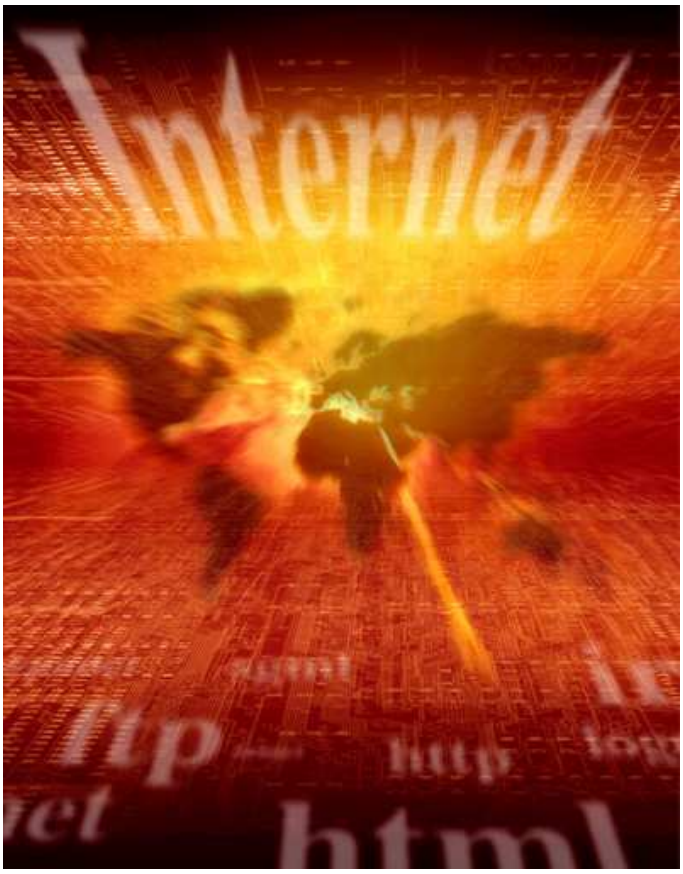
User Centric Media in the Future Internet



November 2009

User Centric Media cluster of EU projects

European Commission
Information Society and Media

Contents

1 Executive Summary

This document is targeted to readers who want to take part in shaping the Future Internet. It focuses on the media and content related issues and considers visions and megatrends for the future as well as research challenges on the way towards the Future Media Internet. The users are always in the centre of these considerations.

In the debate on how the Future Internet will look like, three powerful concepts drift to the surface, vying for attention: Users, Services and Content. Each of the three components presents itself as a powerful force that is able to explain recent evolutionary steps, and that claims the right to drive the Future Internet developments. The **User-Centric perspective** emphasises the end-user multi-sensory, embodied, active experience as the driving force for all technological innovation, observing how today the Internet is a network of active social users rather than a connection of devices. The **Service-Centric view** has roots in enterprise IT solutions as well as in the Web 2.0 mash-up culture, showing how valuable applications can be built faster and more efficiently if service components can be reused in flexible ways. The **Content-Centric view** refers to the central role that rich media content, including metadata of various nature, is playing in attracting users to Internet services, because content consumers are increasingly also content producers. It also refers to how the transfer of media content can impact the network operation.

This White paper on the User Centric Media reflects the consolidated opinion of the User Centric Media (UCM) cluster composed of representatives of ongoing FP6 & FP7 EU funded projects, under the aegis of the Networked Media Systems Unit of the Information Society and Media Directorate General of the European Commission[1].

Taking the end-users with their needs and desires as the initiating force for the design of the Future Internet and the applications it will support, in this paper we aim at describing our vision for the role of the UCM in the Future Internet, from a user-centric perspective.

First the main market trends are checked, and societal megatrends and disruptive technologies are identified. From them the following media related trends and visions are determined:

1. Convergence of real and virtual worlds (“friends meet virtually in social networking websites as well as in the local bar”) allowing for novel paradigms of active and social experience of media.

[1] "The views expressed are those of the authors and not necessarily those of the European Commission or any of its officials"

2. Convergence of different forms of media (films, news, games) and inclusion of more senses in media, such as touch, smell, taste, etc in the media and in the way of accessing it.
3. Parallel existence of consumable media (e.g. video on demand) and interactive media (e.g. games) allowing for active and passive consumption.
4. Increased mobility (at work and in private spheres) and the wish for "virtual togetherness" of families and friends to make up for the lack of real togetherness.
5. Always connected to the network to upload or download media and information "ubiquitous connectivity anywhere-anytime-anyhow" including for learning environments.

After reporting upon the media related trends and visions, in this paper, visionary application scenarios are sketched, one on the work environment and two on the digital lifestyle and home environment.

Moreover, the three dominant concepts of the Future Internet, (User-Centric, Service-Centric and Content-Centric) are analysed and compared, and the opinion of the cluster is given on how they can all harmonically coexist by introducing the concept of Content Objects which lead to a layer-less content-centric architecture design of the Future Internet.

Finally, three groups of potential research challenges are described in detail, which focus on:

Interaction/Personalisation/Identity, comprising

- socially Interactive Virtual Characters with High-Level Cognitive Skills,
- personality and Emotion modeling and Simulation,
- self-awareness, Memory and Interpersonal Relationships,
- expressive Behaviour analysis, modeling, and simulation,
- advanced Multimodal Interaction Technologies,
- embodied social media and the social dimension of interaction.

3D User Generated Content (UGC) comprising

- creation of realistic 3D UGC,
- intelligence in 3D UGC,
- interaction and non-verbal expressive communication,
- collaborative creation of 3D UG worlds,
- interoperability and standardization.

Immersive Media, comprising:

- tele-Immersion environments,
- virtual contact and navigation,
- multi-modal, multi-party interaction among real users and virtual characters,
- sonification and sound-based interaction.

2 Visions and scenarios

Information and Communication Technologies (ICT) are moving forward at an incredible pace. The point of no return into a global modern knowledge-based society has long been passed. Technology cycles are becoming so short that it is difficult for the users to keep pace. Particularly in the networked media sector there are many societal and technological changes ahead. There is a large business related; the size of the global media and entertainment market in 2008 was 1.7 Trillion USD[2] and is predicted to grow to 5.7 Trillion USD in 2024[3].

With no doubt, video/audio market and broadband market are two key business factors to foster competitiveness and create new scales of economy. The provisioned key technologies are, among others, 3D video technology, 3D audio/sound technology, multi-sensory communications technology, perception/ transmission technology of human emotions and sensibility. As shown in Figure 1, these technologies are expected to represent a huge potential market of up 4,900 Billion Euros world-wide by 2020.

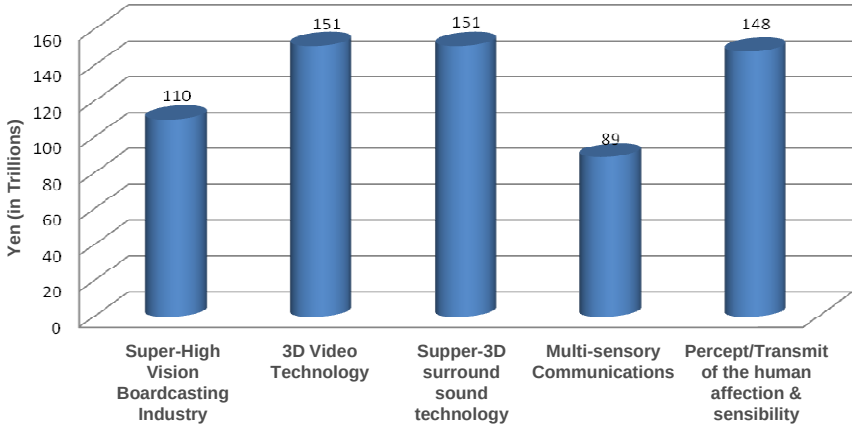


Figure 1. Media delivery technologies market roadmap[4]

There are already many factors that clearly show the great potential of networked media in the audio/video media/content market. First of all, the (user generated) content stored or streamed to home systems is growing exponentially, the users are changing their content consumption patterns and the percentage of users/subscribers

[2] source: PricewaterhouseCoopers

[3] source: Future Exploration Network

[4] K. Enami, "Study Group report on the economic impact of Universal Communications," NEM Summit, St. Malo, Oct. 2008

generating personal content grows rapidly. Another important factor is the significant percentages of Internet users, who are now paying for online content experiences. As more mainstream consumers begin to utilize Internet content services as complements to their existing subscriptions (broadband, television, mobile phone, etc.), it is expected that we will observe a growing number of users paying for these services as value-added services rather than stand-alone subscriptions. Also, the links between content and the platforms from which consumers will search, access and enjoy the abundance of content and producers will advertise and offer their content are also becoming stronger.

Virtual world usage among children in the US is already quite strong and getting stronger. eMarketer estimates that 37% of online children ages 3 to 11 use virtual worlds at least once a month. By 2013, it will be 54% of children (Figure 2). They represent a significant population of Future Internet users.

US Child Internet Users Who Use Virtual Worlds, 2008-2013 (millions and % of total)

	2008	2009	2010	2011	2012	2013
Child Internet users	15.6	16.1	16.6	17.0	17.4	17.8
Child virtual world users	**4.4**	**6.0**	**7.0**	**7.8**	**8.7**	**9.6**
% of child Internet users who use virtual worlds*	28.0%	37.0%	42.0%	46.0%	50.0%	54.0%

*Note: ages 3-11; visit at least monthly; *calculated based on unrounded figures*
Source: eMarketer, April 2009

103112 www.eMarketer.com

Figure 2. US child users who use virtual worlds

Taking into account the aforementioned facts, the experts within the User Centric Media (UCM) Cluster of related EU projects carry out an educated look into the crystal ball and try to predict how the media part of the Future Internet could look like in 2025, and which are the related challenges to R&D and to the society.

The main reason for this exercise is to help developing European RTD work programmes in a way to ensure that Future Media Internet related societal and technical challenges can be addressed in the most efficient way.

2.1 Visions

Societal megatrends and disruptive technologies

It is very difficult to predict how the Future Media Internet will look like in 10-15 years from now. This depends not only on the technological developments, but very much on how society will develop and which possible Information and Communication services the users are willing to finally adapt. The best way for tackling the issue is probably by looking at societal megatrends and at disruptive technologies (*c.f.* the

table below). Disruptive technologies have always been the main driver for changes (for example from centralised steam engines to distributed electro-motors). From those megatrends and disruptive technologies we can derive a list of likely trends in the media related areas.

Societal megatrends	Disruptive technologies
Being always connected	Communication access with high band-width available anytime anywhere, also while being mobile
Increased mobility, both concerning pri-vate and working lives	Wearable communicating computing power
Less distinction between virtual and real worlds	Natural 3D based displays, spatial audio and multi-sensory communications
Generational change towards information and communication technologies-committed people	User generated content available any-where anytime
Increased user involvement and user power; sharing and using user generated information	Personalisation of services, media and devices; natural interfaces, including af-fective interfaces
Life-long learning	Cheap interconnected and communicat-ing things (e.g. RFID). Open access to world wide archives with high quality content
Keeping in touch with others (close & less close friends, communities, etc.)	(Multimedia) Social Networking
Global creative collaboration of individu-als by connecting minds and creativity on a scale never imagined before.	Intelligent Media Recommender to con-nect people with ideas and content. Convergence of intelligent content and connectivity research and technologies

Media related trends and visions:

From the table above we can derive the following media related trends and visions:

- Heavy increase of user generated content (UGC) and new, innovate ways of its use (“everybody can be a director on the Internet”). In the past five years (from 2005 to 2009) UGC increased more than 12 times
- Convergence of real and virtual worlds (“friends meet virtually in social net-working websites as well as in the local bar”)
- Convergence of different forms of media (films, news, games) and inclusion of more senses in media, such as touch, smell, taste, etc.

- Parallel existence of consumable media and interactive media (allowing for active and passive consumption).
- Development of novel paradigms of active and social experience of media in the converged real and virtual worlds
- Access to media through natural, multi-sensory, disappearing interfaces, including affective and social interfaces capturing human non-verbal emotion and social behaviour
- Increased people mobility and the wish for "virtual togetherness" of families and friends to make up for the lack of real togetherness.
- Increased numbers and possibilities of participation in (ad hoc) "knowledge communities". Users will gain knowledge and consumer power.
- Virtual learning environments replace "physical" learning sites.
- Increase of "mobile offices" and "work on the move". Employees will have looser links to a specific employer.
- Augmented reality in development and manufacturing.
- Always connected to the network to upload or download media and information "ubiquitous connectivity anywhere-anytime-anyhow".

The following three scenarios, one focusing on the business, and two on the digital lifestyle, are illustrating these visions.

The functional requirements, challenges, opportunities and possible barriers derived from these scenarios are being used as an input for developing the research challenges, presented in more detail in the further sections of this document.

2.2 Scenarios

2.2.1 Work environment and future business impact

One of the main characteristics of today's economy is globalization. Currently, that means many people from different places are roughly linked together. Advances in media and communication technologies will allow people in these different places to actually feel as if they are together. The following scenario illustrates it (Figure 3).

John is an engineer who lives in Berlin and works for an automotive industry. He is working on the design of a new model of a car. The design is being done collaboratively between different engineers located all over the world. A meeting has been arranged to discuss the new design in a hi-tech room, which produces a real immersive environment. There he will meet and interact with the other designers who will enter similar rooms in other countries.

As John enters the room he puts on a haptic glove and enters the virtual environment. At the same time, the other engineers from Detroit and Tokyo join other hi-tech rooms there and also enter the environment. In this environment all of the designers are able to see, feel and touch the car. They can discuss together the design, as each of them can see and interact with a virtual mock-up of the car. This allows them to quickly see

not only what the car will look-and-feel like at large, but it enables them to inspect also the internal details of the car – the mechanics, the electrical systems, different components, etc. They can easily see these different systems and subsystems and their interconnections while the car is at rest or in operation.

Figure 3. Tele-immersion in collaborative working environments

John and the other engineers not only work more efficiently on the design of the car but also get to know each other better because they feel as they would all working together in the same room. They can see each other, talk to each other and inspect or touch parts of the new model of car in a much more informal, pragmatic setting. They can advance much more quickly the design of this new car, comparing to traditional working methods., Moreover, even though they are from different countries and represent different cultures, they will be able to communicate with each other more efficiently than if they had not had these shared immersive experiences.

This scenario shows some of the possibilities of augmented and virtual reality in the business sphere – more efficiency, less travel, better social interactions, - but the impacts of this will be felt far beyond just the business world, for example less travel means less pollution and lower energy consumption.

2.2.2 Digital lifestyle and home environment

The following two scenarios are concerned with the influences of emerging mega-trends and major disruptive innovations in user centric media, which impact the digital lifestyle and home environment of people living in 2025.

"A day in the life of Sara"

We witness a day in the life of Sara, a British 25 year old student of the social sciences, living in Berlin, Germany. It is a beautiful April morning in the year 2025.

Sara wakes up and is in a very good mood as this is her first free weekend after her last examination. She is looking forward to have a relaxed morning with one hour spare time before she starts a shopping walk in the city and meets with her best friends: Patricia and Marilyn arriving from London later in the evening.

As her home is equipped with the latest media and entertainment technology supporting gesture detection, emotion sensing, advanced 3D visualizations and wireless sensor and actuator network involving almost all human senses, she starts playing her favourite music song just by a snip of her fingers. The musicians are being displayed as a holographic projection inside her living room. The system creates a smell of sun lotion together with a breeze of the Caribbean Sea which makes her feel even more like having some time off from her studies. The system can also analyse her current mood and activities and recommend the most suitable music according to her preferences.

After taking a shower she wants to know about the latest fashion articles for her shopping walk. She directs her home media and communication system by her voice and with a gesture to show a catalogue of great sale offerings in the shop close to the city centre (Berlin Mitte). After making a selection she puts on her body suit which is a location-aware, mobile entertainment and multimedia device. Her body suit allows her to interface with her home media and information system. As she is always connected to the network she does not need to synchronise any data and all her personal information is available everywhere and any time. Her body suits is endowed with an emotion sensing garment, that can measure her emotional behavior and make the home media and information system always suggest the content and information which is best suited at that time.

Figure 4. Emotion sensing garment

After her shopping walk she is notified by her body suit of a music concert that is about to happen at this particular moment very close to her current location - in Berlin Tiergarten. As the information has been propagated by her fellow student network she is delighted to join the jam session as she likes an idea of spending some time together with her fellow students listening to this live concert. In some of the pieces being played, she can even actively participate, together with her friends in the concert. For example, through her non-verbal expressive gestures and movements and her social interactions with her friends she can control music instruments, establish a dialog with the performers, or mould the music which is being played. As the media system in her body suit allows for 3D audio and visual recording together with an unlimited and lifelong storage, she creates a lasting impression of this interesting event that she shares with her fellow students and other friends.

On her way back home she stops at Flughafen Schoenefeld to welcome her friends coming from London and she is disappointed to hear from Patricia that Marilyn could not make it because of her mother sudden sickness. Patricia and Sara decide to stay

at home that night and watch a movie together before they go out to enjoy Berlin night life together. Sara's home media system is capable of providing context aware media recommendations considering the time of the day, Sara's current mood, and the fact that Sara is about to watch a movie together with her friend, and not alone. The newest movies available for them are listed and filtered dynamically based on the Sara's and Patricia's personal preferences. Even though Marilyn could not afford to visit Sara in Berlin because of her mother, she joins their movie session virtually. She is projected as a 3D hologram into the living room of Sara. The friends can talk to each other and watch the movie together in a virtual space as if they would watch movie together in a physical space.

Now it is time to go out in Berlin and Marilyn is a little sad as she can not join her friends. However, Sara will of course capture the night with her media enabled body suit and share the night with Marilyn later on.

The scenario addresses a series of megatrends and disruptive innovations or RTD challenges: the megatrend of being always connected with communication access and high bandwidth available anywhere, anytime and with any device, as well as a megatrend of an increased mobility with wearable and non-intrusive communication and computing power implemented a form of e.g. an emotion sensing garment. The scenario also demonstrates a less distinctive and converging real and virtual world with an increased user involvement using natural 3D holographic displays techniques, spatial audio and advanced methods for generating media content by consumers. Moreover, enhanced techniques and a new generation of recommender systems that allow for an advanced personalization of future media content in a highly dynamic, user centric and context aware fashion are illustrated in this scenario.

2.2.3 Convergence of real and virtual world and impact on social interactions

"Anne meets her friends"

Anne is a forty year old working mother and she is living with her family in a nice house in the suburbs of Paris. After work she returns home and takes care of the children. She loves exercising and as soon she has a break in her daily routine, she joins her virtual 3D Health Club 2025 from home or even sometimes from a hotel room on one of her casual business trips. As soon she enters her health club she gets a notice that three of her friends, Maria, a teacher from Helsinki, Peter, a student from Rome, and John, a consultant from London, are in the club. They greet each other and decide to exercise together. They chose their favourite music and start their aerobics together. They see each other and can talk with each other. The remote friends experience the virtual health club as a natural one, including its temperature, smells, sounds and the lightning conditions. Anne, Maria, Peter and John communicate and interact in a natural and intuitive way with each other. They synchronize their fitness activities and have fun together.

During their exercising break, Peter tells his friends that he has got tickets for the world football championship. The other three are very busy working people and have no time to travel with Peter to the event, but they would love to share the experience

with him. As soon Peter arrives in the host town of the football championship, he connects with his friends and shares with them audio-visually in real time the football fever of local fans in the streets and pubs in the town. Then the three friends beam themselves together in the 3D Football Stadium and enjoy the event with Peter, who sits among the real crowd. The excitement of the live game can be felt by all four of them.

This scenario focuses on the impact on social interactions by converging the real and virtual worlds. It describes the wish for "virtual togetherness" of friends to make up for the lack of real togetherness by meeting virtually in health clubs and sharing remotely the experience of a live sport event by using novel 3D video technology, 3D audio/sound technology, multi-sensory communications technology and perception/ transmission technology of human emotions and sensibility.

An important consideration implied by the scenarios presented above concerns the anthropological, sociological, and psychological issues emerging from such dramatic changes of the living environments for human beings: e.g. the convergence of virtual and real and the novel paradigms for social communication will require careful studies to anticipate behaviour and social diseases possibly caused by technology and complex realities it enables. Novel techniques of analysis, validation and evaluation will be therefore required, grounded, among the others, on the emerging scientific discoveries on neurophysiology and brain sciences.

3 Will the Future Internet be Content Centric?

In the discourse about Internet evolution, assessing and analysing past trends on one hand and imagining and pushing future scenarios on the other hand, three powerful concepts drift to the surface, vying for attention: **User, Service** and **Content**. Each of the three presents itself as a powerful force that is able to explain recent evolution and that claims the right to drive the Future Internet. The **User-Centric** perspective emphasises the end-user experience as the driving force for all technological innovation, observing how today the Internet is a network of active social users rather than a connection of devices. The **Service-Centric** view has roots in both enterprise IT solutions as in the Web 2.0 mash-up culture, showing how valuable applications can be built faster and more efficiently if service components can be reused in flexible ways. The **Content-Centric** view refers to the central role that rich media content is playing in attracting users to Internet services, and the fact that content consumers are increasingly also as content producers, and how the transfer of media content can impact the network operation.

As the three views are emphasising different aspects rather than expressing opposing statements, the User-Centric, Service-Centric and Content-Centric perspectives are not exclusive ones but complementary and should be merged into an all encompassing perspective to make the right design choices for a Future Internet.

3.1 Complementary Perspectives in a Layered Architecture

The classic tool of a layered architecture can be helpful to illustrate the complementarity of the views and to understand cross-layer influences. Figure 5 is an adapted version of a figure in Future Content-Centric Internet Position Paper[5], including the User-Centric dimension as an additional layer.

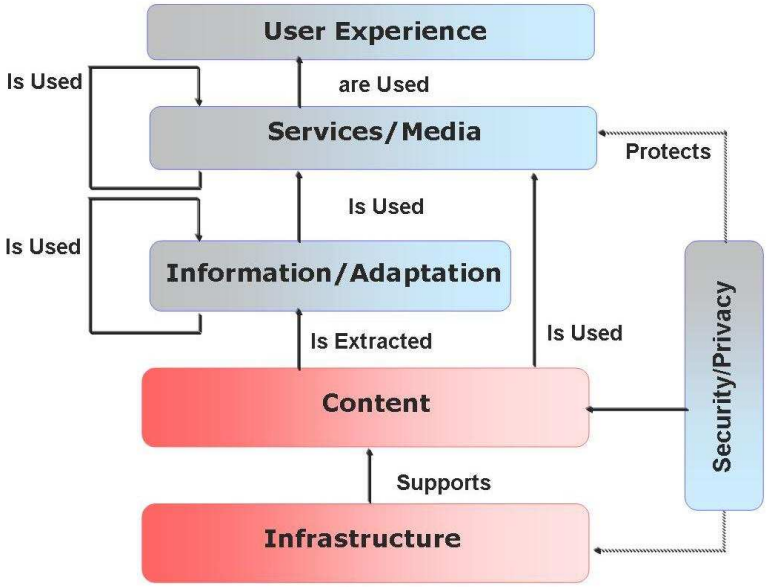


Figure 5. Content-Centric Layered Architecture

The different elements of this architecture are the following:

- ***Infrastructure*** (both private and public) will consist of content transport, storage and processing functions, which are fast, accurate and dependable, and are operational in a highly distributed and scalable manner.
- **Content** is any type and volume of media. Content may be pre-recorded, cached or live, static or dynamic, monolithic or modular. Content may be combined, mixed or aggregated to generate new content and **Media**. It may vary from a few bits (e.g. the temperature that a sensor has measured) to interactive multi-media sessions and immersive complex and multi-dimensional virtual/real worlds' representations. Of particular importance is the role of metadata, which can be of various nature, including data on "how" content may be manipulated, embodied, transmitted and perceived.
- ***Information*** is the product of a number of functions applied to the *content* or recursively to the *information*. By publishing, discovering, combining, mining, rea-

[5] Content-Centric Internet Position Paper: "Why do we need a Content-Centric Future Internet? Proposals Towards Content-Centric Internet Architectures", European Commission, Networked Media Unit, Information Society and Media, May 2009

soning upon, aggregating *content* and pieces of *information*, new *information* may be extracted or generated.

- ***Service*** is the result of a set of functions applied to the *content*, to pieces of *information* or recursively to *services*. By (manually or automatically) handling, managing, combining, personalising, adapting *content*, *information* or *services*, new *services* may be composed or generated.
- ***Media*** is anything that a human can perceive/experience with his/her senses.
- **Security Privacy, and Trust** will be a property of *content*, *information*, *services* and Infrastructure, allowing much more efficient control over *content* objects.
- **User Experience** is the interaction between the service and the user, representing information and retrieving it, satisfying a user's need or desire within an appropriate social/anthropological context.

3.1.1 Impact of the User-Centric Perspective

Taking the end-user with his/her needs and desires as the initiating force for the design of the Future Internet and the applications it will support, we consider the following requirements that will have an impact on the **Service/Media** layer:

- The end-user is the endpoint, rather than his or her device. Users should be able to easily find each other and engage in interactions, even if they use multiple devices in parallel without falling in the trap of limiting the user to one single identity.
- Universal accessibility for services and services evolvement with user experience; various users will approach the offered services with different levels of competency, and this level will evolve over time as they make use of the service.
- Universal accessibility for content generation; users more than ever act as content producers, implying that if a Service-Centric network offers service components for end-user experience creation, they should be usable by all users and not only professional ones
- If the network is enhanced with a certain intelligence to optimize the user experience, this should not lead to a feeling of loss of control with the user, dealing with an unpredictable environment.

3.1.2 Impact of the Service-Centric Perspective

Although the Internet is supporting a wide variety of applications, several functional building blocks are common between large groups of applications. The Service-Centric perspective therefore argues that it makes sense to design a network environment that supports the flexible creation, publishing, discovery and use of common service components. Flexibility here refers to easy location-independent detection and invocation of service components. Just-in-time inclusion of service components – i.e. at the moment of the creation of the end-user experience – allows optimization of net-

work services and supports rapid innovation. The common service components that are typically listed include user identification, authentication and authorization, security and DRM, bandwidth management, storage, power management, payment, location and time context information, user activity, content adaptation, search and indexing functions. Some of these – user identification, authentication and authorization, content adaptation to devices and user context – can be seen as driven by the User-Centric perspective.

3.1.3 Impact of the Content-Centric Perspective

The Content-Centric perspective highlights the driving role that rich multimedia content has played and is expected to play in the growth of the Internet, in terms of usage and traffic. The web has become a true Media Web, and the volume of transferred content will continue to rise sharply, as the quality of the media content further increases (High-Definition and Ultra High-Definition Content, 3D and stereoscopic content, multi-view content etc), as more experience of content becomes active and social and as more users evolve from mere consumers to active creators and/or repurposes of content. The Media Web is furthermore evolving to a Real-Time Media Web with live content streams and multimedia person-to-person or group communication. This can be a separate application experience or it can be embedded in frame experiences like gaming, education and users collaboration.

To satisfy user experience, a Content-Centric Internet will depend on the realisation of a set of content-specific network services. In this way, the Content-Centric perspective adds new service components to the Service-Centric view - indispensable for media experiences, or emphasises already identified ones. Content-Centric services include content distribution networking for both on-demand and live media distribution, content publishing, discovery, adaptation and processing services, DRM services, conferencing services, media annotation, indexing and search services.

Figure 6 schematically depicts the convergence of the three different, yet complementary perspectives.

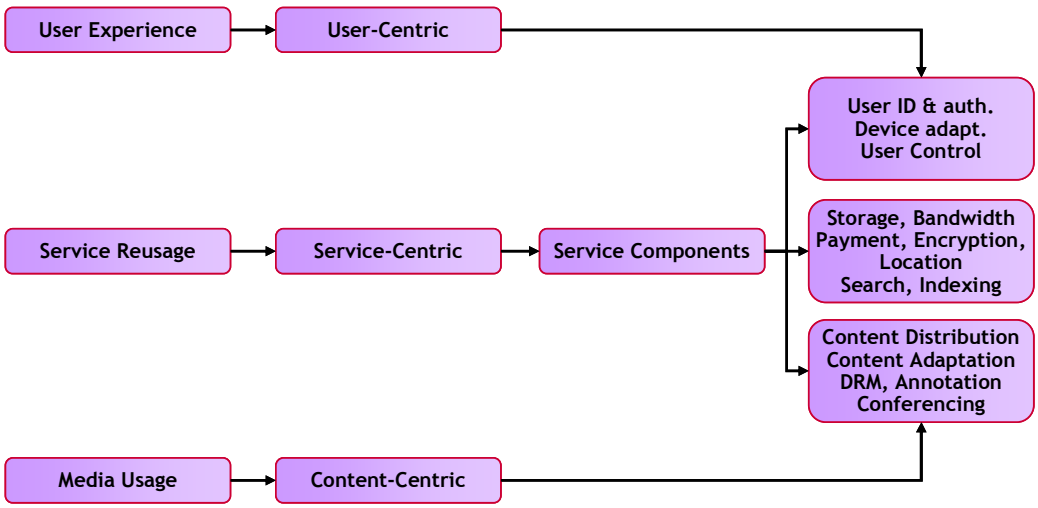


Figure 6. The convergence of three different perspectives

3.2 The Concept of Content Objects

3.2.1 Introduction

The Content-Centric Perspective paragraph above discussed the evolution of media in terms of quality, usage and traffic volume, but it didn't address the format of the media content. Currently, media content is the result of an off-line, cumbersome and lengthy creation process, whereby content components are composed into a meaningful and appealing presentation. The distribution over the network for consumption is then the transfer of the finalised complete media presentation in the form of bit streams, followed by a play-out at the end-user's device.

The key concepts for the Service-Centric perspective as explained above are the identification and separation of meaningful service components and the just-in-time on-the-fly flexible integration of such components into an application experience. It is expected that this evolution for software and network functions will also take place for rich media, i.e. that media experiences will be created as the just-in-time composition of content component objects that are easily located, synchronised, reused and composed.

Such an approach can already be discerned in virtual world applications where users contribute to the content creation: the virtual world representation on the end-user's device is the composition of objects that have been created by various authors and are fetched as they are required for representation.

Figure 7 represents a possible mixed-reality scene for a person-to-person interaction, combining stored and live media objects from a multitude of sources and engagement of many senses..

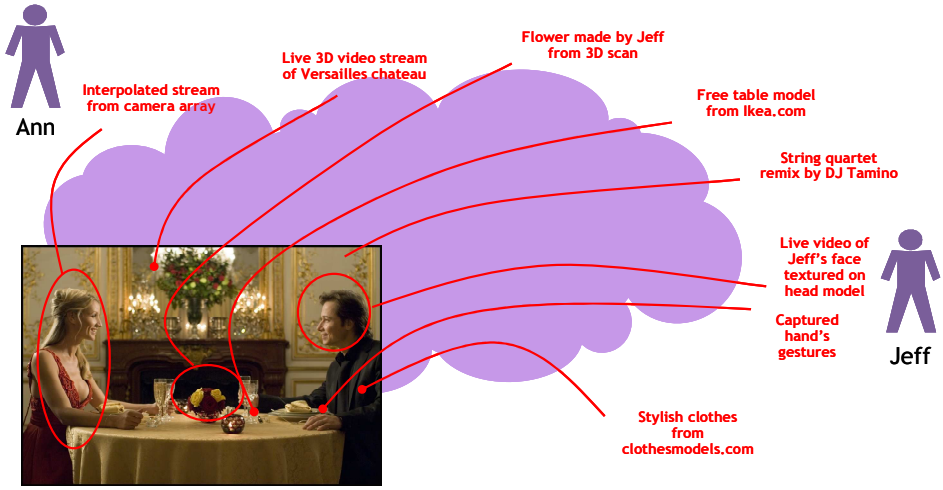


Figure 7: On the fly generation/reconstruction of semantically enriched worlds

The availability of the constituent content objects and their spatial and temporal relationships, rather than an opaque stream of pixels and audio samples, opens up new opportunities for content creation and consumption:

- Re-use of components from existing content for the creation of new audio-visual content becomes much less cumbersome, allowing fast and easy media mash-ups.
- On-line collaboration to edit audio-visual content prepares the way for a true media Wikipedia.
- Personalisation enters a new stage, evolving from a selection of prepared content to a just-in-time composition.
- The insertion of stored audio-visual content into real-time communication is greatly facilitated.
- The combination of captured audio-visual content with synthetic 3D content creates exciting mixed-reality experiences.
- The possibility for the user to actively intervene and mould content components through natural, non-verbal interfaces enhances the experience and provides authoring capabilities to the users, e.g. by enabling them to reshape, personalize, and re-experience in unique ways audiovisual content, including layered metadata.

3.2.2 Repercussions of content objects in a classic layered architecture

Taking a similar approach as above, we can specify more advanced content-centric service components to deal with content objects. Service components in the Service-Centric network will have to deal with the object-nature of content, supporting functions like object publishing and discovery, prioritisation, synchronisation, cloning, scaling and transformation, merging etc.

3.2.3 A forward-looking alternative for content objects: ALLOA (Autonomic Layer-Less Object Architecture)

The classic layered approach may not be the ideal match for the content object vision: the advanced content treatment service functions that are required may exhibit characteristics that differ substantially from the non-content-driven service components, leading to the definition of service components that are positioned in a blurred area between content, service and user layers.

An alternative is a clean-slate approach for the network design, starting from the content object itself, a content-centric network architecture, the **Autonomic Layer-Less Object Architecture (ALLOA)**.

In Figure 7 we have already introduced the concept of content objects, which can ad-hoc, on the fly generate/reconstruct semantically enriched 3D augmented/ virtual worlds in order to create an orchestrated immersive media experience. Here we further expand this concept to "***Content Objects***".

A content ***object*** is a polymorphic/holistic container, which may consist of media, rules, behaviour, relations and characteristics or any combination of the above.

- ***Media*** can be anything that a human can perceive/experience with his/her senses (a dancing person, the second violin in a symphonic performance, a tear on your cheek).

- ***Rules*** can refer to the way an object is treated and manipulated by other objects or the environment (discovered, retrieved, casted, adapted, delivered, transformed, and presented). Rules can for instance be used to specify if the media in the object would allow rescaling and that it would accept a delivery delay of 2 seconds, but that it should certainly arrive for presentation at the end-user side before a child object: the object knows its purpose in the integrated media experience and therefore its priority for transfer. Also the options for manipulation by the end-user at the moment of presentation could be included.

- ***Behaviour*** can refer to the way the object affects other objects or the environment.

- ***Relations*** between an object with other objects can refer to time, space, and synchronisation issues. Relations could for instance describe that an audio object of a singing person is related to an animated 3D model of the singer and that lip synchronisation is required.

- ***Characteristics*** meaningfully describe the object and allow retrieval of its related objects: user interaction with a 'coq-au-vin' object may visualise in the immersive 3D environment the ingredients and their current prices or may lead to the ad-hoc building of 3D replicas of the restaurants where the dish is available.

Objects can be ***hierarchically organised***, like the constituting instrument channels of a music band, and can trigger the generation of new objects. An object can be ***divided/ spit*** into new objects or multiple objects can be ***combined/merged*** and finally create new objects, and these operations may happen while objects are "travelling" over the network.

An object can be ***cloned***. The clone keeps the characteristics of its "parent" object but knows that it is a clone. This is also associated with issues like cashing (object lifetime, check for updates) and Digital Rights Management (DRM).

The cloning has implications in the opening of novel business models around "actively experience audiovisual content": for example, the same audiovisual content can be distributed with different characteristics: for example for a music, ranging from a simple MP3 file to a more complex music content in which the user has authoring capabilities to reshape the music piece, thanks to the availability of further metadata and higher level representations enabling this user a number of degrees of freedom in real-time, e.g. in terms of re-orchestrating the music, re-arranging (post-production), shared (social) active fruition.

The autonomous objects will travel over the network, split and combine to generate the new service or a virtual world object. The Future Content Centric Internet will support the content objects in order to meet their relations.

It is currently very difficult to imagine what a network architecture that support objects would look like. An attempt to map the characteristics of the layered approach which is depicted in Figure 5 into the novel “layer-less” concept of the object is shown in the Figure 8, where one or more layers are mapped to one or more characteristics of the object.

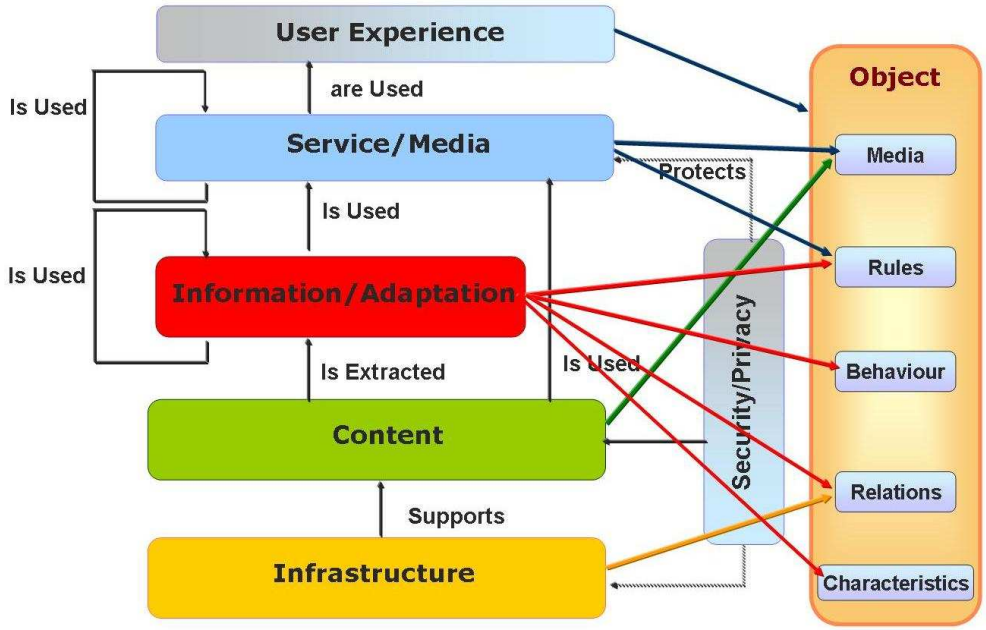


Figure 8: Mapping a layers-based Content-Centric Internet Architecture into “Objects”

More specifically, transfer and integration of objects for the purpose of the creation of an orchestrated “Media” experience clearly demands intelligence that combines application (“Service/Media”) and “Content” information. The intelligence could be embedded in the objects themselves, retrieving information from the network and providing instructions for routing and transformation, or the intelligence could be hosted in network nodes that attempt to satisfy the requests of the objects as they are described in the ”Rules”, “Behaviours” and “Relationships” (which take input from the “Information/Adaptation”, “Content” and “Infrastructure” layers). Finally, the “Characteristics” that meaningfully describe an object take, mainly, input from the “Information/Adaptation” layer.

3.3 Challenges

From a user-centric media perspective, we can deduce the following challenges from the discussion in this chapter:

1. In a classic layered model, and with the currently used content formats, there is a need to fully identify and specify the service components a network will have to support so as to satisfy the requirements from a user-centric and content-centric perspective.

2. Still in the classic layered model, but taking into account the emergence of content objects that support the on-the-fly generation of rich media experiences, there is a need to identify and specify the additional service components requested by the object-nature of future object. This challenge also includes a critical assessment of the suitability of the classic layered approach when confronted with a wide range of advanced content-specific requirements.

3. A further exploration and elaboration of the Autonomic Layer-Less Object Architecture.

4 Research Challenges

Based on the aforementioned analysis we foresee the need for three big groups of research challenges which refer to interaction & personalisation, 3D user generated content and immersive media. For each of these groups specific topics are identified and analysed in detail.

4.1 Interaction/Personalisation/Identity

Socially Interactive Virtual Characters with High-Level Cognitive Skills

Today, the evolution of virtual characters is towards “social companions” that can interact with people in a natural way just like humans and that can develop long-term relationships with them. Recently, there has been strong research on the creation of realistic virtual characters with the improvements in facial and body animation, hair and cloth simulation which has been applied in movies. However, today the challenge is not only making them look like humans but make them behave as natural as humans. Thus, they should be equipped with properties such as social and cognitive intelligence, personality, emotions and memory in order to engage the people to the interaction and to be entertaining or useful enough for them depending on their needs and desires.

Personality and Emotion Simulation

Emotions have proven effects on cognitive processes such as action selection, learning, memory, motivation and planning. Our emotions take role as motivational factors

for our decisions and affect our actions. For engagement during interaction with virtual characters, it is important that they should express emotions and exhibit distinctive personality by using different knowledge and different repertoire and manner of verbal and non-verbal behavior. Emotional awareness and sensitivity to user's emotions through emphatic reactions is also important for many applications such as education and health care.

Self-awareness, Memory and Interpersonal Relationships

As virtual characters live with us in our daily lives at home, at our car or at the office, it is important that they have the memory capability in order to remember our past interactions and have the perspective of self to learn from their experiences and to reason about them. Another very important factor that shapes our every day emotional reactions is the status of our interpersonal relations with other people. This concept becomes important when the conversational partners are coming together multiple times forming a long-term interaction which has high potential importance for motivating people for the interaction with the system.

Expressive Behavior

Virtual characters are simulated via an interrelated assembly of several components such as conversational abilities, facial expressions, hand-arm gestures and eye-gaze control. Their behavior should be expressive enough and should be linked to and driven by their emotional state, personality as well as the social norms that's exist in human-human communication. Various techniques have been developed so far for the simulation of human-like facial expressions and gaze which have produced very good results. However, they are still far from being intelligent and expressive enough for full engagement of user to the interaction.

Advanced Multimodal Interaction Technologies

Adaptivity and Multimodality are two major characteristics of the Future Human-Machine Interface. Estimating the improved human models for understanding user expectations, needs, desires, behaviours and user's natural interaction and how that impacts interface design is a key aspect for the design of Future Internet technologies. Indeed, the introduction of the Future Internet concept poses an additional set of challenges in terms of Multimodality. Conventional modalities such as gestures, speech, and video should be expanded / enhanced with advanced interaction modalities such as emotional user interaction, multi touch interaction, motion sensing and tangible user interfaces. Another important challenge regarding multimodality is the adaptation and integration to mobile devices of the all the next generation interaction modalities taking into account both power and ergonomic issues.

Embodied social media and the social dimension of interaction

While we assist to the worldwide spreading of social networks, many existing multimedia interactive systems and Internet applications are still intended for a single user and social interaction is often neglected. Current research on social interaction does not focus on the high-level emotional aspects, but rather on lower-level issues such as

group cohesion and decision-making. In this framework, pioneering studies developed techniques to measure social signals in scenarios like salary negotiation and friendship. Particular attention was also directed to the recognition of functional roles (e.g., most dominant people) played during small-group meetings. These works, however, are often based on laboratory experiments and do not address the subtlest aspects of social interaction such as emotional synchronization and empathy. Empathy, in fact, has been studied mainly in the framework of synthesis of (verbal) dialogues by virtual characters and embodied conversational agents. Recently, efforts have been devoted to analysis of social behaviour of professional music performers (e.g., a string quartet) and to the development of techniques for social active listening to music by mobile devices, i.e., for allowing a group of users to mould collaboratively a pre-recorded music piece they are listening to. Research is therefore needed on the analysis of expressive behaviour of groups, in order to individuate behaviour descriptors capable to explain group-based social activities, to enhance cooperative and social tasks in user-centric networked media emerging applications. The analysis of non-verbal expressive social descriptors, like entrainment and empathic behaviour in a group, or the individuation of leadership and salience descriptors along the evolution of a group behaviour, are examples of relevant social descriptors which are candidate to greatly enhance future networked media applications.

4.2 3D User Generated Content

User-generated content (UGC), also known as Consumer Generated Media or User-created Content (UCC), refers to various kinds of media content that are produced by end-users. The UGC movement is no longer a fad. In the US, eMarketer projects that the number of UGC creators will rise from 77 million in 2007 to 108 million in 2012. The content is being read, seen and heard, too. The number of consumers of UGC will increase from 94 million in 2007 to 130 million in 2012[6].

Taking into account the following facts:

- UGC can help to keep innovation going within a community, and people's desire to be members of a community (social networks).
- UGC offers the possibility to everyone to act as both a producer and a consumer (prosumer).
- Currently UGC refers mainly to sound (music), text (blogs) and 2D content (pictures, flash animations, video clips).
- There is an explosion of 3D virtual worlds and 3D MMOGs.
- There is an explosion of 3D environments which aim at creating a "convergent reality", where real and virtual life will coexist (this includes serious gaming).
- There is significant progress towards 3D immersive environments and 3D Internet.

[6] http://www.emarketer.com/Article.aspx?R=1006190

We foresee the need for specific research which will provide a basis for easy creation and manipulation of and interaction with 3D UGC. More specifically, the following research challenges are provisioned:

Creation of realistic 3D UGC

The creation of 3D content for immersive virtual/augmented environments requires an extensive knowledge of the existing 3D computer graphics (CG) modelling tools (such as Maya, 3DS max, etc) and is a highly time consuming process even for those who are accustomed to use those tools. However, the realism that is achieved is not always the desired one. Realistic 3D content is the main ingredient for enhancing user's experience and enabling him/her to see the content from different viewpoints.

Realistic 3D modelling of real-world objects and scenes is an open worldwide research topic of computer vision and image processing scientists, since T. Kanade et al. have formulated the concept of Virtualized Reality in 1995[7], today also known as Plenoptical Imaging[8] or Free Viewpoint Video (FVV)[9]. FVV does not use CGI for modelling 3D objects and scenes, but addresses the 3D reconstruction of real objects and scenes by means of computer vision and by representing the reconstructed real-world elements with data formats similar to those from CGI, such that users can interact with real world reproduction as known from virtual reality (VR). The main problem of existing FVV solutions is the gap between complex high-quality approaches that do not run in real-time and simple real-time solutions that do not provide sufficiently high quality. It is self-evident that both, high rendering quality and real-time processing are required for achieving realistic 3D content.

Intelligence in 3D UGC

Intelligence embedded in 3D content can provide personalised reaction to users. This can enable the 3D content to give different feedback to users according to user's context and to enable for more user-centric/adaptive applications. This intelligence can be interpreted as autonomous behaviour of the 3D content (either CG or reconstructed) which could enable users and virtual characters to autonomously adapt their movements/gestures and adjust their emotional behaviour. The latter could possibly lead to tightly coupling of the augmented world to the physical and emotional behaviour of the real user. Specific research is thus needed on how real users can perceive the feeling of their engagement with the augmented world based on visual, emotional, sound and tactile feedback.

---

[7] T. Kanade, P. J. Narayanan, P. W. Rander, *Virtualized Reality: Concepts and Early Results*, IEEE Workshop on the Representation of Visual Scenes (in conjunction with ICCV'95), Boston, MA, June 1995

[8] S. M. Seitz, K. N. Kutulakos, Plenoptic Image Editing, Proc. 6th Int. Conf. Computer Vision, 1998 (ICCV98), pp. 17-24

[9] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards", Int. Conference on Multimedia & Expo, Toronto, Canada, July 2006.

Interaction and non-verbal expressive communication

Interaction with future media (including UGC) needs to be intuitive and natural. Users should be enabled to interact with the (disappearing) systems by means of the usual human-to-human communication mechanisms (e.g., natural language, gesture), including the subtlest emotional aspects of human interaction. Non-verbal communication channels play a paramount role in the latter. However, a deep understanding of the mechanisms underlying them, based on strong experimental evidence, and, as a consequence, the development of robust techniques for a comprehensive and reliable analysis of human expression are still open research challenges. Major research challenges for interaction consist of gaining a deeper understanding of human non-verbal expressive communication channels, of developing techniques for analysis and recognition of a broad range of spontaneous gestures, of developing techniques for analysing the subtlest and more significant human emotional expressions, e.g., empathy and emotional engagement. Results will enable a more natural and intuitive interaction of users with future user generated/centric media and will provide valuable knowledge to enhance socially interactive virtual characters.

Collaborative creation of 3D UG worlds

Given the enormous amount of 2D images and videos that people capture (and upload) when visiting a place that they are really interested in (e.g. museum, castle etc.) during vacation or educational trips and the current trend of social networks, it is not difficult to foresee the shift to multimedia social networking environments. Towards this aim significant effort is needed for allowing the collaborative creation of 3D UG worlds by 2D images and videos which are taken by simple cheap commercial cameras.

Interoperability and standardisation

Interoperability between different 3D virtual worlds is a main issue. Nowadays, the users of a virtual world can create their personalised avatar (by choosing from a predefined combination) and use it only within this virtual world. They can neither use it in a different virtual world nor in a 3D MMOG. In order to achieve the latter, specific research endeavours are required to produce new standards which will help in better defining the field and allowing for more sophisticated “convergent reality” and multimedia social networking environments.

4.3 Immersive Media

Immersive media is a new and emerging field of research that considers the emotional and profound effects of the augmentation, expansion and amplification of media formats and representations. The dramatic increase of the dimensionality of images as well as of sound, changes radically the human relationship to media. The media is no longer an object detached from human; the media becomes an ecosystem that surrounds us, reveals other aspects of our human condition and provides a space for new

emotions to play out. Immersive media can empower new kinds of augmented and virtual worlds in which storytelling and narratives can be experienced from the inside; new forms of games, where direct manipulation of media objects is the main interaction metaphor; and new forms of experience, where human emotional and intuitive apparatus can be directly engaged and challenged.

While the concept of immersion is well elaborated, the technologies of physical immersion are still being developed. We foresee the need for specific research which will hold the promise of creating large-scale Tele-immersive social experiences that are ideal for multiple users, facilitating direct communication and dialogue.

Tele-Immersion environments

In Tele-Immersion environments/rooms users are reconstructed without a direct awareness of the virtual/augmented world. Users are able to see and hear elements in the virtual world but how can they navigate and interact with other reconstructed humans, virtual characters or virtual objects? There are problems related to the users' concurrent interaction with the same virtual character or virtual object at the same moment, in detecting collisions and providing the appropriate physical response e.g. when users hit a wall, touch an object, in trying to virtually walk distances longer than the available space in the Tele-immersion room. In this context, users shall perceive the feeling of their engagement with the augmented world based on visual, emotional, sound and haptic feedback and be aware of how their actions affect the immersive environment (i.e. how their action affect other co-immersed users, virtual characters and/or objects).

Virtual contact and navigation

Novel navigation interfaces will augment a user's movement in the immersive environment (e.g. walking, body posture) allowing them to act naturally (move or express) and unconstrained within the physical limits of the set up of the immersive environment. Force feedback devices will allow the haptic sense and tactile perception from user's engagement in actions of contact with virtual objects. The goal will be to integrate a visual representation of virtual materials (e.g. textiles) with a haptic/tactile interface, thus allowing the user to have the sensation of feeling the virtual surface. Both the visual simulation and the haptic rendering shall be based on the actual physical properties of the simulated surface, taken from specific measurements on real material. This makes it possible for the user to identify different kinds of materials. The achievement of the final goal of creating full haptic interaction with real-time animated surfaces will facilitate the future realisation of such multimodal frameworks, featuring haptic interaction with any kind of virtual objects.

Multi-modal, multi-party interaction among real users and virtual characters

Multi-modal dialogue managers will drive the conversational behavior of reconstructed users and the virtual characters by means of interpreting multi-modal input and in return by producing multi-modal output as speech, facial expressions and gestures. The goal it to create personalised emotional interactions that will facilitate a conversational framework, featuring multi-party dialogues among several reconstructed immersed

users and autonomous embedded conversational agents. Conversational agents will feature virtual senses (vision and hearing), and will communicate with the real users through speech and body language and will guide/navigate the users in the virtual world. Expressive systems will link the emotional state and conversational states with the face, body and gaze behavior of the virtual characters so that the character can speak and behave naturally.

Sonification and sound-based interaction Sound has a great potential in immersive media. From one hand, sound has a leading role as a carrier of expressive and emotional content: it has a high potential in the virtual/augmented worlds to increase the effectiveness of the rendering of the emotional intentions of the remote users. From the other hand, it is well known the importance of 3D sound to increase the perception of immersivity in augmented reality. Auditory displays, 3D sound, sonic interaction design will provide important contributes to user-centric media. Further directions include the following:

Sound synthesis and processing finalised to enhance the interaction and navigation on content.

Strategies and techniques for combining suitably in the virtual/augmented world the sounds captured in different real immersive environments.

the synthesis of sounds, e.g. based on physical models of real and virtual objects, to obtain more effective immersive media.

5 Conclusions

This white paper on User Centric Media in the Future Internet reflects a view from the User Centric Media cluster of ongoing FP6 & FP7 EU funded projects how the Future Media Internet might develop and what the main societal and technological drivers are. The main objective of this work was to identify the challenges that should be realised in order to reach the goal of the User Centric Future Media Internet which is one of the main pillars of the Future Internet.

Towards this aim the greater societal megatrends and the disruptive technologies were identified and two groups of scenarios, one on the work environment and two on the digital lifestyle and home environment, were explained. Then, a discourse about Internet's evolution took place analysing the three dominant perspectives of the FI, namely User-Centric, Service-Centric and Content-Centric. Moreover, the concept of Content Objects introduced which lead to a layer-less content-centric architecture design. Finally, three groups of potential research challenges were described in detail which focus on Interaction/Personalisation/Identity, 3D User generated Content and Immersive Media.

The document is targeted to readers who want to take part in shaping the Future Media Internet. The identified functional requirements, challenges, opportunities and possible barriers can be used as an input for developing an efficient future Internet related RTD work programme for the coming years.

6 Editors & Contributors

Editors

Isidro Laso-Ballesteros (Cluster Leader)
European Commission
Networked Media Syst. Unit
Isidro.Laso@ec.europa.eu

Petros Daras (UCM cluster Coordinator)
Centre for Research and Technology Hellas - Informatics and Telematics Institute
daras@iti.gr

Contributors

No	Partner name	Company	Project	e-mail
1	Isidro Laso-Ballesteros	EC – DG – INFSO – D2		Isidro.Laso@ec.europa.eu
2	Petros Daras	CERTH/ITI	Victory	daras@iti.gr
3	Marianna Panebarco	PANEBARCO & C	VirtualLife	Marianna@panebarco.it
4	Helena Bijnens	Europace	Beacon	helena.bijnens@europace.be
5	Jan Bouwen	Alcatel-Lucent	TA2	jan.bouwen@alcatel-lucent.com
6	Peter Stollenmayer	Eurescom	TA2	stollenmayer@eurescome.eu
7	Nadia M. Thalmann	Miralab	Intermedia	thalmann@miralab.unige.ch
8	Dimitris Protopsaltou	Miralab	Intermedia	dimitris@miralab.unige.ch
9	Carmen Guerrero	UNIV. CARLOS III MADRID	Content	carmen.guerrero@uc3m.es
10	Kate Wac	UNIGE	Playmancer	wac@cui.unige.ch
11	Antonis Alexiou	Systema	Playmancer	Antonis.Alexiou@systema.gr
12	Alex Shani	EXENT	games@large	ashani@exent.com
13	Doug Williams	BT	TA2	doug.williams@bt.com
14	Antonio Camurri	Univ. of Genova	SAME	antonio.camurri@unige.it

No	Partner name	Company	Project	e-mail
15	Gualtiero Volpe	Univ. Of Genova	Same	Gualtiero.Volpe@unige.it
16	Alexander Voss	MICROSOFT	MyMedia	avoss@microsoft.com
17	Roman Ver-raest	K.U Leuven	Beacon	Roman.Verraest@avnet.kuleuven.be
18	Carmen Mc Williams	Grassroot	CitizenMedia	carmen@khm.de
19	Vytautas Čyras	Vilnius Universi-ty	VirtualLife	Vytautas.Cyras@mif.vu.lt
20	Paul Moore	Atos Origin	My-e-director 2012	paul.moore@atosresearch.eu