

Kim H. Veltman

Infrastructures for Ways of Knowing in E-Europe

European Commission. Directorate F. Emerging Technologies and Infrastructures. Applications FP7 Exploratory Workshop. Workshop 3. Knowledge Infrastructures for Researchers May 10-11 2004. cf. http://www.cordis.lu/ist/directorate_f/f_ws3.htm

Background

Questions determine answers. To ask what infrastructure we need for a knowledge society risks assuming that there is a unique infrastructure that is needed to address a single kind of knowledge. If we accept this question then one expects the same answer in the United States, Japan, China, Europe and throughout the world. If we accept this question then it is superfluous to ask what are the special things we need to do to create a unique E-Europe. The question and the idea behind it is attractive: If we build “it”, the objective infrastructure, the architecture, then “they” the subjective applications will come and fill whatever we build. So the challenge seems easy: the technologists ask us to outline a list of need and requirements. They produce the “goods” and then, almost as an afterthought we can “do” a few add-ons in the form of user interfaces so that inexperienced users can manage the system. Simple economics and common sense make this approach even more attractive. We cannot afford and it would be totally impractical to build a different infrastructure for every application. So somehow we have to accept that there is going to be only one infrastructure and make sure that we get it right. We have heard all this before. As Innis, McLuhan and more recently Matellart have reminded us, similar arguments were made with the advent of the telegraph, the telephone, radio and television.

The developments that are underway are fundamentally different and much more profound than these earlier technologies for four reasons: 1) convergence; 2) quantity; 3) scales of application; 4) speed of development and adoption. First, new communication technologies of the past brought their own new methods of creation and distribution broadcast: telegraph wires, radio stations, cable and satellite television. By contrast, Internet is something much more than adding new wires and machines. It is bringing a new convergence such that the same the new digital ICT technologies will be used for internet, telephone, television and indeed all communication. In the past, getting the infrastructure right meant getting right for one set of applications such as telephony or television. Today, getting the infrastructure right means getting it right for the entire range of human and machine communication.

Second, there is the challenge of quantity. If we listen to technologists and especially computer salesmen, then every year machines are bigger, better and cheaper. The good news is that there is enormous progress. Google has expanded to over 4 billion pages and over 880 million images in the past four years. The bad news is that this is hardly enough. What the technologists do not tell us is that for well over a century there has been an increasingly enormous chasm between information/knowledge production and knowledge organisation. The 20th century produced an estimated 200 million hours of radio and television archives of which less than 1% is readily accessible. The EU’s SERENATE enquiries have revealed that even in fields such as radio astronomy, scientists have access to about 1% of the available data. In the realm of culture, books and paintings and books are being scanned it at ranges between 30MB to 30GB per image, reconstructions are being made at 150MB to 5+terabytes. There is no infrastructure to share this quantity. The technologists have been kept busy creating new knowledge. There has not been sufficient attention to organising, maintaining, preserving that knowledge.

Third, the effects of the earlier technologies were largely in the 1:1 scale of everyday experience even though the science underlying them required increasing study at the microscopic level and even though the field of radio astronomy took us increasing towards the limits of the known macroscopic universe. By contrast the new ICT technologies are affecting interaction at every scale from the very small (10^{-15}) to the very large (10^{25}). Knowledge is no longer about only one scale but about relations between these scales.

Fourth, the earlier technologies took decades to develop and further decades before they affected a group of at least 50 million persons. Even 10 years ago it was fashionable to look at the growth of the Internet in

these terms. These comparisons are no longer useful. Between 1995 and 1996 the Internet grew from 5 to 50 million in one year. In 2000 there were 200 million and in March 2004 there are over 800 million. In 2005 there will be over 1 billion users and there is every reason to believe it will continue to grow. This paper explores visions, challenges and R&D priorities that these changes might bring.

Visions

Ever since Lisbon there has been a central vision of an E-Europe. Building grids for E-Science is an obvious place to start because the T(echnologies) part of IST play a vital role in these developments. On the other hand, E-Science alone is effectively a T programme. E-Culture needs to be part of the vision if we are have an IST programme that allows us move in the direction of E-Society and E-Europe. We are fortunate that the Ruffolo report of the European Parliament has given us new insights to look beyond some of the traditional “baggage” of the culture word and recognize that new links between local, regional and national are our only hope of ensuring that their uniqueness is represented in the emerging big picture of E-Europe. Mastering complexity is not just about mastering the small as well as the large (different scales) in science. It is equally about ensuring that the unique details of local culture continue to be acknowledged, recognized and studied alongside the big structures of major cities and political capitals. “Empowering users” is not just about new tools for scientific and technological progress. It is equally about new tools for creativity and cultural expression. Expanding knowledge is not just about a “knowledge infrastructure for empowering research” into new machines and industries. It is also about recognizing that the notion of “knowledge as power” that leads us to speak of “empowering” is but one of many definitions of knowledge.

The Anglo-Saxon notion of knowledge and power is linked with the German verb “to be able” (*können*), but ignores another dimension of the German tradition whereby this same verb is linked with art and creativity (i.e. *können* and *Kunst* as Panofsky showed in his study of Dürer). It also ignores the German distinctions between doing and knowing in a deeper sense (*können* vs. *wissen*), which is reflected differently in the French distinction between *savoir (faire)* and *connaissance*. It overlooks entirely Sanskrit roots whereby knowledge is linked with giving birth and life, or Hebrew traditions where knowledge is traditionally linked with carnal knowledge. Some knowledge is fleeting and of the moment (the magic of performance especially in dance and music). Other knowledge strives to be eternal, as in the case of astronomy. Scholars have reminded us that civilisations have chosen different paths. The Egyptians focussed on eternal dimensions, the Greeks on beauty, the Romans on a legal framework.

This leads to a first fundamental question: Are we trying to create a Knowledge Society where knowledge is defined in the Anglo-Saxon and particularly the American definition of this word, or are we trying to develop a society that respects, explores and nurtures the many definitions of knowledge of an expanding number of European countries? If we choose the narrow definition then we are aiming at an e-modern technological system that is no different than American or some Anglo-American solutions. If we choose the larger definitions, then E-Europe has new challenges of creating technological infrastructures and human(e) contexts to explore how we maintain and develop diverse and sometimes contradictory notions of knowledge. Historically Europe has changed its position. At times it tried to define itself in terms of a single religion: Christianity. At other times it has acknowledged and respected and studied seriously other religions and other cultures. If the Digital E-Europe tries simply to replace A(nalog) Old Europe, with a good slogan type answer of the day, it will lose all the richness of its past. Alternatively if E-Europe embraces this past in its evolving visions of itself, then historical and cultural dimensions of knowledge must become central to its models.

From this emerges a vision that recognizes both the enormous changes of the present and the immense riches of our past to create an E-Europe, which remembers its past mistakes as well as its successes in building a new future. This combines E-Science and E-Historical Cultures in arriving at new visions of E-Europe.

Challenges

1. DEER

As noted above four factors make developments in ICT fundamentally different than earlier technologies: 1) convergence; 2) quantity; 3) scales of application; 4) speed of development and adoption. These factors are so complex, so wide ranging and changing so quickly that our existing models of carefully collecting (user) requirements, then building and evaluating is no longer enough. FPs 1-6 have increasingly emphasized the need that users must be central to the development process, but in practice this has meant that a few social scientists and usability experts are added almost as icing to the architecture cake. When the project delivers its project in the form of “Work Packages”, these are typically in the form of written reports rather than in the form of working prototypes that can be shared and made available to researchers. The challenge is to structure the IST programme such that the processes used (including things that did not work) and the resulting products are accessible for the entire community. The entire IST programme should effectively become a series of collaboratories wherein researchers can share experiences. In theory this is what the NOEs should be doing and what CORDIS should co-ordinate. In practice even EU employees admit that finding things on CORDIS is practically impossible because there is no permanence to different sections of the site. In the analog world we have created memory institutions (libraries, museums, archives), for all enduring knowledge. We need the digital equivalent of memory institutions in the form of a Distributed European Electronic Resource (DEER), which fosters our collective memory through permanence, visibility and access. In addition to distributed repositories of knowledge, this would entail virtual reference rooms and a virtual agora for sharing knowledge (cf. challenges 4 below).

2. Multi-lingual, Multicultural Access.

If we accept that E-Europe must embrace, reflect and foster our rich historical and cultural heritage and traditions, then there is a range of challenges in developing multi-lingual and multicultural repositories. As noted above in our discussion of knowledge a simple translation from English into the different languages of the union is not enough. We are fortunate that the AMP (Accès Multilingue au Patrimoine) Consortium has explored some of the challenges. These need to be further explored and to be made a part of evolving databases that are integrated into the DEER.

3. Pedigree and Versioning

There is general agreement that Intellectual Property (IP) and their attendant Rights (IPR) are important ingredients in the digital world. There are obvious economic considerations that can be linked with concepts such as digital watermarks. In addition, there are challenges of creating tools to identify the pedigree of images and texts. Is the digital version in front of me a direct “copy” of an original painting (or document) or is it based on a photograph, which was based on a copy of the original? If there have been changes, we need tools for determining the precise version with which we are dealing. Without such tools we shall lose our ability to distinguish between scholarly discussions based on originals and wild claims based on tampered versions. In the analog world these tools are in the form of clearly identified editions. We need digital equivalents to maintain a defensible notion of academic excellence and indeed the fundamental idea of authenticity.

4. New Structuring Tools

McLuhan helped to make us aware that new media do much more than simply moving the contents of earlier media into the new form. The advent of printing began by moving the handwritten contents of manuscripts into the typeset form of books. Gradually it separated the production of (primary) sources from the creation of (secondary) literature and journals about those sources in the form of commentaries. Print media has made further distinctions between magazines, articles in journals, articles in refereed journals, articles in books, monographs etc. We need to have digital equivalents of these distinctions. At the same time the new media call for further distinctions in knowledge structuring: between internal analyses, external analyses, restorations and reconstructions. We need to build these into our tools for structuring enduring knowledge. In practical terms this means providing new structures to our distributed repositories of knowledge and using these in equivalents of virtual reference rooms.

One fundamental reason for this effort is that they allow us to benefit from previous attempts to value and evaluate quality. Rhetorically it is attractive to speak of knowledge anywhere, anytime. But this assumes that the knowledge has been vetted and digested. The great libraries aimed to collect all true knowledge but not everything that was written. In a world where the equivalent of 7 million new pages are being added daily to the Internet we need new tools for knowledge structuring, and orientation.

In the longer term something more fundamental is needed. Ever since Antiquity there has been a debate between the approaches of Plato and Aristotle, between universals and particulars, between top down deductive systems and bottom up inductive systems. Both seemed to point to static hierarchies. The 20th century helped us to recognize that these subsumptive hierarchies are dynamic. How can we construct such dynamic systems which allow us to go up and down inductively and deductively and learn to see these activities as belonging to a larger set of processes?

5. Personal Structuring Tools

A first enormous implication of the networked society is that even loners can have access to the collective memory of our time and achieve insights that go far beyond their personal horizons. In the digital age this is the equivalent of going to one of the greatest libraries such as Alexandria, the Vatican, or Library of Congress without needing to leave one's home. We need new tools as we go from handling a few books on our desk shelf to the tens of millions offered by these collections and the hundreds of millions when they are combined in networks.

Personal organisers are a well-established concept. We need tools to organise our appointments and our everyday activities. Even more so do we need tools to help us organise our ideas, our basic ideas. Yes, there are tools such as mind maps and tools for visualising clusters of ideas. But I need tools to remember that what I classed under topic A last year and yesterday relates to what I class under topic B today. I need tools to relate my personal concepts to existing fields of knowledge, to trace my ideas in the context of emerging fields of knowledge.

6. Multiple Awareness Profilers (MAP)

In the past there was a recurrent tendency to limit our identity by fixing it to a single geographical reality. In the 15th and 16th centuries, this identity was frequently in terms of a given local area, a city-state such as Florence. In the 17th and 18th century this identity increasingly extended to regions in the form of what we now call provinces or states. In the 19th century this sense of identity became linked with the national level. For a while it seemed that these were either-or allegiances: as if becoming an Italian or a member of the European Union would need mean abandoning one's sense of being a Florentine. Experience has proven otherwise. It is perfectly possible to be a proud Florentine, a good Italian and an excellent European. Now the challenge lies in finding new tools to express this new level of complexity. We have databases for Florentines, for Italians and for Europeans. We need databases to show Florentine, Italian Europeans; databases which reflect local-regional-national not only about persons, but also about objects, events and everything else.

We have created wonderful GIS tools, which will soon be linked with GPS and UMTS. We need tools which help us to trace not only exiting boundaries, but how they have changed over time, how Poland was a small country in 1000AD, the largest country in Europe in 1440 and today a smaller one in a larger entity. We need not only Poland's maps of itself, but also Germany and Russia's maps of Poland. We need not only the existing chronology but tools that help us to move between chronologies in different countries at different times. We need comparative, dynamic cartography and chronology.

7. Collaborative Sharing and Learning

History should prepare us to accept that a preference for study by lone individuals is not likely to disappear soon. Even so, the trend to work in groups, to develop ideas in teams is destined to play a greater role as we develop tools to make this possible. It is easy to predict that teachers will move from the sage on the stage to the guide by the side. Ironically those who predict this are typically still sages one stages. Even so there is an enormous challenge to create new learning and working environments where such sharing is not merely feasible or possible, but a reality of daily practice.

8. Collaborative and Personal Creation

We have mentioned the need for tools that give access to existing knowledge. Equally we need tools that help us to create new knowledge. Yes, we have word-processing tools that pretend to do this with text. Yet these force us to use keyboards with all the limitations of typewriters of the past century. We have tools to create simple collections of images with basic text (e.g. Powerpoint). But where are the tools that allow us to do this in a multi-media and omni-media environment; that allow us to link text, hypertext, still images, video, film and television clips in a single collection? We have editing tools for almost every medium? But where are the editing tools that allow me to combine media to make new personal versions of television programmes or films and then new combinatorial forms of artistic expressions that do not yet exist? These tools need to be at both the collaborative team and the personal level.

In the past FPs there have been billions spent on aspects of all these eight challenges. Alas, the new tools have largely been to allow major corporations to develop products to the exclusion of contributions by individuals and small groups. The rhetoric speaks increasing of users being empowered, of the importance of Small and Medium Sized companies (SMEs) and yet the reality remains that the great advances remain almost exclusively in the hands of an ever smaller number of international corporations. Such solutions have important financial potentials. But so too do individuals and smaller groups. If e-inclusion is to be more than polite rhetoric then we need to use taxpayers' monies for projects, products and insights which serve the whole of society and not simply to help a few companies dominate our society with their hidden or even open agendas.

R&D Priorities

1. Distributed Repositories as a first step towards a DEER
2. Multilingual Access for a Virtual Reference Room as a second step towards a DEER
3. Tools for Pedigree and Versioning
4. Tools for Knowledge Structuring and Value Indication
5. Personal Structuring Tools
6. Multiple Awareness Profilers (MAP)
7. Virtual Agora as a third step towards a DEER
8. Virtual Creation Tools

Policy Implications

- 1) The most fundamental policy implication is that the result of all IST projects will become part of permanent, cumulative, distributed repositories. In this sense the whole of FP7, in addition to many independent contributions, will contribute towards the growth of a Distributed European Electronic Resource (DEER).
- 2) At least two of the priorities (1-2) should make contributions to integrating the authority names of memory institutions (cf. projects such as GABRIEL, Bibliotheca Universalis) in a multilingual European-wide framework.
- 3) High level examples from memory institutions (e.g. Louvre, British Library) and research (e.g. NUME) should be used to explore the potentials of the Virtual Agora for teaching, research and new creativity.

Appendix. Further Thoughts

It may be helpful to provide some of the background to the thoughts by focusing on two themes: 1) DEER as a unifying vision for FP7 and 2) the shift from ICT to UCT to get a handle on what happens beyond FP7.

DEER as a Unifying Vision for FP7

In E-Culture Net the members identified three essential components to a future Distributed European Electronic Resource (DEER, figure 1).¹

VIRTUAL REFERENCE ROOM -

comprising, interfaces, discovery tools and navigation systems for portraying e-culture, for searching and content retrieval.

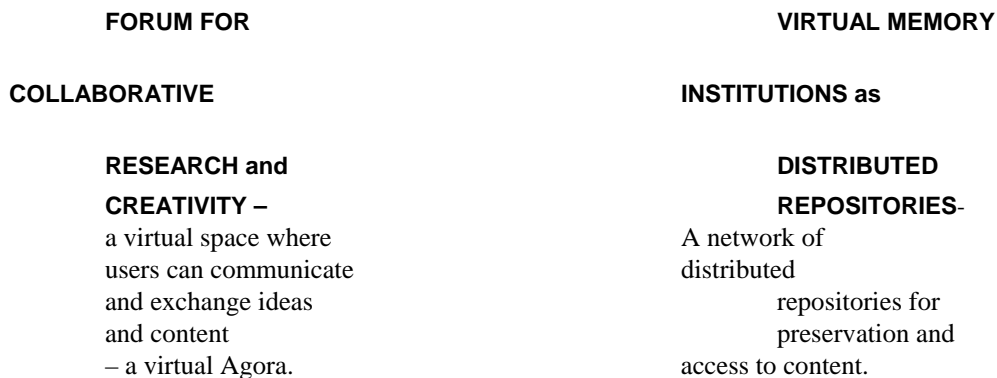
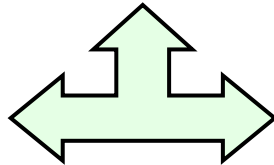


Figure 1. Essential ingredients for a Distributed European Electronic Resource (DEER) as a first step towards a World Online Distributed Resource (WONDER).

This idea is too big for Luxembourg given their budget. So there has been a feeling that this should become a dimension of the knowledge infrastructure visions. What you called the Knowledge Organisation layer in your schema corresponds at a macro-level to the Virtual Reference Room in figure 1. The Distributed Repositories correspond to what your circles think of digital libraries but then expanded to include all memory institutions.

The virtual agora is actually two things:

- 1) new kinds of classrooms/work areas where persons can learn together. Here the key technologies involve online access to historical objects, sites, routes and their reconstructions (the CINECA-RAI blue rooms).²
- 2) new e-creativity/innovation spaces, the high level versions of which will be the digital film, video, animation, game and television studios of the future. This will require online access to film, television and editing tools for purposes of visual quotations. In the industrial realm the high versions will be distributed virtual design studios, which links with virtual organization concerns of some of your members.

Our figure covered culture alone. E-infrastructure will eventually need to cover all fields. As you have indicated we need roadmaps to go from today to the story of everything. One obvious step is to take the

application areas identified by SERENATE and use them 1) as initial repositories which can be linked with memory institutions
 2) as initial subsets of the virtual reference room which need to be mapped from one to the other.
 3) as initial materials which can be used for collaborative workspace and collaborative creativity/innovation spaces.

A second obvious step is to insist on permanence in the results of EU projects: even permanent records of the so-called failures, so that we can understand where we went wrong and what worked. If this is done at the policy level in FP7, then by the time we reach FP8 the Commission should have enough of an online “demo” of the DEER project to convince member states to pay for this as a permanent investment. A third step is to align the Commission’s E-Content initiatives so that these also contribute cumulatively to the DEER. A fourth step might be to align the CORDIS database such that it integrates this.

Reference (Pointers)	<ol style="list-style-type: none"> 1. Names, Concepts, Terms (Classification Systems) 2. Definitions (Dictionaries) 3. Explanations (Encyclopaedias) 4. Titles (Catalogues, Bibliographies) 5. Partial Contents (Abstracts, Indexes)
Objects (Primary Sources)	<ol style="list-style-type: none"> 6. Full Contents (Books, Paintings, etc.)
Interpretations (Secondary Sources)	<ol style="list-style-type: none"> 7. Internal Analyses 8. External Analyses 9. Restorations 10. Reconstructions

Figure 2. Basic Levels of Knowledge linking virtual reference rooms (1-5) with virtual libraries (levels 6-10).

If it is approached properly the DEER could serve as a unifying vision for the whole of the FP7. It would give a framework for e-Science to become more than a buzzword; a framework that would allow E-Culture to benefit from many of the basic commonalities while also helping to identify what we need above and beyond the needs of specialist groups in science.

An obvious example is authentication and security which the physics and HPC (High Performance Computing) circles call AAA (Authentication, Authorization, Accounting). This is something that we all need. The High energy physics persons need it mainly behind a firewall. The cultural community, concerned also with public heritage needs ways of identifying the authenticity of images and objects (proof that they have not been tampered with). We don’t want to discourage persons tampering with images to create new ones. We just need to be certain that when an image claims to be an original it is what it claims to be. Else persons could be writing all kinds of theories about Mona Lisa and other paintings which bear no relation to the original object. (This past week Geoff Hoon has realized that uncertainty about the provenance and authenticity of images can have serious political and in turn economic consequences).

A second example, if we use figure 1 we see that the science community focuses on distributed repositories and the collaborative workspace of the Virtual Agora. Their notion of middleware focuses on these two bits. The world of high industry, architecture and the broadcast worlds (including film) emphasizes the e-creativity/innovation spaces. The world of culture (arts, humanities, social sciences, history) reminds us of the need to include the virtual reference room idea as something far more complex than linking a few glossaries. One of the unique challenges for Europe may well lie in developing a DEER for all three of these communities so that they can have synergetic effects. I think Michael Smirnov’s term layered knowledge (with the “services” bit later) is exactly what we need. If we build our road map in this way we can see exactly where we are.

From IST/ICT to UCT

The other big point in our section was this shift from Information Society Technologies (IST) or Information and Communication Technologies (ICT) to Universal Convergent Technologies (UCT). Figures 3 and 4 give a Japanese and German take on these developments. The American way would be to say industry is already “taking care of this.” But if Siemens offers another version than IBM etc. then there will no way that we can arrive at a serious version of a knowledge society. My recommendations are twofold.

1) This should become a FET theme which worries about producing the future equivalent of what UMTS set out to do for telephony. Effectively this would be a Universal Ambient Mobile Telephony, Internet, Television (UAMTIT does not seem a good buzzword so UCT is preferable.

2) Since we already have examples of IP telephony, digital film, telephony we should have testbed projects in FP7 which explore and experiment with the enormous implications of this shift.

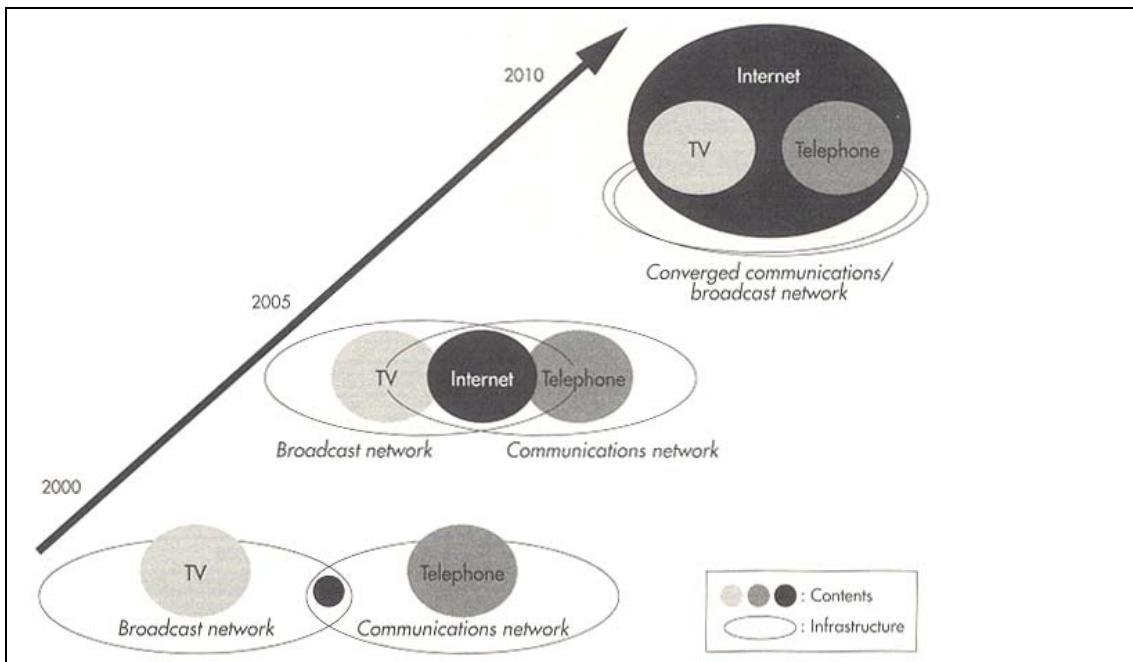
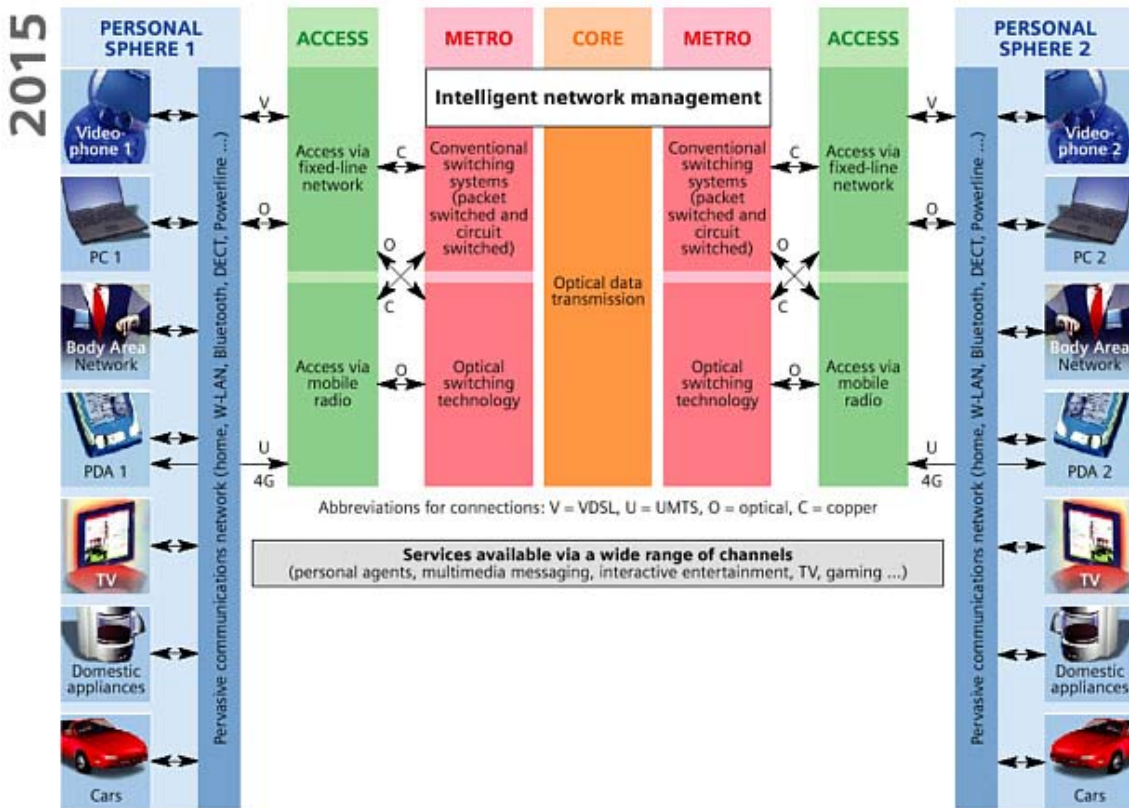
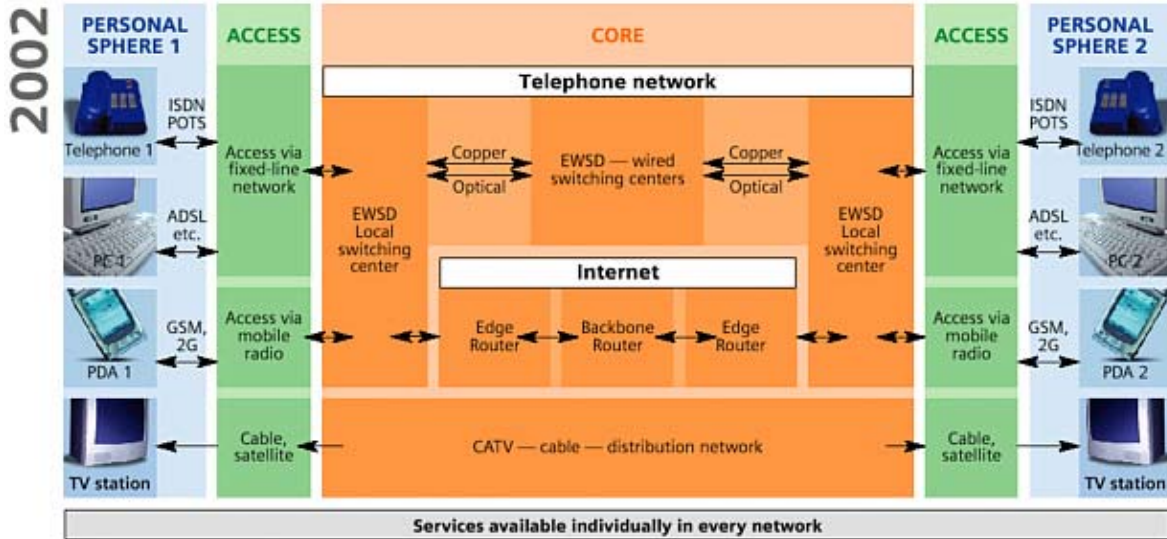


Figure 3. Japanese scenario for convergence of Internet, broadband and telecommunications networks according to Hiroshi Fujiwara, Internet General Research Centre (2000).



Notes

¹ DEER. See: <http://www.eculturenet.org/FP5/publicPDF/deliverable11b.pdf>.

² For an idea of existing examples cf. an article and a thesis that was written for e-culture net: http://www.reuna.cl/redID/docs/docs_internacionales/veltman.pdf <http://www.eculturenet.org/FP5/publicPDF/deliverable10c.pdf>

³ *The State of the Japanese Internet Market 2000 Digest*, Tokyo: Impress, 2000, p. 59. Cf. Ken Ichiriki, *Study Group on the Next Generation Internet Policy, e-Japan Initiative for the IT Revolution*, June 2000. The Siemens descriptions come from an article by Pease at Siemens.

See: http://w4.siemens.de/FuI/en/archiv/pof/heft2_02/artikel11