Storage and Entertainment Magazine, January 2003

Data Storage for the Digital Content Value Chain

Thomas M. Coughlin
President
Coughlin Associates

1. *Introduction*

The digitization of human content is now firmly established. Most content is now being created digitally and there are massive efforts underway to digitize the large content archives that exist. The digital content value chain consists of content creation (and conversion from analog to digital storage), content distribution, and content reception. Rich media such as movies and video are being shot using digital technology. Editing of the content is done with digital non-linear editing technology. The distribution of rich content is being done digitally over the web and at digital projection theatres. This newly digital human content requires data storage and lots of it. Without the development of lower cost, high performance data storage systems the explosion of digital content could not take place. In fact the growing existence of digital content storage makes new revenue opportunities with content archives possible as well as allowing better access and exploitation of these new digital treasure troves.

Rich digital content will be a major driver in the proliferation of digital data storage. Rich media is any combination of voice, data, video, and other technologies to create an otherwise unattainable user experience. Data storage will be used in the content creation process, the content distribution process, and in the reception and use of this digital content in homes, businesses, and in mobile devices. The growth in the availability of high bandwidth Internet access is accelerating this process. In 2001 about 5 Exabytes of disk storage and 28 Exabytes of tape were produced. Much of this new data storage is being used for rich content throughout the content value chain.

In 2001 the rich media market was about $10.8 B. By 2004 it is estimated to be over $30B (see Figure 1-1) with an annual growth rate of 53%. Data storage is currently about 20% of this market with an annual growth rate of about 70%. To give an idea of how big the rich media opportunity is here are some statistics:

- If 5% of all Internet users posted a couple of short movie clips, it would require 1.2 Petabytes.
- A high-resolution photograph from a standard digital camera is about 500Kilobytes. There are an estimated 52 billion photos taken each year and these are rapidly being converted to digital photographs.
- Forrester predicts that 92% of online consumers will communicate via rich media by 2005.
- One half TV show compressed for delivery over the Internet consumes 1 Gigabyte and 22 episodes are typically produced each year.

- There are 1600 TV stations in the US that create 21 Petabytes of airtime per year.
- If Digital Video Recorder users had on-line access to syndicated programs from the last four TV seasons that would be 320 Terabytes.

Including TV, movie, commercial production, and video distribution the annual video production storage market may be 740 Petabytes by 2006. Currently Hollywood creates an equivalent of 3.5 Petabytes annually, it also creates about 10 Petabytes in dailies and CGI cycled in production annually. Worldwide movie requirements are roughly ten times these numbers.

The digitization of hundreds of thousands of hours of analog archived content will require greater than an Exabyte of data storage. Major networks are in the midst of major digital conversion projects. CNN is in the midst of a project with IBM and Sony to digitize 20 years (about 100,000 hours) of analog video recording. 50 years of MPEG-2 quality digitized television content is estimated to require 912 Petabytes by 2006. With 10% annual utilization this would result in about 91 Petabytes downloaded or viewed per year. With mirroring this on-line content might require 3.6 Exabytes of data storage by 2006. Rich content is a significant growth driver for data storage growth at every level of the content value chain.

2. ***Digital Content Creation***

Digital video cameras (mostly using digital tape) with the most popular 24 frame per second professional format have begun to supplant traditional celluloid film cameras due to their lower cost, $1.25 per minute vs. $90 per minute and the greater ease with which the digital material can be transmitted and edited using digital non-linear editing equipment. Digital content creation and editing is done using non-compressed digital content since any compression will result in some loss of the original data. Digital content distribution is usually done using some amount of digital compression in order to reduce the data storage requirements for content distribution.

800 Terabytes of data may be required for a complete digital movie production. According to Jim Wheeler, Video and Film Archivist, standard definition NTSC video archiving requires 128 GB per hour (1 MB per frame plus audio). Some videos are two hours long so 256 Gigabytes are required. Film archiving requires several times more capacity than standard definition video (24 MB per frame is what Cinesite uses for digitizing movies) or 2.1 TB for a two hour movie. Dr. Lukas Rosenthaler, Imaging Researcher at the University of Basel in Switzerland recommends 2.5-5 TB per hour for movie archiving. Thus digital archiving of a complete two-hour movie requires as much as 10 Terabytes. Figure 2-1 shows the data storage requirements in GB/hour for various compressed and uncompressed digital formats. In addition to entertainment video and movies there are also considerable amounts of data storage involved in the creation of digital video for commercials and business uses.

According to JPA the professional and prosumer 3D digital content creation market should be greater than $1 billion by 2003. Non-linear editing (where the film edits are not necessarily done sequentially) of films requires special data storage characteristics and work with large files of uncompressed data. The various editing functions are encoding, rendering, ingest/output, and asset management. Connections must be isochronous and deterministic. Latency must be low and extremely predictable. Hardware and Software components must be optimized, features such as cache integrity checks by RAID controllers must be eliminated. Read/Write behavior will be random due to multiple layers of video and audio and random cuts from one file to another on each layer. In addition, audio and video want to share data yet exercise the system quite differently. Because of rapid changes in commands during editing the data storage system shouldn't read too far ahead. Bandwidth must be high to handle multiple users who may play the same files simultaneously. Multiple clients on different platforms will want read/write access to the same partitions and dynamic resizing of these partitions allows the most efficient use of the data storage assets. As any demanding enterprise application the storage system must be highly redundant and have no single point of failure. Drive contention must be handled gracefully using mirror set steering. Figure 2-2 is an example of a non-linear editing system implemented at a TV station in Washington, DC.

Content creation and distribution data storage system requirements can be very different. Table 2-1 compares the data storage system requirements for Video on Demand (VOD) content distribution data storage systems and those used in Professional Video Editing.

Table 2-1. Comparison of Video on Demand and Professional Video Editing Data Storage System Requirements

Video on Demand (VOD)	Professional Video Editing
One Continuous Stream of Data	Many Overlapping, Disjoint Streams of Data
Large Sequential Reads	Continuous Asynchronous Reads
Host Buffering	Hardware Buffering
Progressive Download	Real Time Random Access
Highly Compressed data	High Data Rates

Digital content can be either dynamic or static. Static content doesn't change much once it is written to the data storage system (like a recording of a movie) while dynamic content is subject to frequent changes (such as a database). The characteristics of a data storage system created for dynamic vs. static data will be significantly different. A dynamic data storage system needs rapid access to random data to change updated files whereas a static data system may not need the extra expense of the fastest SCSI or Fibre Channel disk drives and instead may be built with ATA/IDE disk drives organized to deliver content in a continuous streaming fashion. This has lead to Nearline content addressable storage systems (CAS) that are tailored to such static content. As these products begin to incorporate serial ATA (SATA) technology such as the recently announced ASACA FireFly the performance of these systems will begin to encroach upon traditional SCSI drive performance.

3. ***Digital Content Distribution***

Figure 3-1 shows one depiction of the broadband content distribution value chain (from an upcoming book by Scott Kipp about Broadband Entertainment). Ultimately digital content whether dynamic or static will need to be accessed by the ultimate users, in the case of Figure 3-1 this is a well-equipped multimedia home. Digital content in a home or office will be accessed wherever it is needed whether this is through a fiber-optic link, DSL or cable modem, or through wireless data transport systems such as various flavors of 802.11x or Bluetooth. It is likely that within a home or business many of these data channels will be used, often in parallel.

About 1 Terabyte of data storage is required for a typical multi-screen digital theatre complex. There are 61 of these in the US and 117 worldwide at this time. Many of these digital projection systems use the QuVis DVD. A movie required about 8 two -layer DVDs and 16 in some cases. Handling such a large number of DVDs may be troublesome and eventually these digital projection systems might use low cost RAID-based data storage systems to store this data in the theatres. With disk drive capacities as high as 250-320 Gigabytes today this storage system might take up no more space than a medium-sized rack mounted box and cost less than $20,000. Distribution of the digital content may be through a secure Internet connection, via satellite, or by shipping the content on specially packaged ruggedized hard disk drives.

The bandwidth requirements of digital storage systems in content distribution applications will increase with time as customers require ever richer and higher resolution media. Figure 3-2 shows the bandwidth requirements for various compressed and uncompressed video formats. As the amount of digital video resolution increases the bandwidth requirements increase.

Figure 3-3 is an example of a streaming media content distribution center using various makes of hardware and data storage for different types of content. Such a data center might be used to support VOD system or music and video subscription services. ATA-based data storage in the ASACA, EMC Centera and StorageTek based products provide low cost data storage for static or seldom changed content while a high performance EMC Symmetrix data storage system is shown in this streaming media distribution center to support more dynamic content. Other data storage companies involved in this content distribution space include niche product companies such as DataDirect and Ciprico, as well as the traditional data storage systems companies such as IBM, HP, and Hitachi Data Systems.

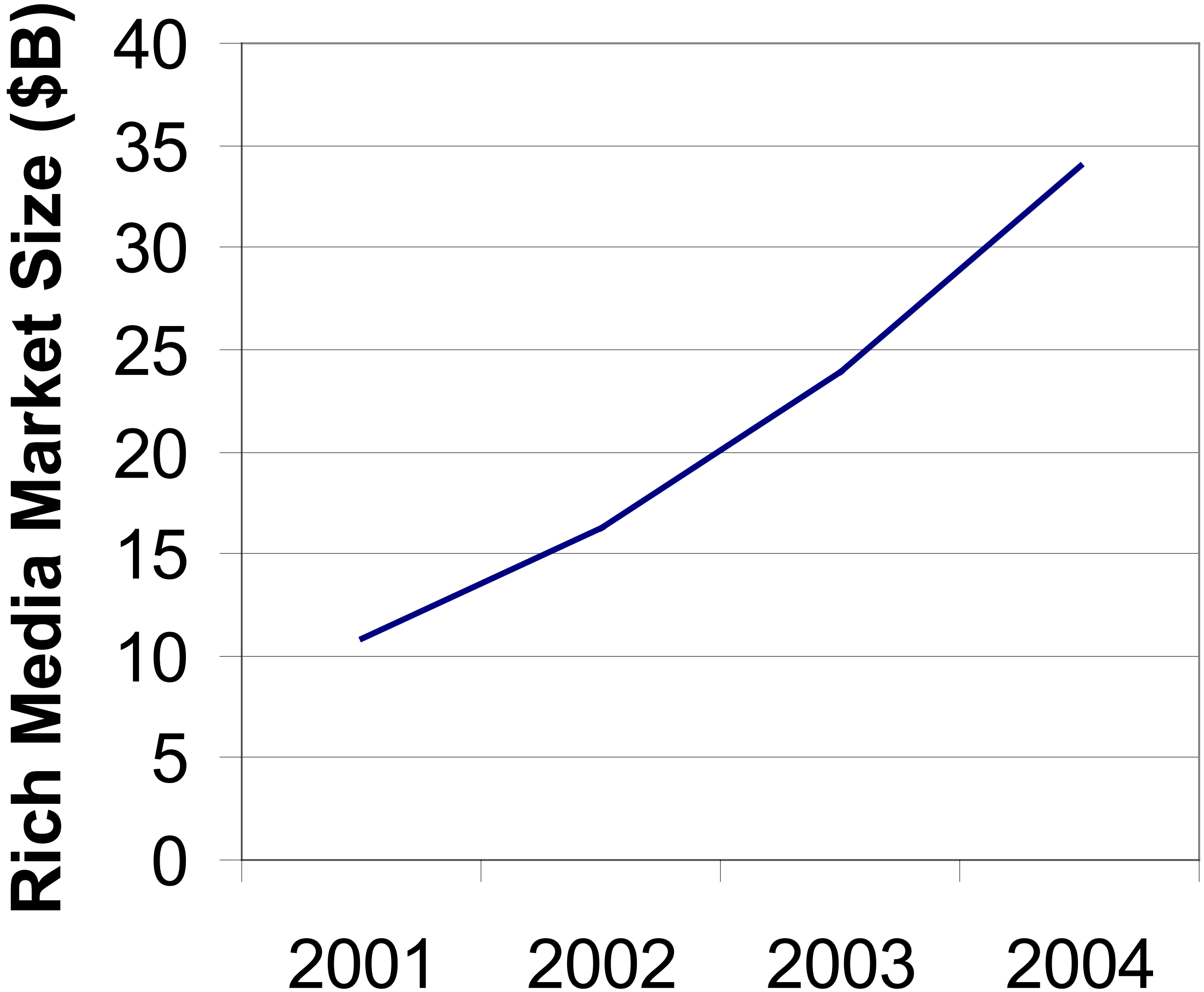
The resulting data storage requirements at the receiving end of this content distribution value chain will also increase as the bandwidth capability and availability increases and as more digital content is available to be accessed and used in the home or office. Table 3-1 shows current data storage requirements for several consumer electronic devices as well as projections for the data storage requirements for these devices by 2005. Inexpensive data storage is enabling ever greater access to human content and is the natural complement to the increasing data storage at the content creation and distribution links in the content value chain.

Table 3-1. Consumer Electronic Device Current and Future Data Storage Needs

Product	**2002 Storage Needs**	**2005 Storage Needs**
PVR/STB	40-180 GB	100-250 GB
Digital Audio Jukebox	20-100 GB	60-250 GB
Home Gateway/Media Center	100-500 GB	1-20 TB
Digital Camera	16 MB-1 GB	128 MB-10 GB
Smart Handheld (Smart Phone, PDA, Tablet)	64 MB-1 GB	128 MB-10 GB
CE Devices (Mobile MP3, GPS, etc.)	64 MB-5 GB	128 MB-20 GB

Digital content distribution makes nearly perfect copies of that content possible. Although the debate over content protection for the content owners and fair usage rules for those who purchase or lease this content have still to be worked out to everyone's satisfaction it is likely that changes in data storage hardware and software will result. It should be noted that some digital movies are now being distributed using hard disk drives that are shipped to digital projection facilities. Such data will probably include some form of encryption to protect it but additional security of content can be created by implementing some of the ATA and SCSI standard commands such as secure erase. Invoking a secure erase command causes all the data on the disk drive to be overwritten, including any reallocated sectors. If after transferring data to a digital theatre RAID the transport disk drive were secure erased then that digital content would not be accessible when the disk drive is returned to the content distribution center.

Figure 1-1.
Rich Media Market Growth



Source: FAC/Equities, 2001

Figure 2-1
Digital Storage Requirements (GB/hr)

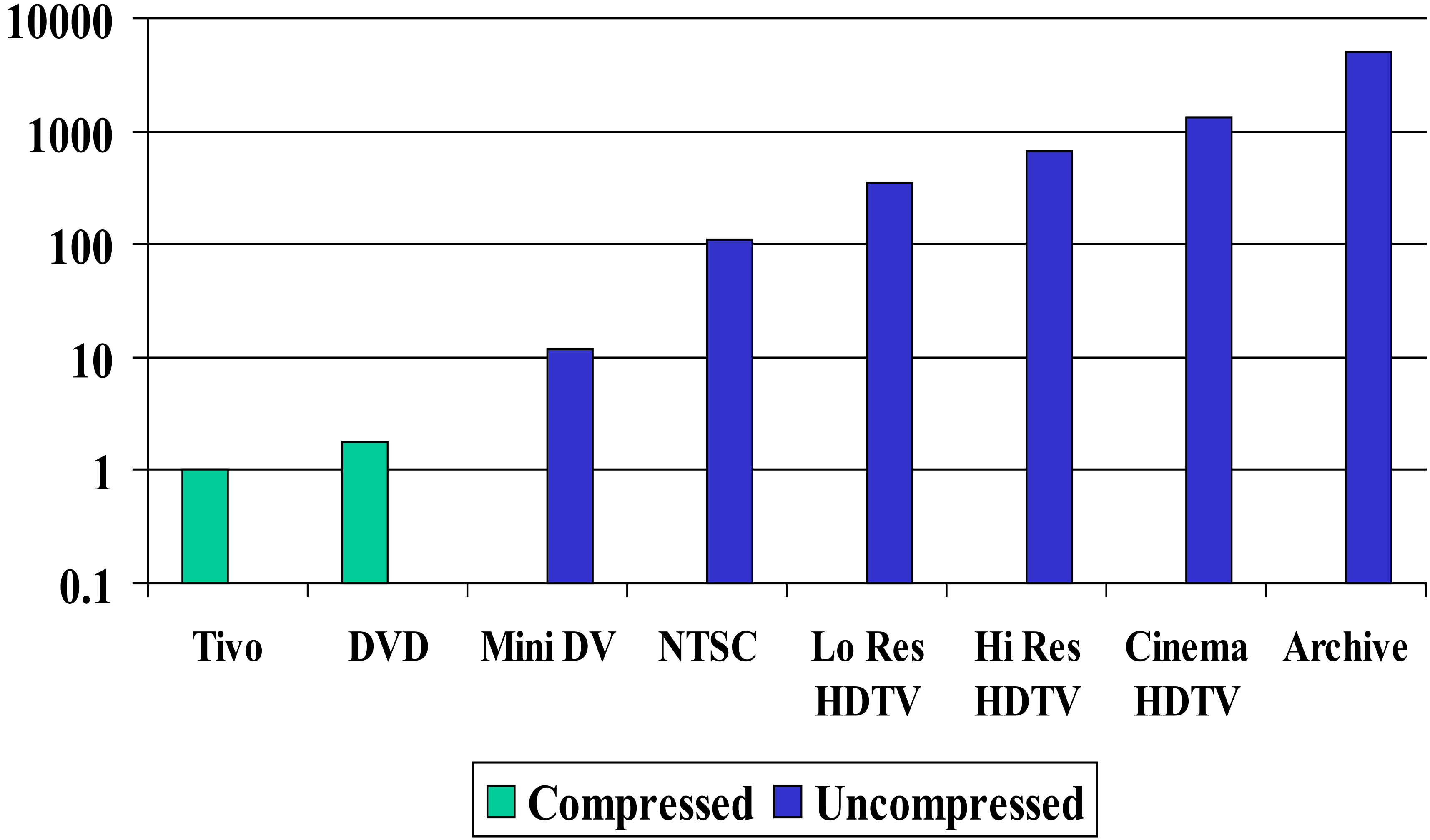
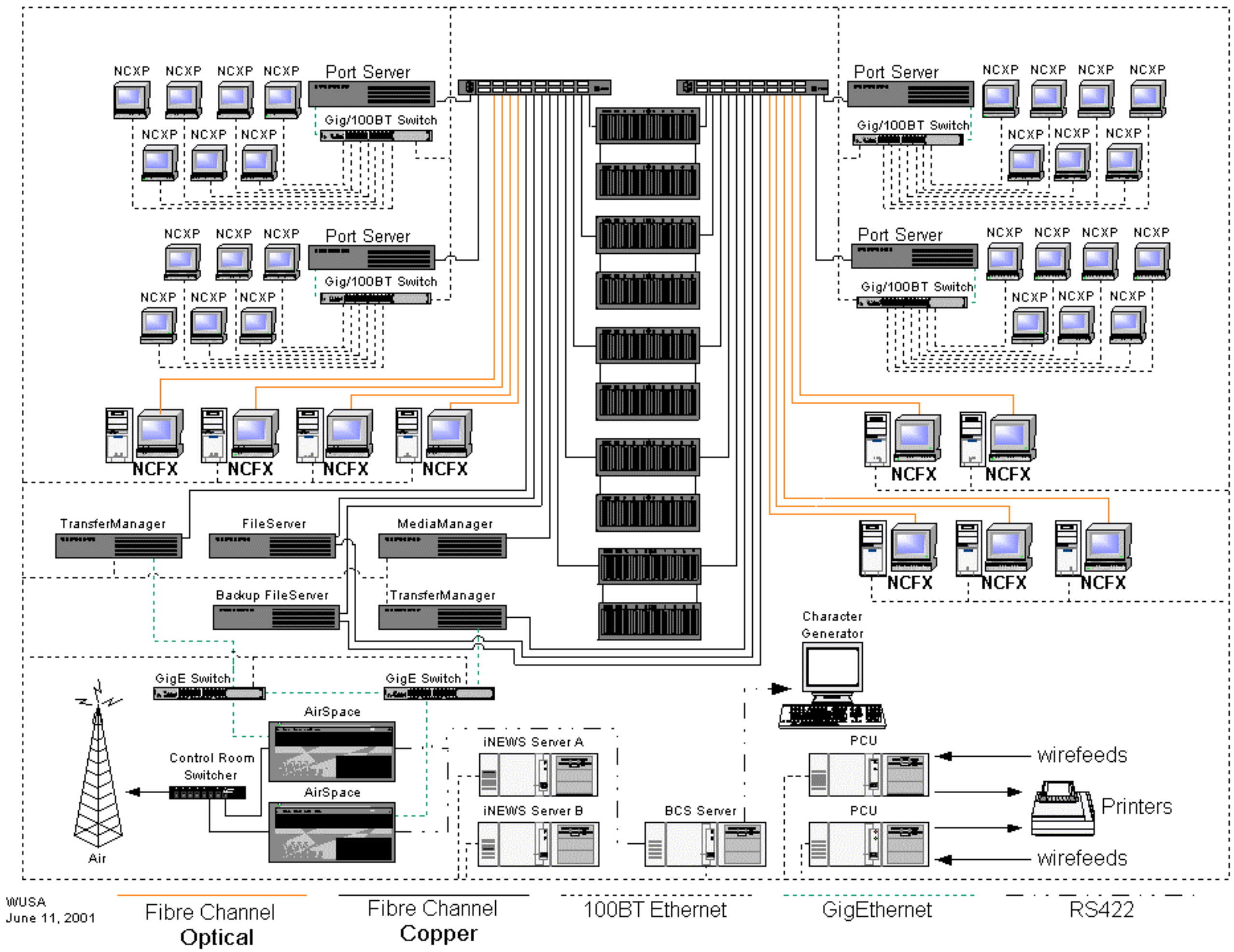
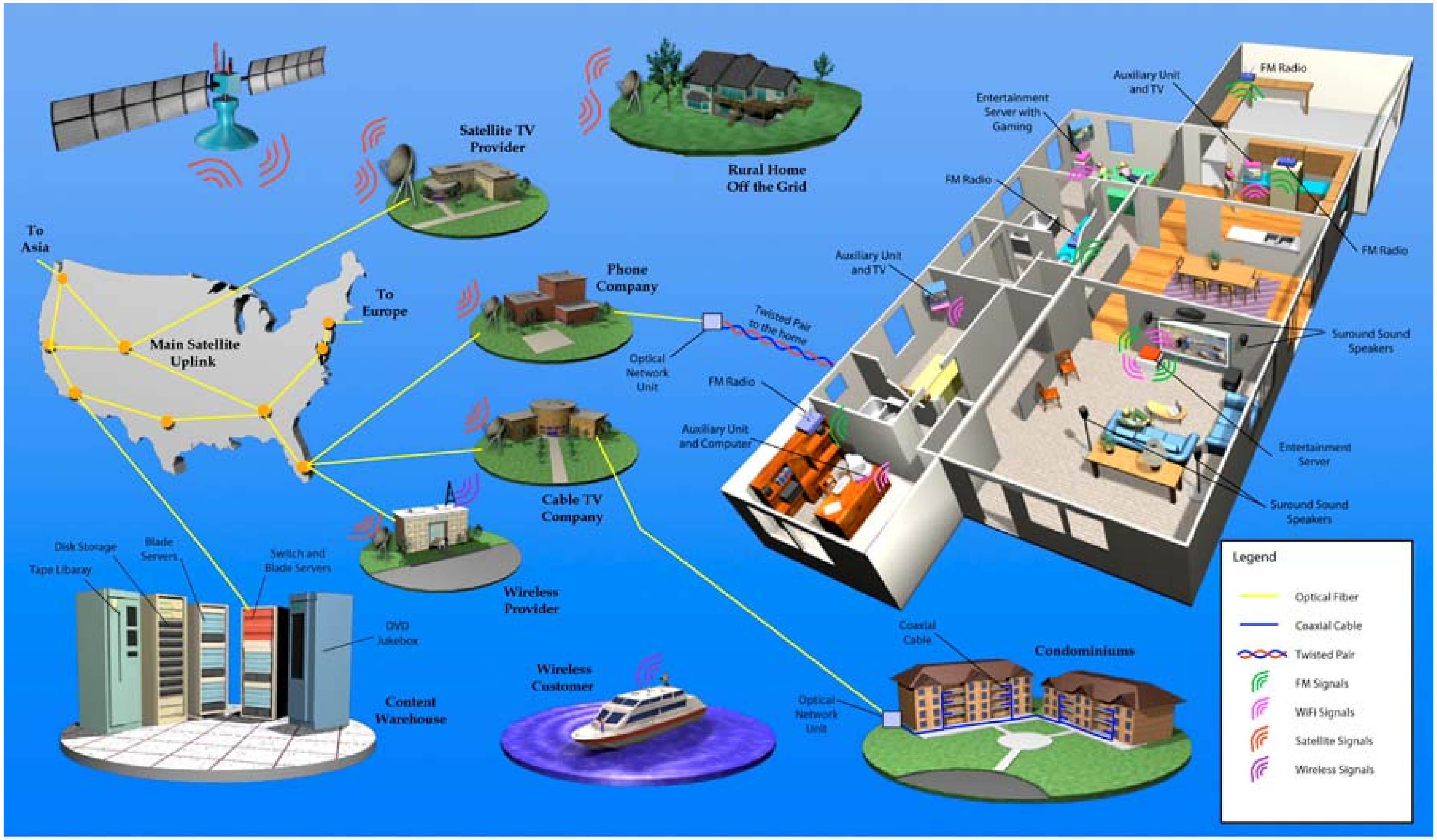


Figure 2-2. Non-Linear Digital Editing System WUSA in Washington DC



From Avid Technology, 2002

Figure 3-1.
Digital Content Distribution Chain



From Scott Kipp's upcoming book on Broadband Entertainment

Figure 3-2
Minimum Bandwidth Requirements
(MB/sec)

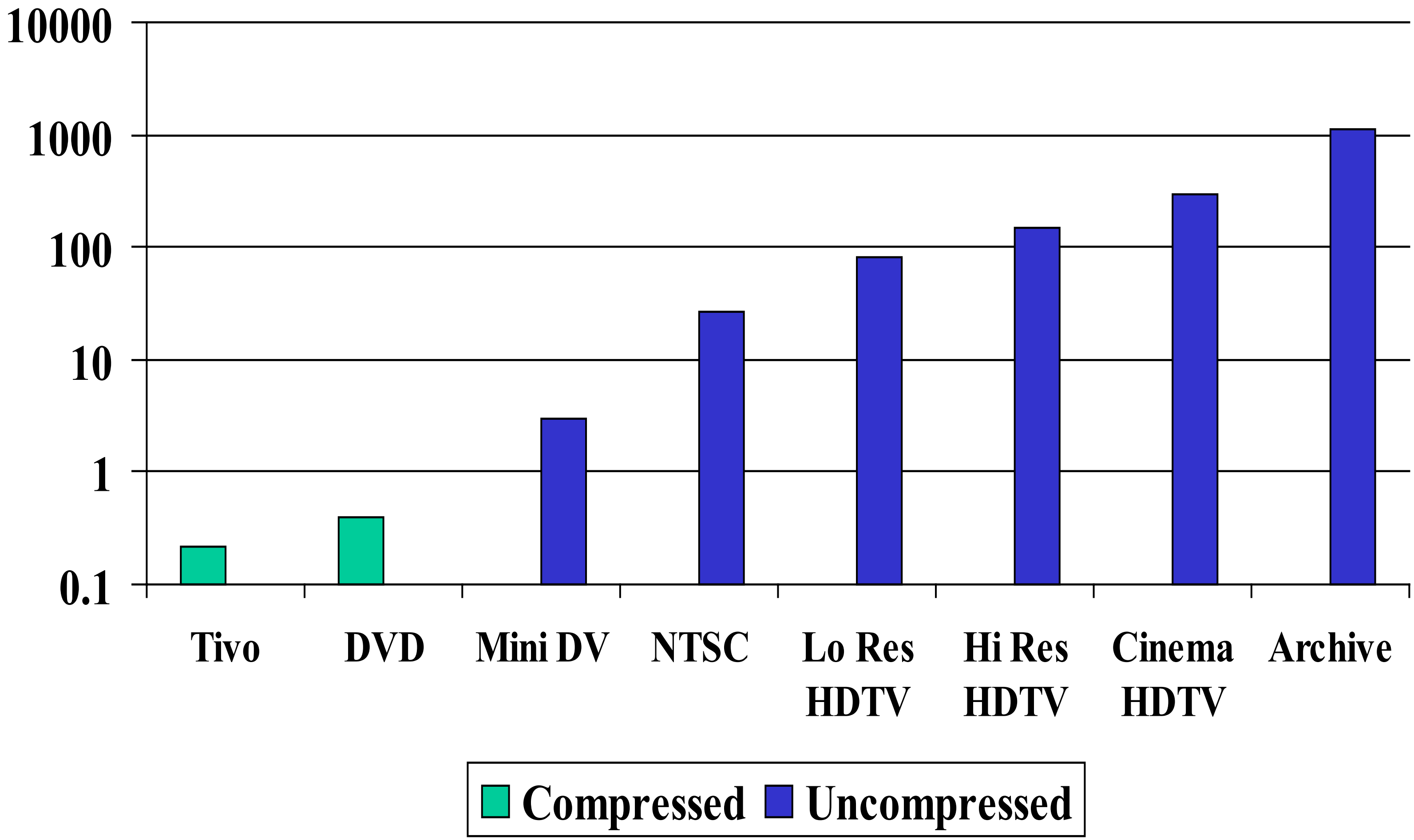


Figure 3-3. Example of a Streaming Media Datacenter