination and tight feedback from prototypes and tests that will enable both communities to meet their goals. The four physics experiments are about to enter a new era of exploration of the fundamental forces of nature and the structure of the universe. The CMS and ATLAS experiments at the Large Hadron Collider will search for the origins of mass and probe matter at the smallest length scales; LIGO (Laser Interferometer Gravitational-wave Observatory) will detect the gravitational waves of pulsars, supernovae and in- spiraling binary stars; and SDSS (Sloan Digital Sky Survey) will carry out an automated sky survey enabling systematic studies of stars, galaxies, nebulae, and large-scale structure.”

The spoken language community may wish to investigate the applicability of data grids to the problem of sharing access to dispersed language data sets.

3.2 Copyright

A discussion of spoken word and copyright must focus on three issues: whether these materials are likely to be protected by copyright, whether other rights must be taken into consideration when considering a digitization initiative, and, all rights notwithstanding, whether an argument can be made to proceed with digitization and delivery. Copyright legislation has changed dramatically over the past decade, both in the United States and in Europe. The rise and demise of Napster and other online file-swapping services have focused the attention of the technology, content, legal and consumer advocacy communities on the issue of digital audio distribution. Despite this attention and debate, clear rules have failed to emerge, and are unlikely to arise in the near term, particularly for the non-commercial enterprises contemplated by many libraries and archives.

3.2.1 Extent of copyright protections for spoken word materials

As signatories to the Berne convention, the United States and the European Union member nations have established reciprocity in copyright protection so that materials created or published in one nation will enjoy the same protections in other nations. Copyright statutes generally reserve for the copyright holder the exclusive right to reproduce, display, distribute copies of and perform or broadcast the work. The European Union issued a new copyright directive in 2001 which closely matches many of the technical provisions introduced in the United States’ Digital Millennium Copyright Act (DMCA) of 1998. Both extend encryption protections with harsh anti-circumvention language. How principles in the EU Copyright Directive will be implemented through the laws of member nations remains to be seen.

In general, sound recordings have historically been accorded fewer protections than have other types of works, though some recent initiatives have had the effect of increasing protection. In the United States, sound recordings were not protected by federal copyright law until 1972, and recordings made before that date are still not federally protected (though they may be under state copyright laws), though works fixed after 1977 receive at least 70 years of protection. In the United Kingdom, copyright for sound recordings

---

6Directive 2001/29/EC

7see http://www.law.asu.edu/HomePages/Karjala/OpposingCopyrightExtension/legmats/HarmonizationChartDS K.html, “Chart showing changes made and the degree of harmonization achieved and disharmonization exacerbated by the Sonny Bono Copyright Term extension Act (CTEA)”

8In the United States, in order for works to qualify for protection, they must be fixed in some physical medium. This requirement has been clarified to encompass digital publication, as well.
was established in the 1911 law\(^9\) and lasts for 50 years, 20 fewer years than granted to creators of print works. The 1979 revision of the Berne convention\(^10\) likewise established a 50-year duration of copyright, a term also endorsed by the European Union in its 1993 directive\(^11\).

Most of the signatory nations require either some form of fixity, in the case of the United States, or availability to the public, in the case of Germany and the United Kingdom, in order to claim copyright protection. However, France's copyright law is much more generous toward authors, stating: "A work shall be deemed to have been created, irrespective of any public disclosure, by the mere fact of realization of the author’s concept, even if incomplete.”\(^12\)

There may be several layers of authorship embedded in a single sound recording, and each act of authorship may be subject to separate protection. For a musical work, the composition and arrangement might both be protected even if the physical recording itself were not. A more relevant example of layered rights may be seen in observing several separate acts of creation that might be said to be encompassed within a sound recording of a news broadcast: a typescript, background music, and interviews with news subjects. It is unclear how stringently these protections are likely to be pursued and enforced. The Berne convention singles out certain types of works and suggests that signatory states may wish to exempt them from copyright protection. The article reads in part:

1. It shall be a matter for legislation in the countries of the Union to exclude, wholly or in part, from the protection provided by the preceding Article political speeches and speeches delivered in the course of legal proceedings.

2. It shall also be a matter for legislation in the countries of the Union to determine the conditions under which lectures, addresses and other works of the same nature which are delivered in public may be reproduced by the press, broadcast, communicated to the public by wire and made the subject of public communication as envisaged in Article 11bis(1) of this Convention, when such use is justified by the informative purpose.\(^13\)

It might therefore be argued that, in some countries, recordings of lectures, speeches and courtroom oral arguments, and other such public speech enjoy fewer protections under the law. [Discussion point: can we claim that the spoken words of public officials are NOT copyrightable?]

### 3.2.2 Other rights: privacy, publicity and moral rights

Beyond the set of rights reserved by copyright laws, other rights may come into question with audio digitization projects. Among these are the rights of privacy, publicity and so-called “moral rights.”

Privacy and publicity rights in the United States are primarily protected under state and local laws. As such, specific provisions will vary, but generally speaking, a right of publicity prevents “the unauthorized commercial use of an individual[s] name, likeness, or other recognizable aspects of one’s persona. It gives an individual the exclusive right to license the use of their identity for commercial promotion.”\(^14\) It might reasonably be argued that creation and provision of digital audio archives for research purposes are a non-commercial use and therefore unlikely to cause any archive to run afoul of such privacy laws.

In order to protect against invasion of privacy claims, collecting agencies should make a determination about whether the subjects have granted permission for a recording to be made, either implicitly or explicitly. A signed consent or permission form will always be the best safeguard, but such instruments are not likely to be available for most recordings, particularly older works. It is not likely that a public figure, such as a politician or a known lecturer, could substantiate an invasion of privacy claim were his speech to be recorded. The more public the citizen, the less likely he is to be able to make a claim. The right of privacy is a right entirely separate from copyright, and care must be taken not to conflate the two.

\(^9\)The Copyright Act, 1911

\(^10\)Berne Convention for the Protection of Literary and Artistic Works

\(^11\)“Council Directive 93/98/ECC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights”

\(^12\)see http://clea.wipo.int/clea/Ipext.dll/Folder/Infobase/17e8f/180067 fn=document-frame.htm&f=templates&2.0


\(^14\)http://www.law.cornell.edu/topics/publicity.html
Moral rights are established in the copyright laws of several countries, but are not universally supported and protected. The United States, for example, grants rights of attribution and integrity only to authors of works of visual art, and extends them to the end of the author’s natural life. However, German and French copyright law extend these moral rights to authors of all works and allow them to be transferred to heirs. Moral rights allow the author to associate or disassociate herself from works, including derivative works, and, in the case of French law, to prevent release of or remove from public availability already published works. It is possible that moral rights will play a role in evaluating spoken word collections, particularly in the cases of unscripted or extemporaneous speech in oral histories, interviews, meetings, and the like, where it is perhaps more likely that a subject will wish to retract or withdraw.

### 3.2.3 Making the argument to digitize

Although several countries either mandate or strongly encourage legal deposit, it is by no means true that physical ownership confers ownership of the underlying intellectual content, unless a deed of gift or some other condition of acquisition explicitly transfers copyright along with the physical artifact. Therefore, even though national and depository libraries and archives have wonderful, unique and precious audio collections at their disposal, they must look carefully at exemptions in the copyright law to make an argument in favor of digitizing, for most audio content is likely to be subject, in some degree, to copyright protections.

The “fair use” or “fair dealings” clauses are a natural starting point. Most countries have made some provision for reproducing copyrighted works for certain purposes. Those specifically mentioned included teaching, criticism, news reporting, and parody. In all cases, the language of the copyright law is non-specific as to the particulars. The United States copyright law’s Fair Use clause cites “amount and substantiality of the portion used in relation to the copyrighted work as a whole” as a factor, but states that it must be balanced along with three other factors and offers no specifics about what a “substantial portion” might be. In practice, however, fair dealings clauses have been problematic. Their vagueness, rather than opening the door to interpretation, has led instead to self-censorship captured by the old adage, “Floors have become ceilings.”

Although incompletely illustrated here, specifics of copyright law vary from country to country, even among the signatories to the Berne convention. It may be, then, that a more productive collaborative activity will be to establish some reasonable “acceptable risk” policies and practices, which need not be overly concerned with a true reading (or worse, strict interpretation of case law) of any one copyright statute. We can take heart in the example set by the Australian National Archives through its recent decision to begin digitizing and making freely available archival materials, regardless of copyright status, to help overcome the “tyranny of distance”:

The approach adopted by the Archives was to look realistically at the nature of the material in question, and to look at the overriding purpose for which the Archives was planning to digitise and publish this material online. Generally the material in which copyright is held privately is of no commercial value. The records are mostly between 30 and 150 years old and, because of the passage of time, current copyright holders often cannot be identified or traced. The principal objective of our digitisation initiative is to fulfill our statutory function of encouraging and facilitating the use of archival material, and to that extent we have determined that it is in keeping with the spirit of the “fair use” provisions of our Copyright Act. On these grounds, the Archives felt that the public interest lay overwhelmingly in favour of proceeding with the initiative.

Questions to ask when considering whether or not a risk is acceptable [This section needs some flesh]:

- How old is the material?
- Was it produced for commercial or non-commercial purposes?

---

15Title 17 U.S. Code, Chapter I, Section 106A
16France, Copyright (Part I), Code (Consolidation), 01/07/1992 (03/01/1995), No. 92-597 (No. 95-4), Title II, Art. L. 121-1.
17Title 17 U.S. Code, Chapter I, Section 107
18http://www.rlg.org/preserv/diginews/diginews6-4.html#feature1
• Do deeds of gift or consent forms transfer any rights?
• Can a copyright holder be identified with a reasonable amount of effort?
• Can access be brokered in any way to mitigate potential concerns ("thumbnail" equivalents or other deprecated versions only delivered to the public at large, full network access granted to researchers who request certain materials, etc.) about worldwide distribution?
• Can 'digital liability' insurance be obtained?
• Will the U.S. Safe Harbor provision help?

4 Research Agenda

5 Conclusion

A Selected Spoken Word Audio Collections

B Standards

References
