User Requirements and Technologies for Automated Storage and Retrieval

Interim Report of the SMPTE Study Group on User Requirements for Future Automated Storage and Retrieval Systems (V16.09)

Jürgen Heitmann, Chairman of the Study Group V16.09

Recently, users are displaying increased awareness of archiving-related problems. First of all, the limited lifetime of any physical storage medium and the necessary reproduction equipment sometimes poses insurmountable problems for the archivists. On the other hand, there is a growing realization that the real assets of all broadcast companies are the contents of their archives. Multiple re-use of audiovisual program material is steadily gaining economic significance.

The first broadcast of a program currently covers only approximately 60% of the production costs. This is of huge financial consequence for all broadcasters. Efficient archiving is becoming an economical force. The targets of archiving are moving from solely preservation to a scenario centered around easy access and use of the archive. The preservation of archived material is just an evident prerequisite for it's use. Archives are becoming libraries for audiovisual program material. In other words: archives cost money, libraries can earn money through wider internal use or expanded external use.

Realizing these facts, in December 1995, the SMPTE Committee on Television Recording and Reproduction Technology (TRRT) formed Study Group V16.09 to analyze user requirements for future automated storage and retrieval systems. The work statement says in detail:

The Study Group will document user requirements for future automated storage and retrieval systems employed in the archiving and use of audiovisual programs and related information. The study will cover all areas that will have be brought together in new archiving systems. These areas are as follows:

- Library management system(s).
- Browsing facilities.
- Automated extended term archival storage.
- Networking to production and play out diskservers (user access).
- Exchange of library management data.
- Storage media.

Based on the user requirements gathered, the Study Group will compare available and foreseeable technology and, by way of example, illustrate possible configurations in order to provide guidance to users and manufacturers of storage media and storage and retrieval hardware/software.

The Study Group shall provide specific recommendations to the parent technology committee as to which areas new Recommended Practices, Engineering Guidelines, and Standards will be necessary, to meet the requirements of both users and available or emerging technology.

This paper should be considered as an interim report of the work of Study Group V16.09. Finally, the report points to open questions and requests further contributions for the continuation of the group's work.

Present Situation

Today's archives contain a mixture of film and videotapes that are stored on shelves. To view an archived program for possible reuse, the material has to be taken from the shelves manually and then transported manually to the suitable reproduction equipment. Depending on the film or tape format, different equipment and personnel with various qualifications are required to reproduce the recorded program information. That very often means a long, disruptive time delay.

In the case of program carriers stored at reduced temperatures, there is an additional time delay due to the required conditioning to the higher temperatures of the reproduction process. That means an increase of the access time that will never be acceptable for news applications. Storing at reduced temperatures on the other hand prolongs the lifetime of the physical program carrier due to the deceleration of destructive chemical processes.

Film Archiving

Older nitrate-based films require special means of storage due to their high inflammability. This danger does not exist with film based on acetate; but these materials deteriorate due to the vinegar syndrome. It is therefore necessary to regularly check the amount of acid within the film containers. Moreover, the dyes of color films are not stable. Only an optical copy of the film can save the program information once deterioration has begun. Loss of quality is unavoidable despite the high expenditure for optical copying, and can only be avoided by a lengthy and expensive computer restoration, as was done several years ago for Disney's production *Snow White*.

Videotape Archiving

All over the world archivists of videotapes are fighting against problems. Due to chemical substances used for lubricants and binders within the magnetic layer, crystallization in the form of a white powder was found on the surface of tapes more than than ten years old. After expensive

A contribution received from Dr. Jürgen Heitmann a consultant in audiovisual media technology.

cleaning of the tapes, requiring special equipment, a reproduction of the recordings was possible. But don't worry — the duration of life of magnetic tapes is still much higher than the lifespan of the equipment needed to reproduce the recorded programs.

The retirement of experienced personnel makes servicing of older equipment even more difficult. Old recordings have to be copied, out of necessity, to newer formats all the time. This means a huge amount of (manual) work. Quality losses cannot be avoided during this process if the transfer takes place from one analog format to another analog format.

New Technologies and New Thinking for Modern Archives

The ideal of all archivists is a physical program carrier with unlimited lifespan and long-lasting reproduction equipment. Recognition is gaining ground that these demands are totally unrealistic due to innovative cycles that always become shorter. Technological progress on the other hand is offering solutions as well.

The following key technologies will be available for the design of future archives:

1. Storage and retrieval of audiovisual information in digital form as bits and bytes. Any copying that might become necessary at the time of storage can be done in a totally lossless way.

2. Computerization and automation. Robots will take over the transportation of the program carriers to and from the recording and playout stations. Library management systems are organizing the administration of the archive.

3. Objective quality check of the archived material.

The first two technologies combined enable continuous checking of the state of preservation of the archived material. This is done by analyzing the error protection information that was recorded together with the audiovisual information. The result of this analysis provides information about the state of the reproduction equipment as well. If necessary, automatic copying processes will be initiated. These copies can be made on the same type of program carrier or may use newer storage technologies if they are available.

4. Networking for catalog data and program data. Catalog, archive, and user can be separated locally.

The application of these new technologies for archiving audiovisual program material requires a totally new way of thinking. This new thinking says that the first priority of archiving should not be the preservation of the physical program carriers like film, tape, or disk, but the preservation of the stored program information, that means of the bits and bytes.

The value of an archive will become higher the more the owner as well as external third parties use the stored programs and pay license fees. Third parties can only use an archive provided that they have direct access to the catalog (library management system) of the archive through networking. In addition to networking, standardization of the catalog structures is required.

Data Compression

The application of data compression techniques to the audiovisual information to be archived can drastically reduce the costs for storage media and storage room. The compression factor accepted by the archivists may vary with the type of audiovisual information to be stored. MPEG-2-quality may be considered acceptable for news material about a big demonstration, but inadequate for the archiving of a Wagner opera.

A special demand is heard more often: The quality of the original recording of the video and audio information has to be retained through the archiving process. This means that information which was originally captured uncompressed should be archived in an uncompressed or at least lossless compressed way. On the other hand it is uneconomical to archive information with a high uncompressed date rate that was already captured or processed by use of compression techniques.

It can therefore be expected that within future archives we will find a variety of different audiovisual data formats. The handling of identical structured data files, which represent audiovisual information with different compression factors or even different compression algorithms, poses no problem to the archive itself. The variety of different compression formats will be of greater concern for the production area. But that's not our topic today.

Catalog

The primary tool of the archivist is cataloging, called "data base management" or "library management system." The card-index box of former times is replaced by a computer with special software that makes possible an effective search for specific information based on inputs of some keywords or key-data. A major problem is still the correct and complete logging of catalog data and key-words at the time of the archiving of new programs.

Technologies like automatic voice recognition and translation of a spoken text to an ASCII file will support automated cataloging in the future. Automatic identification of key-frames at scene changes will be added. The still to be standardized audiovisual data format must be supplemented by a header with a standardized form that will continuously be completed during every processing step of an audiovisual production by acquisition and/or post-production teams. Only such a procedure can guarantee that when moving an audiovisual data file from one library to another — or from data server A to data server B — the respective library management system is supplied with all necessary information about the moved bits and bytes.

The request for ease of program exchange was the driving force for a standard for television tape recording formats yesterday; tomorrow it will require standardized catalog forms. Online access to standardized catalogs will become the prerequisite for expanded program exchange.

Metadata

Catalog data are part of the so called "Metadata," the bits about the bits. Metadata have to become an integrated (standardized) part of an audiovisual data file format.

Metadata as defined by the Joint

EBU/SMPTE Task Force (TFHS) are all sorts of data that relate to program material but not directly to content, such as:

- Time code.
- Details of technical conditions when material was created.
- Scripts used.
- Descriptions of shooting locations.
- The publicity materials created.
- Standardized descriptive data to help in locating the material through various database entries.

There are several categories of Metadata, such as:

- Format — any information necessary to decode the content.
- Descriptive — all information used in the cataloging, search and retrieval, and administration of content.
- Association — any information necessary to achieve synchronization between different content components, and to achieve appropriate interleaving of the components.
- Composition — information required on how to combine a number of content components (e.g. video clips) into a sequence or structure.
- Other (anything not included above).

Browsing

In a production environment all creative people should have online access to the library (asset) management system. The search for old archived programs to create new ones with the help of such a system is only the first step.

As a second step, program producers would like to have a quick look at the video and audio material preselected by key-words, to decide which parts of the preselected material to dub to a production server for further processing. This "browsing" is similar to the function of visible fast forward and reverse search that is offered by television tape recorders, but cannot be provided with the required speed and functionality by a tape-based mass storage system.

It is therefore necessary to store a highly compressed mirror of all archived programs on a second parallel disk-based program carrier that can offer the necessary, much faster access. To minimize the required storage capacity (and costs) the mirror image will be highly compressed due to greatly reduced demands on picture quality for browsing regarding resolution and size. A compression scheme is required that supports the function of viewable, uninterrupted fast forward/reverse browsing. The only task of such a signal is the evaluation of program content. It will therefore be sufficient to display a video signal with one quarter of its common resolution (VHS-quality) within a quarter-sized window of a PC (or workstation) terminal. A representation of all 25 frames per second is requested by users to enable the creation of an edit decision list.

Data Format for Audiovisual Information

The answer to all the above-mentioned requirements is first of all the conversion of all video and audio signals into data signals. The format of these data signals should be independent of the program carrier used and of the TV or film format as well. Such a format is not yet standardized.

Table 1 — Archiving Systems

	Applications				
Functions (selected)	Broadcast Programs	Broadcast News	Film Stock for Broadcast	Historical Film Stock	Medical
Online access to catalog	Important	Important	Important	Important	Important
Browsing	Important	Important	Not Important	Perhaps important	Not important
Short access times	Important	Very important	Not important	Perhaps important	Important
Simultaneous access for multiple users	Less important	Very important	Not important	Less important	Important
Networking	Important	Very important	Not important	Perhaps important	Important
Compression	Lossless - high	High	Lossless- medium	Medium	Medium
Size of files	Big	Medium	Big	Medium	Medium
Play out via disk server	Mandatory	Mandatory	Not required	Perhaps required	Not required
Storage medium	Tape	Tape/ optical disk	Tape/film	Tape	Optical disk
Frequency of access	High	Very high	Medium	Low	Low

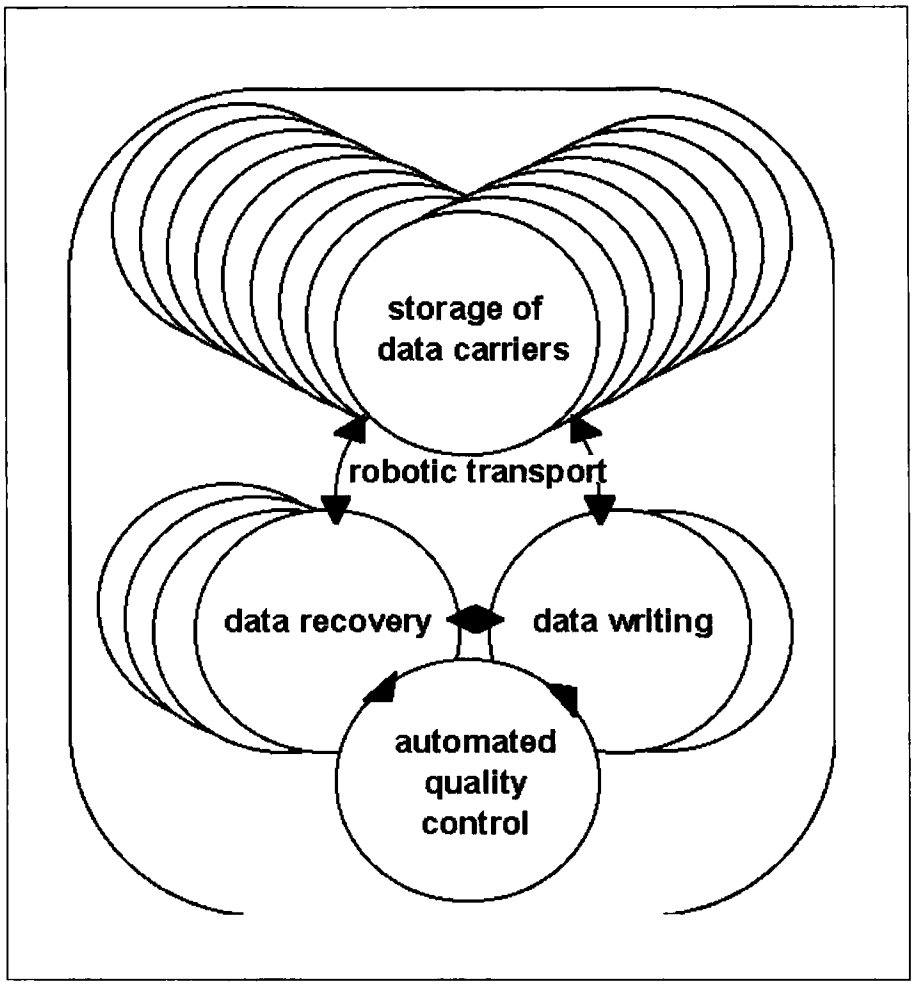

Figure 1. Automated storage system.

but discussion is going on in the joint EBU/SMPTE Task Force on Harmonized Standards (TFHS), (see *SMPTE Journal*, June 1997).

Recording Format

The selection of a recording format for a mass storage system is not a decision of primary importance, nor is it required that all archives of different users or all archives of a single user utilize one and the same recording format. It is advisable to use different types of program carriers even within one archive, for example, a storage medium with short access times for news material or a very cost effective medium for the long-term storage of soap operas.

The physical program carriers never leave the archive; program access and exchange take place via a data network or via copying to videotape. Previous demands for a common standardized recording format can be dropped. In principle any recording format and equipment can be used as long as the loading and unloading of the data carrier can be automated. This offers the freedom to select the most economical, most appropriate storage medium and storage equipment for the expansion of an archive or the replacement of worn out storage equipment at any time.

In contrast to the opinion outlined above, some members of the Study Group have requested a new recording format that is optimized for archival purposes, and the appropriate specialized tape as well.

The costs of a mass storage system are determined by the costs of the program medium per hour of storage and by the expenditures for the necessary storage room (cm^3 per hour of storage). Based on its extremely low thickness, a tape-type program carrier exceeds any other type of storage medium regarding the volume density of stored information. Magnetic tape is the choice today; optical tape may offer an interesting alternative in the future.

Optical tape promises improved long-term stability. It is a WORM medium (write once, read many). This, however, poses no restriction at all for archiving purposes, but protects the recordings against unintentional erasing as well.

Application Determines Functions

No archive equals another. In addition to archives for broadcast programs, there are several other applications: libraries for broadcast news material, film archives for the purpose of later broadcast use, film archives for the preservation of new productions for either theater rerelease decades later or by new improved electronic distribution means, medical archives, and archives for customer data of big insurance or mail-order companies.

The requirements and the resulting structures differ for all these applications. This can be seen by a selection of functions and parameters that are valued in relation to some typical applications (Table 1).

Different applications require different functions or parameters. A short access time is more highly valued for the compilation of news programs than for the access to a stored soap opera.

This paper deals especially with broadcast archives. Based on some overlapping with other applications, it is evident that you will find components of a broadcast archive at other applications.

Basic Structures of Future Archives Performing as Online Libraries

Figure 1 shows the structure of an automated storage system. The incoming audiovisual data are written to data carriers and are moved to their final storage place by a robotic transportation system. Depending on the demand for writing new data into the archive, multiple writing equipment may be used in parallel. These paral-

lel units must not necessarily use the same type of data carrier.

This means that today such an archive can work with magnetic tape, but can evolve to another more modern type of data carrier, for example, optical tape that may be used in the future for some time in parallel to magnetic tape. Any type of data carrier can be used — including disks — as long as the writing/reading and the transportation to/from the storage place can be automated. The amount of reproduction equipment will be chosen according to the demand.

The quality of the reproduced data is automatically checked at regular time intervals. This is achieved by counting the number of defective bytes that are eliminated by the internal error correction system. Prior to the point of overloading the error correction system, lossless, error-free copying will automatically be initiated. This copy can be made on the same type of data carrier or on another, newer, and therefore, more powerful type.

Figure 2 shows the function of the automated storage system explained above within a complete archiving solution. Three components have been added: the library management system, the browsing system, and a disk-server for the play in/play out of the audiovisual program data.

The disk server has to perform several tasks. It offers simultaneous access for a multiplicity of users and acts as buffer storage to smooth traffic peaks that exceed the available resources. Another major task of the disk server is the transformation of different data speeds. The data transfer rate of the network on one side and the data rate of the recording and reproduction equipment on the other side are generally not identical. Technological progress of reproduction equipment and networking will lead to continuous change of these parameters anyway. The structure of the archiving system is such that it will be able to follow technological progress without any problem. The disk server for play in/play out is connected to external servers in the post-production area, the news area, or to a distribution server via high-speed data networks like ATM or Fibre Channel.

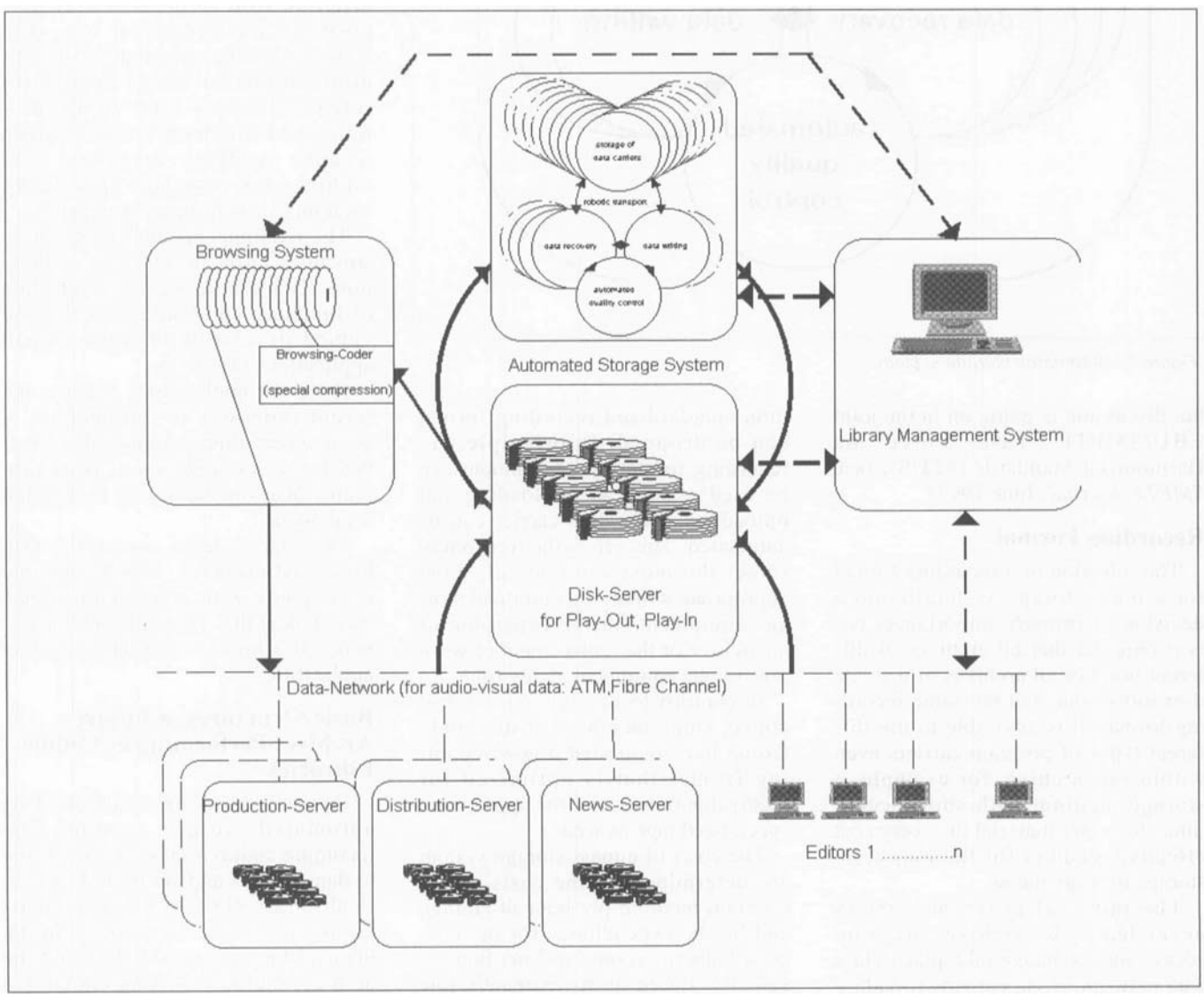

Figure 2. Structure of future archives with online library performance.

The programs stored within the archive can be accessed only via these high-speed data networks. The program carriers will never leave the archive. The requested audiovisual program will be copied to outside post-production disk servers via the high-speed network. It will not be possible to view the program content during this transfer process due to the fact that the transfer will be done slower or faster than realtime depending on the compression factor used and the speed of the network. However, this does not imply additional restriction in comparison with today's manual transportation of program carriers. They too cannot be viewed during a manual shipment.

During a long transition period, a major part of the requested programs will be copied for further transportation to videotape at the output of the disk buffer.

The library management system (or asset management system) is the administration center for the catalog data. It has to provide simultaneous online service to a multiplicity of users who are searching for a certain program. The library management system is connected with the browsing system.

The core of the browsing system is a disk-based server again. According to my assumption, this server will use magnetic hard disks due to the short access times for news applications, but may also use optical disks for other applications. That disk server is complemented by a browsing encoder that creates a highly compressed mirror image of all programs stored in the main archive.

The browsing system and library management system are connected to various editing workstations via data networking.

Protection of Today's Archives

Today's broadcast archives with their mixture of film and analog videotape can only be protected efficiently by moving the stored information from the analog carriers to digital ones. For moving all the contents of an old analog to a new digital archive, it will be necessary to provide the new archive with sufficient resources along with the normal workload.

Long Term Availability of Reproduction Equipment

It has already been stated that the lifespan of the physical program carrier, and of the reproduction equipment as well, is important in an archive's operation. I was told about archives that have stored unused but carefully sealed tape players in addition to the recorded tapes. At the time of cancellation of service and spare part support by the equipment manufacturer, and the retirement of the last service technician familiar with that old equipment, all programs of the affected program carrier will already have been copied to new program carriers. In future archives this will be done automatically, but will place a heavy load on the reading/writing resources. The expected lifespan of reproduction equipment is to be considered carefully in selecting a certain type of mass storage system.

Economic Significance

The value of new broadcast programs created per year in Europe exceeds $15 Mrd. The value of new productions worldwide is about $50 Mrd. per year. A major part of this sum, 40% or $20 Mrd., has yet to be recovered by multiple reuse of the audiovisual program material. Efficient archiving is becoming an economic force. Efficient means that archives will have to become libraries that should earn money through wider internal use or expanded external use by worldwide data nets.

Conclusion

Realizing a new archiving system will still take some time. But it will need a new way of thinking today. This new thinking says that the first priority of archiving should not be the preservation of the physical program carriers like film, tape, or disk, but the preservation of the stored program information, that means of the bits and bytes. Preservation shall not be the main goal but rather a requisite only to providing easy and fast access to the stored information. Online access to standardized catalogs will become the prerequisite for expanded program exchange. Today's archives cost money: libraries will earn money.

Open Questions and Work Still to be Done

It was one of the tasks of the Study Group to make specific recommendations to the parent technology committee for new Recommended Practices, Engineering Guidelines, and Standards necessary to meet the requirements of both users and available or emerging technology.

The following areas were identified.

Robotic Storage Systems: A draft for an RP is already under consideration by the SMPTE WG on Tape Archiving (V16.07).

Metadata: A WG on Metadata (PT20.07) has recently been formed by SMPTE.

Standardized cataloging formats: Although closely related to the subject of Metadata, the Study Group recommends determining how this may be documented and clarified.

Standardized browsing formats: The recommendation is to initiate a study on user requirements to determine the standards required to enable various user applications.

The Study Group would like to encourage all users and manufacturers to comment on this interim report and provide contributions for the future work of the Study Group or in the areas identified above. To submit a contribution or for any questions you may have, please contact either Juergen.Heitmann@t-online.de, or Eddy Zwaneveld, Chairman of V16.07, at E.HA.-ZWANEVELD @nfb.ca.