IIEEJ **Invited Paper**

The Future is More Media Convergence

Leonardo CHIARIGLIONE†

†CEDEO.net

Introduction

The Webster's defines convergence as "the act of converging and especially moving toward union or uniformity", and as "the merging of distinct technologies, industries, or devices into a unified whole".

In the Information and Communication Technologies (ICT) domain, convergence is better defined as the conversion of different industries from a "vertical" arrangement, where products and services are industry-specific, to a "horizontal" one where the structure of products and services is layered and each layer is potentially shared by different industries. The horizontal arrangement induced by convergence, far from implying uniformity, increases diversity as industries continuously expand or shrink depending on how many industry-specific technologies are shed and how many "common" technologies are added to offer new products or services that respond to customer needs.

This paper elaborates on the following media convergence elements.

1. It starts from the seminal role of the Moving Picture Experts Group (MPEG) in fostering media convergence with its first standards issued 20 years ago;
2. It explores the growing number of dimensions in which media convergence is happening or is expected to happen, often ushered in by new generations of MPEG standards;
3. It shows that convergence is not a state where uniformity rules or where industry opportunities are reduced because they merge, but a state where increasing rewards are offered to industry and consumers;
4. It proposes to apply the rules and practice of technology-driven convergence to other non-technical areas.

Convergence starts with audio and video

Most of the original research on video coding was driven by the investments started by the telecommunication industry half a century ago, but a significant role was also played by Broadcasting and Consumer Electronics (CE) research. MPEG became the channel that converted those research results into its first MPEG-1, MPEG-2 and MPEG-4 industry-agnostic standards of the 1990s. Eventually the MPEG digital media representation standards created a layer of technology shared by all industries with a current or a planned stake in audio and video products and services.

MPEG-1, targeted at interactive video on "digital storage media", was the means to bring the Telecommunications, CE and silicon industries together because of the need to share silicon technology for video coding between CE and Telecommunications. MPEG-2, targeted at digital television, added the Broadcasting industry in its three terrestrial, satellite and cable instances. The "generic coding" attribute of the joint ISO/IEC-ITU MPEG-2 standard, became the symbol of industry convergence. MPEG-4, targeted at fixed and mobile internet applications and stressing the need for software solutions, file format and streaming, added the IT industry to the other converging industries. The Advanced Video Coding (AVC, MPEG-4 part 10) standard reiterated the convergence symbolism.

Most of the original research on audio coding was carried out by the Broadcasting and CE industries, but Telecommunications did also play a role. MPEG-1 and MPEG-2 Audio were specifically for music and television. MPEG-4 Audio targeted music and speech coding in a single standard with two sets of algorithms that covered the two audio types. Only recently the Universal Speech and Audio Coding standard (USAC) has succeeded in combining the capability to handle music, speech and arbitrary combinations of the two audio types in a single algorithm. USAC can be taken as the best example of convergence: the algorithm works better than the best music coding algorithm for music and better than the best speech coding algorithm for speech.

Convergence continues with transport

Achieving convergence at the media information level was sort of easy. The tough job began when digital media had to be interfaced with delivery. MPEG-2 Program Stream (PS) and Transport Stream (TS) provided ideal technologies for integration of digital audio and video for

error-free channels and digitised analogue channels, respectively. Success breeds success and MPEG-2 TS is used today for environments that are as far as one can think from analogue media, as in the IPTV specification of OIPF.

For some time, after the crowning of the internet as the universal digital network, MPEG did not need to devote attention to media transport. Recently, however, the intrinsic lack of bandwidth guarantee of the internet has prompted the need for standards that make up for some of the shortcomings of the internet. An example is offered by Dynamic Adaptive Streaming over HTTP (DASH), a recently published MPEG standard where suitable logic at the receiver drives the flow of information from the transmitter and MPEG Media Transport (MMT), a standard oriented to broadcasting applications (second screen). The convergence of DASH and MMT as far as segment format is concerned is of primary importance and the convergence momentum must be kept with continuous attention to the transport of media information.

Convergence success is not guaranteed

The original MPEG-7 standard targeted metadata unification, but industry decided otherwise. In spite of the generous efforts of some organizations such as TV Anytime, the number of metadata schemas has only increased while each industry, one could even say each company, has decided to stick to their own proprietary metadata.

Convergence in the studio

For a long time the studio used to be the place where signal perfection was pursued disregarding what the signal would become under the effect of the distribution network. The definition of the MPEG-2 4:2:2 Profile was the first great event that brought consumer (distribution) technology into the studio through the commonality of video compression tools. The process has been further consolidated by several intra-frame-only AVC profiles and the current work toward extending the HEVC standard is continuing the trend.

Convergence of synthetic and natural media

For a long time the world of media generated by computers and media captured from the real world remained separated, contiguous or at best slightly overlapping. The early work done by MPEG in the late 1990s yielded the MPEG-4 Scene Description and Application Engine standard that enabled creation and playback of 2D and 3D scenes composed of synthetic and natural audio and video.

The drive continues today under the general name of Augmented Reality (AR). This aims to create applications where real scenes are integrated with synthetic objects that typically bring additional information enhancing specific real objects in the scene. AR content may also become "active" if the scene is designed to react to the appearance of a particular object represented by an image.

MPEG's Augmented Reality Application Format (ARAF) standard, under development through various versions, integrates a range of MPEG technologies and will foster convergence between a large number of application areas.

Convergence of the user interface

As content becomes richer with more media types and users can enjoy more combinations of media, user interfaces, too, evolve to become more functionality-rich, e.g. by including media types such as audio, video, 2D/3D graphics and rich media.

MPEG-U is a standard for widgets, small and dedicated applications that can be retrieved from different sources and aggregated to build effective and user-friendly interfaces. The MPEG-U technology is neutral to the application domain and offers more opportunities to extend convergence in the typically application-specific domain of the user interface.

Convergence of "other data"

When there is an established set of media types, one can always think of adding "other" media. This was the case of motion pictures added to photos, audio added to pictures, content matched to user's taste added to give interactivity and so on. This is also the case today when a whole new series of media types are added that can influence the environment in which content is consumed or to sense the environment and provide content that is more suitable to it.

The Media Context and Control (MPEG-V) standard includes a range of novel technologies to interface media consumption environments with sensors and actuators providing a more engaging user experience. Some environments – especially mobile – are particularly well suited to an early exploitation of such new media. No doubt others will follow sharing the same converging technologies.

Video processing convergence

So far successful digital media technologies have cleverly exploited the nature of audio and video information without digging too deeply in the nature of those signals. MPEG-7, a standard with a lamentable lower success so far than other MPEG standards, offer hints about the semantics of the audio and video information and provides a wide set of descriptors created from automatic audio and video analysis techniques.

Things are changing with progress of technology. MPEG is currently developing the Compact Descriptors for Visual

Search (CDVS) standard. This aims to extract parameters from an image that are closely connected to the objects in that image and can be transmitted efficiently to a remote end. For instance a user can take an image from the real world, have his mobile handset extract and send CDVS descriptors to a service with a request for information related to the image as represented by descriptors. An alternative use of the technology is the creation of automatic links between video scenes. Still another is replacing the image triggering an AR application with its CDVS descriptors.

Another convergence trend is looming. Autonomous vehicles may still be a thing of the future but equipping cars with cameras that bring pre-processed visual information of the environment around the car to the driver is within reach and can achieve wide deployment if suitable standards exist. While the CDVS technology is being designed for communication purposes it can also serve as a basis for extensions that can support automotive applications.

Convergence in the third dimension

The dream of giving users at a remote place a full 3D audio-visual experience has been around for a long time and a range of technologies have been tried with different levels of response on the part of users. Holography does not appear to be around the corner but devices for some environments could be the first to offer satisfactory user experiences to consumers and more devices for more environments are bound to follow. Seconding that trend MPEG is actively developing a range of standards for 3D Video and 3D Audio balancing the need for practical solutions now versus support of more ambitious features that anticipate technology.

Convergence of networks

The race to the global digital network of the early 1990s ended with the unqualified success of the internet. Originally designed to let computers exchange handfuls of characters asynchronously, the network is now more and more used to deliver real time video and audio. Unfortunately the internet is unable to natively offer a match between location of content and the means to delivering it in a guaranteed fashion and such overlay technologies as peer-to-peer and content delivery networks try to make up for the absence of these key network functionalities.

Information Centric Networking integrates some current overlay functions in the network. The Euro-Japan GreenICN (http://www.greenicn.org/) project is a great opportunity to accelerate the normal technology development cycle in the traditionally slow infrastructure domain and provide a robust future-proof networking solution that responds to the current trend of more real-time media at higher bitrate.

Software convergence

Progress of programmable components is so fast that a lot of functions that used to be implemented in rigid silicon architectures can now be more easily implemented in software running on a CPU. The MPEG Reconfigurable Media Coding (RMC) standard allows a transmitter to communicate to a receiver the algorithm that will be used to transmit video or 3D graphics information and constitutes the ultimate frontier of media coding.

A similar software convergence approach is witnessed in the lower layers with the Software Defined Radio (SDR) technology that offers further flexibility in the exploitation of the physical layer of communication channels where the transmitter can design the most suitable modem for the specific channel and communicate it to the receiver.

Middleware convergence

As software implementations keep on extending their reach, the role of the software layer called middleware, part of or above the operating system and that includes most technologies considered in this paper, becomes more important.

The Multimedia Service Platform Technologies (MPEG-M) standard represents the point of media technology convergence because it provides Application Progamming Interfaces (API), a defined set of media-related Elementary Services and the means to create new services by aggregating Elementary Services.

User description convergence

Media exist because humans are keen to consume them. So far the relationship between media and humans has been in the hands of the individual consumers and, increasingly, in the hands of service providers characterizing media consumers with profiles. More recently social network services have been attracting billions of users who seem to be happy to share information on intimate aspects of their lives while service providers collect and re-use wealth of information in their hands.

This liberal use of subscribers' profiles on the part of service providers is the object of countless controversies spanning all sorts of services and countries, and the level of friction is only bound to increase. The MPEG User Description (MPEG-UD) standard under development will offer the opportunity to create a level play field and return to users the management of their most intimate and valuable assets.

Energy consumption convergence

Energy is what drives billions, and in the not-so-distant future, trillions of devices. So far industry and consumers have tapped energy as a practically inexhaustible asset, but the limited nature of our environment is slowly changing

consumer perception and will eventually change also industry behaviour.

MPEG has been investigating the requirements of a standard that would offer a video decoder the means to tell an encoder that video information should be sent using a suitable subset of coding tools that requires "less" energy for decoding and display. The "Green MPEG" technology – as it is called – can become the universally deployed converging device technology for increasingly energy-aware consumers.

Payment method convergence

Monetary value, be it in implicit or explicit form, drives many business scenarios involving digital media. Today the means to handle money are typically dependent on ad hoc interfaces and technologies, but lack of convergence actually undermines the capability of enticing more users to handle more media.

The MPEG-M standard that includes a Transaction Service that can be used to virtualise a transaction interface is a first step in this direction.

Industry convergence

Today telecommunication companies can realise their plans to provide video services to their subscribers. Broadcasting companies can serve their users better by squeezing more television programmes in the same bandwidth. Manufacturers of telecommunication devices can become manufacturers of CE devices. IT companies can add CE functionality to their products and have the chance to become successful CE companies. A single manufacturer can supply set top boxes for terrestrial, cable and satellite broadcasting, but also for IPTV and OTT services.

All this is possible because there is such an amount of standard, non-industry and non-country/region specific technologies that are available at Fair Reasonable and Non-Discriminatory (FRAND) conditions. Those planning products or services aiming at interoperability can do so quickly and economically from a plurality of competing suppliers.

Legal convergence

While rules to achieve successful technology-driven convergence are known and practiced, the same is not true for other non-technical domains. In the legal space we encounter examples that show how there is a long way to go to achieve such convergence, even within a single country.

In the United States the Audio Home Recording Act (AHRA) of 1992 is applicable only to CE devices and not IT devices. This may have made sense when the law was enacted, but its application to the RIAA v. Diamond Multimedia of 1999 no longer did, even though it allowed the manufacturer of the portable MP3 player Rio to escape the clutches of the plaintiff. This case also shows another trend working against convergence: judges, not elected members of parliaments, decide how existing national laws should be adapted to the digital world.

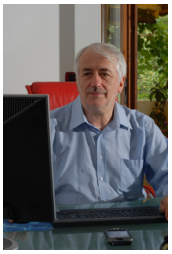
The record of making global laws has encountered mixed success. The WIPO Copyright Treaty, an international treaty laying down global rules on the use of copyright management and protection, has been ratified and is being enforced by a large number of countries. On the other hand the Anti-Counterfeiting Trade Agreement (ACTA), developed in secret negotiations without participation of civil society groups, developing countries and the general public, has come to a standstill. As much as MPEG experts could foster convergence by developing global technology standards with an open process, legal experts should develop legal standards of global impact in the open.

Another aspect working against convergence is the fact that patents protect an invention for 20 years while copyright generally protects a creative work for 70 years after the death of the author. It is ironic, but also a source of concern that, in spite of the very generous protection afforded by law to copyright, the global size of the patents business dwarfs the global size of the copyright business.

Conclusions

Digital media keeps on expanding thanks to the uniform technology basis created by MPEG that is accessible to all parties and keeps on adding more innovative technologies. This technology basis is fostering convergence by permeating traditionally different areas and extending its reach to new areas.

Technology-driven convergence in digital media has brought many benefits: an innovation-prone environment, many opportunities for diverse players and an ever-growing business fuelled by a constantly improving user satisfaction. The same cannot be said of other aspects that are integral to digital media use, such as legal aspects. These could and should be handled in a similar open process yielding advantages that are comparable to those obtained from technology and possibly more so.



Leonardo CHIARIGLIONE

Leonardo Chiariglione obtained his Ph. D. degree from the University of Tokyo in 1973.
During his career he launched several standardization initiatives, such as MPEG in 1988, DAVIC in 1994, FIPA in 1966 and the Digital Media Project in 2003.
He is currently CEO of CEDEO.net, a technology company developing new digital media services using its technology portfolio.
Dr. Chiariglione is the recipient of several awards: among these the Eduard Rhein Foundation Award, the IBC John Tucker Award, the IEEE Masaru Ibuka Consumer Electronics Award and the Kilby Foundation Award.