Journal of Universal Computer Science, vol. 8, no. 9 (2002), 848-868
submitted: 4/2/02, accepted: 1/7/02, appeared: 28/9/02 © J.UCS

Webs, Grids and Knowledge Spaces
- Programmes, Projects and Prospects -

Hans-Georg Stork[1]
(European Commission, DG Information Society
Hans-Georg.Stork@cec.eu.int)

Abstract: Many believe that today's Web has not yet reached the full potential which globally distributed systems may achieve in terms of information access and use. Realizing this potential may indeed turn the Web into a vast knowledge and service space. We discuss some of the issues involved and present a number of activities initiated and supported by the European Commission that are likely to make significant contributions towards attaining this goal.
Key Words: semantic web, grid computing, knowledge technologies, European research funding
Category: A.m, H.0, H.4, K.4.1

1 Introduction

In less than ten years the World Wide Web, based largely on HTTP and HTML, has evolved into a vast information, communication and transaction space. Its features differ greatly from those of traditional media.

Its development has been driven by a variety of interests and needs: by commercial interests certainly, in that "the Web" represents a new global platform for advertising, selling and trading; but also by the need to manage an ever growing body of documents of all sorts, and to give specific and general audiences access to these documents.

Indeed, these needs were apparently the prime motive in the early days, when Tim Berners Lee conceived the nucleus of the World Wide Web. His original *"proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system."* ([W3C_TBL 1989])

Not surprisingly, the CERN is one of the focal points of yet another major initiative to harness the power of distributed systems. Known as "the *Grid*" and launched a few years ago (cf. for instance [Grid 1998]), this initiative is about distributed computing while "the Web" is - at least *on the surface* - about distributing and accessing digital content.

[1] The views expressed in this note are those of the author and do not necessarily engage the European Commission.

However, "the Web" itself has become the target of an effort the objective of which is to enable software agents to perform complex transactions and other operations on Web content without extensive human intervention ([W3C_SWA 2001]). This objective is referred to as the *"Semantic Web"*.

Both the *Grid* and the *Semantic Web* initiatives have been taken up by the European Commission's Information Society Technologies (IST) Programme that started in late 1998.

In this note we review some of the issues pertaining to these initiatives and discuss how Webs and Grids may relate. We present those parts of the IST Programme that address these initiatives directly, and introduce a number of relevant projects that are being co-funded by the European Commission.

We contend that specific technologies underlying the *Semantic Web* and the *Grid* belong to the core of what may be called *"Knowledge Technologies"*. The latter will in fact be the label of one of the principal domains of the next IST Programme under the European Commission's 6th Framework Programme for Research and Development (2002 - 2006). These technologies may help to turn globally distributed systems into vast knowledge and service spaces.

2 Webs

To most people "the Web" is simply the World Wide Web (WWW), which does indeed have a direct impact on the everyday life of a steadily growing number of people.

Yet, the WWW is but one instance, albeit the largest, perhaps most important and certainly the best known, of a technology the principles of which were invented quite some time ago (e.g. Vannevar Bush's MEMEX [MEMEX 1945] or Ted Nelson's XANADU [XANADU 1960]). They have been experimented with, based on comparatively 'primitive' precursor technologies (Interactive Videotex was one of them), a long time before 'the Web' became the most popular application of the Internet ([W3C_Dybvik 1993]).

This technology can be described most appropriately as *distributed hypermedia systems* whose actual distribution may vary greatly in scale. The WWW is the outstanding example of a *very large scale* distributed hypermedia system, owing its size to the underlying Internet[2].

While the World Wide Web as well as many local or company Webs (over intra- or extranets) are based on the *HTTP*-protocol for data transfer and *HTML* presentational markup for content display, there are a number of other distributed hypermedia systems built on top of the Internet infrastructure, that use different protocols and content description schemes. Early examples developed

[2] One may also argue that it has been the World Wide Web that nourished the growth of the Internet.

before or around the time the WWW started to gain ground and momentum are MICROCOSM (University of Southampton, [Microcosm 1997]) and HYPER-G (Technical University of Graz, [Hyper-G 1995]).

Largely due to the alleged 'simplicity' of the WWW (e.g. in terms of ease of 'putting content on the Web') and its ensuing dominance, alternative designs never really took off. Yet some of them offered sophisticated features well ahead of the original WWW structure and functions. HYPER-G, for instance, offered bidirectional linking, links and content description separated from content, content management through client interfaces, ..., features that are only now entering the WWW scene at large, partly based on a growing set of W3C recommendations, including the *XML* and *RDF* families.

Of course, saying that Webs are 'distributed hypermedia systems' only defers an explanation. The most concise way, perhaps, of characterizing such systems would be as interlinked *digital content* that resides on servers and that can be accessed, represented and interacted with through specific interface clients, known as 'browsers'.

Digital content can be almost anything. With their capability of interpreting various forms of 'telesoftware' (e.g. Java applets, Java code or Javascript) and of hosting so called plug-ins, browsers can deal with a large variety of transaction requirements and content types. Servers, on the other hand, are capable of assembling content on the fly, from all kinds of sources, including data base systems, document management systems and computing facilities in general, thus reacting to whatever request a user may issue through her browser. Processes running 'behind the scenes' can be of any degree of complexity.

Thanks to this generality Webs (and "the Web") lend themselves to all sorts of applications. From a user's point of view a Web is simply an *interface to resources*. These can be: documents, data residing in data bases, computing facilities, data capturing devices, sensors, applications, ... Applications are resources that make use of resources.

This is why *'resource description'* ([Miller 1998]) is all important on Webs. It is a prerequisite for effective and efficient *resource discovery* and use, just as comprehensive catalogues are needed to make full use of brick-and-mortar libraries.

Although Web technology greatly surpasses its - in retrospect - rather clumsy predecessors many believe that it has by no means reached its full potential. Whatever this may be: at least two issues have to be addressed, one being the *"semantic" access and use* problem (i.e. access to and use of content and services, based on semantically sound resource description); the other being the *universality* of *physical access* via high-bandwidth local loops and broadband wireless channels. These are certainly moving targets.

While *universal physical access* is an issue that concerns the Internet and networking in general, *semantic access and use* is about making webs (and of course "the Web") more usable (as distributed systems) and "user-friendly" (as interfaces).

3 The Semantic Web

The *Semantic Web* initiative addresses precisely this latter issue (see also [W3C_Sem 1998] and [W3C_Data 1999]). As pointed out in the introduction to this note it is about automating a range of tasks within the context of distributed systems (such as the WWW): from (chains of) business transactions to searching, filtering and integrating relevant and trustable information on whatever subject a user may be interested in.

Software agents performing these tasks must be able to make sense of Web-based resources regardless of who or what is providing them.

"Making content machine-understandable" is therefore a popular paraphrase of the fundamental prerequisite for the *Semantic Web*. This is to be taken very pragmatically: for content (of whatever type of media) to be "machine-understandable" it has to be bound (attached, pointing, ...) to some formal description of itself (often referred to as metadata[3]).

However, while formal resource description is necessary it is not sufficient. Descriptive terms are "understandable" only if their meaning has been defined somehow, somewhere. This is usually done through *ontologies* (e.g. [Gruber 1993]) which provide meaning that can be operated upon. They embed terms in contexts (of other terms) and/or stipulate rules a given term (or set of given terms) must obey.

Hence formal ontologies are instrumental in achieving the *Semantic Web*. Rooted in a long tradition not only of formal logics and artificial intelligence but also of more mundane endeavours such as the setting up of data dictionaries, classification schemes, thesauri and controlled vocabularies, they are currently the most promising candidates for a sound semantic ground of descriptions of digital content.

Of course, saying that software agents can "understand" something is only a figure of speech. What they really do is expedient data processing, serving a given purpose, no more, no less, in accordance with the intentions of their designers. On the *Semantic Web*, however, the interpretation of data is not entirely hard-coded in an agent's software but readily retrievable from some ontology, itself a resource, located and accessible somewhere ("factored out", so to speak).

[3] Classical examples of collections of metadata are library catalogues consisting, for example, of MARC records describing books and other items belonging to the 'Gutenberg Galaxy'. By contrast, the metadata we have in mind when talking about the *Semantic Web* pertain to all kinds of digitally representable objects.

Ontologies do indeed represent *knowledge* if we concur to describe knowledge roughly as something that is needed to "do the right thing", to "draw the right conclusions" or to "take correct decisions" (what agents are supposed to do). Whatever is "right", "correct" or "useful" must, in a final analysis, hinge on implicit or explicit agreements among human beings[4].

Thus the *Semantic Web* will be a *"Web of knowledge"*, quite apart from all those other forms of knowledge it may contain (such as papers, articles, reports, recipes, still and moving images, graphs, product and service descriptions, ...), and that are meant for direct human consumption. In fact, it will be *"meta-knowledge"* (about digital content of all sorts) that makes all the other forms of knowledge (and occasionally nonsense!) more accessible and usable.

But how do we get this (meta-)knowledge on the Web? In view of the vast amounts of content already out there and constantly being produced this is likely to be a Herculean task. Clearly, this must be the remit of specific technologies, *"Knowledge Technologies"*, to be precise, drawing on various Computer Science sub-disciplines such as formal modelling, logics and languages, information retrieval, (multimedia) databases, image analysis, ..., but also on "trans-disciplines" such as Cognitive Science.

4 Grids

Grid [Grid 1998] stands for an allegedly new paradigm of large scale scientific computing (or "research networking"): the application of co-ordinated computing resources, interconnected via high-speed networks, to the solution of problems in fields such as high energy physics, astrophysics, nuclear physics, geophysics, meteorology / climatology, neurobiology, molecular biology, earth observation, operations research, ... Grids are *large scale distributed computing systems* providing mechanisms for the controlled sharing of computing resources. (The term has been borrowed from the electric 'Power Grid' that enables the sharing of energy resources.)

Grids need generic 'middleware' components, that shield specific applications from the details of accessing and using a configuration of heterogeneous resources, such as processors, storage and network connections. They guarantee resource interoperability through the use of standard protocols. The term *Grid technology* ([Globus 2002]) usually refers to this kind of middleware.

[4] For obvious reasons we do not engage in any discussion of the elusive notion of knowledge beyond this rough description; we do, however, maintain that there is a set of operations applicable to whatever knowledge may be. This set includes: *acquisition, elicitation, discovery, representation, communication, inference* and *access*. We also contend that knowledge can be more or less precise, more or less pertinent and hence more or less usable in a given environment and for a given purpose (*"ex falso quodlibet"* is a well known worst case).

An important function of *Grid* 'middleware' components is to 'discover' resources and information about these resources in order to optimize their use. Middleware components also provide a range of directory and file services.

Most *Grid* applications (e.g. within the above mentioned areas), apart from requiring enormous processing power, deal or are expected to deal with huge to gigantic datasets (peta - exa orders of magnitude).

So far, industry involvement in *Grid* application developments has been fairly limited. But this is already about to change and more rapidly so as generic *Grid* middleware becomes more widely available.

Less resource demanding applications (commercial or not) that nevertheless require or would lend themselves to large scale resource sharing, may then benefit as well. The increasing popularity of *"peer-to-peer (P2P)"* computing schemes, generalizing the traditional *client-server* paradigm, appears to corroborate this conjecture.

5 Grids and Webs

What do Grids and Webs have to do with each other? Well, everything and nothing. 'Everything' because in the digital world everything can somehow be related to everything else ([Negroponte 1996]). 'Nothing' because they address - at least at first glance - entirely different problems, tasks and functionalities.

While both are being operated on the Internet they are currently driven by different needs and interests: Grids by 'eScience' (including - more and more - 'industrial science'), Webs by 'eCommerce' and 'eContent' (of which more and more will be multimedia, e.g. for enter-, edu- and infotainment). Grids (and *Grid* technology) have fairly limited user communities; by contrast Webs (and "the Web" in particular) address and potentially serve millions of people, i.e. the general public. Grids are about doing special computations on huge to gigantic datasets whereas data volumes flowing across Webs are far more modest, ranging from very small (e.g. a transaction request) to very large (e.g. high-definition streamed video), also depending on the capacity of physical access paths.

Of course, there are some common basic problems which may have common solutions; *Grid* and Web developers may actually benefit from each other, for instance in the area of metadata codification. One has to bear in mind, however, that the characteristics of resources are quite distinct between Grids and Webs. Table 1 summarizes the 'differences' highlighted so far.

To take 'nothing' for an answer to the question introducing this section may indeed be a bit too little. And it is certainly not necessary. The clue to a possibly correct understanding of the relationship between Webs and Grids lies in the statement "Webs are mainly interfaces to 'behind the scenes' applications" (cf. Section 3). We noted that these applications can be arbitrarily complex. And

	Grids	Webs	*Comments*
main drivers	(big) eScience, eEngineering	scientific communication (initially, now:) eCommerce, eContent (multimedia)	*there is some overlap and there may be more in the future*
main functions	high performance computing, sharing of computing resources	information, communication, transactions	
applications	computationally hard and data intensive problems in science and engineering (e.g. realistic simulations)	I&C services, education & training, eBusiness, eCommerce (B2B, B2C, B2A, etc.), etc.	*Webs are mainly interfaces to 'behind the scene' applications*
data volumes	XXL (and bigger)	S - XL	*future Grids may also work on smaller volumes*
resources	storage (incl. caches), bandwidth, processor time, data files, ...	digital content and related services	*containers, conveyors & processors vs. content & applications*
users	special user groups (scientists, engineers)	general public, businesses, public administrations, etc.	*these are only the main target groups*
standards	middleware standards need to be agreed	many standards and recommendations exist	*Grid and Web communities are still fairly separate*

Table 1: A "comparison" of GRIDs and WEBs

we do not usually care who or what is working 'behind the scenes'. So it may be Grids (or isolated high performance computers or just an ordinary PC, or whatever). Indeed, *Grid* applications could render invaluable services even to the general public, via specialised professionals such as medical practitioners. These applications would be accessed via a Web and their output (e.g. visual representations of complex objects or simulations) translated into standard Web formats.

We shall argue that Grids could indeed provide the Web (or Webs . . .) with 'knowledge', assuming the pragmatic interpretation of that notion put forward in Section 3.

6 The Knowledge Grid

While the Web community, led by the World Wide Web Consortium (W3C), and with substantial contributions from researchers in the fields of Artificial Intelligence, agent and database technologies, has been developing and refining the concept of a *Semantic Web*, *Grid* proponents - notably in the United Kingdom ([DeRoure et. al. 2001]) - extended the basic architecture of the *"Computational Grid"* by adding two layers, called *"Information Grid"* and *"Knowledge Grid"* respectively.

Roughly, the first two layers of this model make up the technology explained in Section 4. By contrast, a major role ascribed to processes running within the "knowledge layer" of a *Grid* is to assist in making sense of the huge amounts of data generated by, say, scientific instruments such as particle accelerators, gene sequencers, telescopes, satellites and a gamut of sensors. And they should do so by making use of the computational power and the services of the underlying layers.

As pointed out in Section 4, current *Grid* development is mainly driven by the needs of *"data intensive"* science. *Knowledge Grid* processes can therefore be understood best as special applications of the computational *Grid*, supposed to enhance scientific and other *"problem solving environments"*. They make use of a variety of techniques, such as those that can be described broadly as *algorithmic content analysis* and *algorithmic learning.*

7 Semantic Webs and Knowledge Grids

From the foregoing it should be clear that the relationship between *Knowledge Grids* and *Semantic Webs* may be characterized as *complementarity. Knowledge Grid* techniques are indeed among those the *Semantic Web* calls for in order to render its contents meaningful to software agents (e.g. by creating semantic annotations or by mining data - representing text or other forms of content - for the purpose of establishing and maintaining ontologies). On the other hand the formal framework and (possibly) the organisational underpinning of a *Semantic Web* would be needed to make full use of *Knowledge Grid* resources and services.

While many of the basic ideas underlying both the *Semantic Web* and the *Knowledge Grid* initiatives are not new, the sheer size, capacity and dynamics of today's global networks (notably the Internet and whatever it will develop into) provide a strong incentive to turn these ideas into large scale reality. This in itself may require a major research effort. In a manner of speaking the evolution of the Internet and the Web brings some of the *"good old fashioned Artificial Intelligence"* research, results and approaches (some of which do require powerful computing resources) down to earth, begging for new approaches to solving new problems.

8 Programmes

However, as yet neither *Semantic Webs* nor *Knowledge Grids* exist as envisioned, let alone do they co-operate. How could they come about? *"Growth"* may be an appropriate metaphor to describe the emergence and evolution of networks (for whatever kind of traffic). But growth does not happen out of the blue. It needs seeds in the first place, then possibly fertilizers and irrigation if nature does not make the necessary provisions herself.

In our case nature's role would be that of the business world, and commercial interest would be the main driving force. However, commercial interest in starting and sustaining pioneering research may be weak if no substantial benefits can be made out in the short or mid term. And indeed there are examples of technologies that would probably never have succeeded the way they did if their development had depended solely on "commercial interest" in the first place.

The Internet itself is a case in point. Its initial development depended largely on public funding. And while commercial interest had been strong enough to make "the Web" grow exponentially for a number of years this may not be as obvious for the *Semantic Web*, the *Knowledge Grid* and their possible "marriage" (regardless of what the new family name may be: *Semantic Grid* or *Knowledge Web* or whatever). The underlying concepts are after all not so easy to grasp, and their potential benefits (e.g. in terms of creating mass markets) are not so easy to sell given the current perceived slump in online business.

Moreover, a *critical mass* problem has to be solved for instance for the *Semantic Web*: adding semantics to content (and services) does not pay off if no tools are available to make good use of it; developing tools, on the other hand, does not pay off if there is little semantically-enriched content to work on.

These may be some of the reasons (apart from the obvious research challenges) why both initiatives, the *Semantic Web* and the *(Knowledge) Grid*, have been given firm places on the agenda of the European Commission's IST programme 1998-2002 ([Cordis_IST 2002]): as *Semantic Web Technologies* under *Key Action III (Multimedia Content and Tools, Action Line III.4.1)* in the IST work programme 2001 and as *Grid Technologies and their applications*, a *Cross Programm Activity (CPA9)* in work programme 2002.

Action Line III.4.1 offered four broad inter-related R&D areas as an orientation for submitting project proposals:

Creating a usable formal framework in terms of formal methods, models, languages and corresponding tools for semantically sound machine-processable resource description (e.g. content characteristics, properties of repositories, capabilities of devices, service features, ...);

fleshing out the formal skeletons by developing and applying techniques for knowledge discovery (in databases and text repositories), ontology learning, multimedia content analysis, content-based indexing, ...;

acting in a semantically rich environment, performing resource and service discovery, complex transactions, semantic search and retrieval, filtering and profiling, supporting collaborative filtering and knowledge sharing, ...;

making it understandable to people through device dependent information visualisation, semantics-based and context-sensitive navigation and browsing, semantics-based dialogue management, ...

While the second track of Action Line III.4.1 did not stipulate in any way the underlying computational platform, CPA9 (*Grid Technologies and their applications*) was very specific about it. It invited proposals to apply *Grid* technology to "*knowledge discovery in (multidimensional and multimedial) large distributed datasets, using cognitive techniques, data mining, machine learning, ontology engineering, information visualisation, intelligent agents, ...*"

Neither of these *action lines* prescribed a particular application domain. The very title of Action Line III.4.1, *"Semantic Web **Technologies**"*, made this quite explicit. And as far as Grids were concerned the term *applications* referred to the implementation *on top* of the basic architecture of computational Grids, of solutions pertaining to fairly general classes of problems.

Yet, clearly, technologies must not be developed for the sake of developing technologies. They should respond to real needs and they will be successful (commercially and otherwise) only if they do so. Therefore proposers were advised to make sure the solutions proposed would not benefit a limited constituency only, or solve just one isolated problem. Rather, projects submitted under a generic action line such as *"Semantic Web Technologies"* should, in a final analysis, yield more widely applicable results.

Calls for submitting proposals to these *Action Lines* were published in July (AL III.4.1) and November (CPA9) 2001, respectively. Both Calls met with considerable interest in relevant R&D communities across Europe and drew altogether nearly one hundred submissions involving several hundred participating organisations. The "success rate" (in terms of acceptance for funding) has been close to 25%.

It must be noted that the modules of the IST programme we have described so far are not the only ones designed to deal with the objectives and problems related to knowledge discovery, acquisition, management and use in the context of large scale distributed systems. Other *IST Key Actions*, in particular IV (*Essential Technologies and Infrastructures*) and the *IST FET domain* (*Future and Emerging Technologies*), but also II (*New Methods of Work and Electronic Commerce*), invited and are hosting relevant projects.

While *Key Action IV* and *FET* are also "neutral" as far as applications are concerned, projects under *Key Action II*, are indeed required to focus on particular application domains which could be broadly described as *corporate knowledge management* and *"e-business"*.

9 Projects

As the current IST programme (1998 - 2002) is coming to an end this is the time to make a first assessment of the extent to which retained projects are contributing or are expected to contribute technically, to creating, managing and using the *"knowledge spaces"* that could be spanned by Webs and Grids. Summaries for a selection of these projects (in alphabetical order) are provided in Appendix A to this note.

Talking about *space* insinuates *dimensions*. Obviously, we cannot discuss the dimensions of *"knowledge space"* (there may be infinitely many) but only some of the *"problem and solution spaces"* at issue. Here we can at least identify more or less orthogonal subspaces accommodating the various aspects of relevant IST projects. These subspaces correspond roughly to the areas outlined in Section 8.

The provision and usability of a formal framework for dealing with the semantics of distributed digital content is of general concern and the main focus of a number of projects, in particular ON-TO-KNOWLEDGE and WONDERWEB. Not surprisingly (in view of our remarks in Section 3), ontologies take centre stage in these projects whose workplans include ontology language definition and an analysis of the requirements to be met by "ontology servers". While both projects contribute (directly and indirectly) to the W3C "recommendations process" that organisation actually takes the lead in SWAD-EUROPE, a "bottom-up" experimentation and implementation project designed to showcase the viability of *Semantic Web* model and language recommendations.

There are two large subspaces that can be labeled *"making content semantics explicit"* and *"acting on explicit semantics"*, respectively. Usually, *explicit semantics* means metadata grounded in a firm semantic domain, such as a formal ontology. But it also refers to the ontologies themselves.

"Making semantics (i.e. metadata and ontologies) explicit" can happen in many ways, depending largely on content types and usage environments. There are, however, two main categories of approaches: either through (automated) content analysis or by interactive capture, at content production time. These categories represent extremes. A middle way would be to provide means of interactive "knowledge capture" on the fly, directly from workflow processes for instance, where users would not be required to make the extra effort of entering semantic metadata. This approach appears to be particularly appropriate in the context of "corporate knowledge management". But there is in fact an entire spectrum of "semi-automation".

The second subspace (*"acting on explicit semantics"*) can also be partitioned depending on who is acting, *software agents* or human beings. Software agents are certainly main characters in a *Semantic Web* scenario. They always act, in a final analysis, on behalf of humans.

They appear in several roles: as service providers, discoverers, mediators or composers. Hence, they also need *"service semantics"* (i.e. ontology-based descriptions of prerequisites and effects) in order to do what they are supposed to do.

Agents and processes handling queries can build on reasoning and inferencing capabilities made possible by ontologies. However, a particularly serious problem in this context is scalability: of ontologies, ontological reasoning and ontology (change) management. This problem ranks high on the agenda of several projects, including MOSES and the aforementioned WONDERWEB and ON-TO-KNOWLEDGE.

Explicit content (and service) semantics may greatly enhance the quality, effectiveness and efficiency of man-system interaction and of system-mediated communication among people. Topics of particular interest include: "Navigation and browsing", "query construction support", "dialogue management", "personalization, profiling and customization", "information visualisation", "semantic Web portals" and "collaboration support".

Table 2 presents a classification of the projects listed in Appendix A, based on the above outlined scheme. This classification must be fuzzy as a given project may well address problem areas and propose solutions that belong to different subspaces: *"making content semantics explicit"* for instance, is never an end in itself while *"acting on explicit semantics"* presupposes the existence of explicit semantics. Hence, our classification merely reflects the perceived gist of a project, its main thrust.

The number of projects allocated to the groups (I)-(IV) respectively seems to indicate the relative urgency of the issues involved. *"Making content semantics explicit"* (groups (I) and (III)) appears to be a dominant objective, and probably rightly so. It comprises both, ontologies (e.g. through ontology learning and emergent semantics in peer-to-peer networks in SWAP, GRACE and MOSES) and ontology-based metadata (e.g. through semantic annotations in ESPERONTO, through image analysis in SCULPTEUR, and through the extraction of domain-specific metadata in SPIRIT and WISPER). The projects in group (I) aim to achieve this objective through more automation than interaction whereas group (III) projects emphasize interaction, including system-mediated human-to-human interaction (cf. above).

Group (II) projects focus on services, their description and discovery, and on other service related operations (cf. above). SWWS for instance, will be about the implementation of a fully fledged Web service modeling framework (WSMF),

	making semantics explicit	acting upon explicit semantics
automatic tools	(I) ESPERONTO, GRACE, MOSES, SCULPTEUR, SPIRIT, SWAP, WISPER	(II) IBROW, MONET, SEWASIE, SWWS
interactive tools	(III) COG, COMMA, FF-POIROT, ONTO-LOGGING, SPACEMANTIX	(IV) INDICO, VICODI, WIDE
general framework	(V) ON-TO-KNOWLEDGE, WONDERWEB, SWAD-EUROPE	

Table 2: Semantics projects in problem&solution space

put forward in [Fensel, Bussler 2002]. MONET will offer mathematical solvers and SEWASIE concentrates on semantic search and inferencing. IBROW is a brokering service, configuring "knowledge components" (ontologies and generic algorithms) according to stated specifications of user needs.

The main objectives of the three projects in group (IV) are related to what may be called *"interfacing with knowledge"*, aspects of which we have already mentioned. In the projects in question these are mainly *browsing* (INDICO), *context visualisation* (VICODI) and *cooperative work* (WIDE) support.

Given the crucial role of *ontologies* within "semantic systems", these constructs appear in one way or other in virtually all projects. Similarly, the agent paradigm that has been quite popular in the field of distributed computing already for some time, is gaining new ground in "semantically-enriched environments". Agents also appear almost everywhere: as constructors of ontologies, extractors of metadata, as service composers and as assistants at the user interface.

Several projects (e.g. MOSES, FF-POIROT and SEWASIE) also address multilinguality issues that bear on the creation and use of ontologies. And despite the fact that non-textual content poses much harder "semantics problems" than even unstructured text, some projects (e.g. SCULPTEUR, SPACEMANTIX and ESPERONTO) have taken on that challenge. We do note, however, that there still seems to be a fairly wide gap between, say, the *Semantic Web* communities proper and those who do research on multimedia semantics (e.g. image understanding).

It would be presumptuous to claim absolute novelty (or uniqueness) for any of the approaches taken by the projects discussed in this section. (One may argue that science and technology proceed by piecemeal research and engineering; and that *"radical breakthroughs"* are never so *radical* when seen against the backdrop of their birthing grounds. The Web itself, as explained in Section 2, corroborates this statement.) However, all of these projects do provide an opportunity for researchers in Europe to explore new territory, to prove or disprove the viability of existing approaches, to establish the need for new ones, and to contribute to making worldwide distributed systems more usable.

10 Prospects

The evolution of basic digital technologies has been going on for more than half a century, characterized by ever increasing values of parameters such as processor speed, storage and memory capacity, bandwidth and connectivity. Given its current momentum (occasional ups and downs notwithstanding) it is likely to continue for quite some time. It has brought about many novelties relative to the pre-digital era. At least three classes of applications are relevant within the context of this note:

1. Digital technologies allow us to create, maintain and use content of all types and media in hitherto unreachable dimensions, thanks to tools that are many orders of magnitude more powerful than pen, paper, the printing press or library catalogues.

2. The digital technologies have enhanced drastically our ability to analyse what is going on in the world (in both nature and society), to peruse vast amounts of data, searching for structure, thus refining our models of the world[5].

3. Digital technologies allow us to build machines that can learn and - to a certain extent - act autonomously in limited environments. (In a way this may be considered a special case of the second class of novelties.)

Developments corresponding to the first and second of these classes have led directly to the more or less recent phenomena discussed in this note, to Webs and Grids. And they will perhaps lead on to *Semantic Webs* and *Knowledge Grids*. The third novelty refers to autonomous "intelligent" agents and *robotics*, a field of applications of basic digital technologies that may not yet be as fully

[5] These data are, by the way, mainly being collected through devices that owe their existence, effectiveness and efficiency to digital technologies.

visible as other application domains are. All three are about creating and using representations of knowledge (in the sense of footnote 4). They can be subsumed under the heading *Knowledge Technologies*.

Their impact is steadily growing: they are transforming industrial production processes, the way we create and distribute content for human consumption, the way we do science, the way businesses are managed, the way public administrations work, ...

We note, however, that in the past attention has always been focused on technologies that would bring about changes in precisely these areas. We remember: Management Information Systems (MIS), Office Automation Systems, Decision Support Systems, Expert Systems, Computer Supported Cooperative Work (CSCW) systems, Corporate Information Systems, Computer Integrated Manufacturing (CIM) systems, . . . (not to mention the multitude of isolated or linked business application systems and tools for building such applications).

So the obvious question to ask is: what is going to happen next? Will there be a next *Big Thing* and if so, what will it be?

The evolution of technology appears to be driven by at least two interacting processes: the emergence of needs and the sophistication of tools. Usually, there is "positive feedback" in the sense that the increase in sophistication of tools goes hand in hand with an extension and more detailed specification of needs.

What then are the problems to be solved by *Knowledge Technologies*? And what are the problems created by these technologies? Some answers to such questions have been outlined in previous sections. But they do require much greater attention.

Will the visions of a *Semantic Web* where ontologies would be as crucial as plain documents are for the current Web, and of a vast virtual computer called *Grid*, hold solutions in stock also for big multinational companies? For the multitude of small and medium-sized enterprises? For the general public? For scientists? For engineers? For professionals from all walks of life? The knowledge workers? What exactly are their needs, how do these needs change as technologies change?

Forecasting the future has never been a very gratifying undertaking. Not to predict the future but to create it, is perhaps a more rewarding task. Yet there have been few large scale joint research and development activities that were truly vision-led or taking as (seemingly) straightforward a path as for instance the man-on-the-moon project (not to mention military objectives, of course).

As pointed out in Section 8 publicly funded research programmes have a key role to play here. Ideally, they would provide some guidance and focus, based on a sufficiently broad consensus among relevant R&D communities. The design of the multi-annual research programmes of the European Union reflects this objective.

One of the *Priority Research Areas* of the European Commission's forthcoming *6th Framework Programme* will again be *Information Society Technologies (IST)* which in turn will offer two main research foci: *applied IST for major societal and economic challenges*, and *generic IST research and technology development*. The latter will also cover *knowledge and interface technologies*.

The *Council Decision* concerning the specific programmes implementing the Sixth Framework Programme of the European Community for research, technological development and demonstration activities (2002-2006) addresses *Knowledge Technologies* as follows ([Cordis_RTD 2002]):

> The objective is to provide **automated solutions** for **creating** and **organising** virtual **knowledge spaces** (e.g. collective memories) so as to stimulate ... new content and media services and applications. Work will focus on technologies to support the process of **modelling** and **representing**, **acquiring** and **retrieving**, **navigating** and **visualising**, **interpreting** and **sharing** knowledge. These functions will be integrated in new **semantics-based** and **context-aware** systems including **cognitive** and **agent-based** tools. Work will address extensible **knowledge resources** and **ontologies** so as to facilitate **service interoperability** and enable next-generation **Semantic-web applications**. Research will also address technologies to support the design, creation, management and publishing of multimedia content, across fixed and mobile networks and devices, with the ability to self-adapt to user expectations. The aim is to stimulate the creation of rich interactive content for personalised broadcasting and advanced trusted media and entertainment applications.

This includes the technologies and research directions addressed in this note, but goes clearly beyond. A number of action lines pertaining to several components of the IST programme under FP5 have already set the scene (cf. Appendix B).

The funding instruments foreseen at the time of this writing have been designed with a view to giving research communities a real opportunity to formulate and pursue visions of their own and to build strong bridges across disciplines if necessary or desirable. These instruments have also been designed with a view to supporting the emergence of a true European Research Area (ERA). Applying them to Knowledge Technologies may indeed complement this ERA with a vast European Knowledge Space.

A Some IST projects on semantics for distributed systems

COG

Corporate ontology grid (CPA9) Collaborative ontology development in a

corporate environment (automotive industries); automatic scripting for transformation and query; creating 'virtual views' of data from disparate sources.
(http://www.cogproject.org/)

CoMMA
Corporate Memory Management through Agents (KA2)
Multi-Agent System, based on semantic enterprise and user models, and ontologies, applied to Corporate Memory Management, using techniques for learning from user behaviour.
(*http://www.ii.atos-group.com/sophia/comma/HomePage.htm*)

ESPERONTO
Application Service Provision of Semantic Annotation, Aggregation, Indexing and Routing of Textual, Multimedia, and Multilingual Web Content (KA3)
Focuses on "legacy" Web content and develops ontologies to support multimedia and multilinguality.
(*http://esperonto.semanticweb.org*)

FF-POIROT
Financial Fraud Prevention-Oriented Information Resources using Ontology Technology (KA3)
Interactive construction of multilingual ontologies through domain modelling, (automatic) text-mining and (semi-automatic) validation and alignment, as a basis for Semantic Web services for knowledge storage, management, retrieval and sharing.
http://www.starlab.vub.ac.be/research/projects/default.htm# Poirot

GRACE
Grid Search and Categorization Engine (CPA9)
Develops a decentralized search and categorization engine for unstructured textual information; builds on-top of Grid-technology (peer-to-peer), uses locally computed indexes.
http://pertinax.cms.shu.ac.uk/projects/cmslb2/

IBROW
An Intelligent Brokering Service for Knowledge-Component Reuse on the World-Wide Web (FET)
Configures distributed, heterogeneous applications using pre-existing components (ontologies and problem solving methods - for information filtering, automatic classification and design problems) retrieved from distributed digital libraries.
(*http://www.swi.psy.uva.nl/projects/ibrow/home.html*)

InDiCo
Integrated Digital Conferencing (KA3)
Develops semantics-based multimedia indexing and browsing methods for conference and distance learning applications.
http://indico.sissa.it/

MONET
Mathematics on the net (KA3)
Ontologies for mathematics services description, querying, explanation and use.
http://monet.nag.co.uk

MOSES
A modular and scalable environment for the Semantic Web
(CPA9) Focuses on the scalability and linguistic aspects of ontology construction and evolution through content (mainly text) structure analysis.

On-To-Knowledge
Content-driven knowledge management through evolving ontologies (KA4)
Design of languages (OIL) and implementation of tools for ontologies, for automatic derivation of semantics of semi-structured data (text-mining), knowledge maintenance, view definitions and agent supported information access.
http://www.ontoknowledge.com

ONTO-LOGGING
Corporate Ontology Modelling and Management System (KA2)
Distributed formalization of corporate ontologies; dynamic optimisation using agent technology for user modelling and category extraction.
http://www.ontologging.com

SCULPTEUR
Semantic and content-based multimedia exploitation for European benefit (KA3)
Constructs a semantic layer enhancing search in distributed digital libraries of images of 3D objects, by linking low and high-level features; implementing agents for classification and search of structured and unstructured content.
http://www.sculpteurweb.org

SEWASIE
Semantic Webs and agents in integrated economies (KA3)
Designs a distributed agent architecture for semantic search and inferencing based on multilingual ontologies.
http://www.sewasie.org/

SPACEMANTIX
Combining spatial and semantic information in product data (KA3)
Enriches 3D graphics in product catalogues with semantic information (e.g. assembling instructions) for easy and natural access to and manipulation of 3D models.
http://www.agc.fhg.de/uniGoethe/forschung/SpacemantiX/

SPIRIT
Spatially-aware information retrieval on the Internet (KA3)
Derives/extracts ontology-based geo-metadata from Web pages and digital map datasets, for spatially-aware search engines.
http://www.research-projects.unizh.ch/math/unit70600/area20/p2563.htm

SWAD-Europe
W3C Semantic Web advanced development for Europe (KA3)
Informs W3C work on new “Semantic Web” recommendations, through research, open source implementation and testing
http://www.w3.org/2001/sw/Europe/

SWAP
Semantic Web and Peer-to-Peer (KA3)
Realizes a “Semantic Web” based peer-to-peer system, building on available Open Source peer-to-peer solutions, for sharing individual views on knowledge through emerging semantics.
http://swap.semanticweb.org

SWWS
Semantic Web enabled Web services (KA3)
Develops semantic means for describing, recognizing, configuring, combining, comparing and negotiating Web services, supporting Web service discovery and scalable

mediation.
http://swws.semanticweb.org

VICODI
VISUAL CONTEXTUALISATION OF DIGITAL CONTENT (KA3)
Provides mechanisms for contextualising distributed multilingual digital content (European history), taking into account topics (category, hierarchy), location and time, through semantic indexing and ontological markup and using neural classifiers; development of a suitable SVG-based visualization interface.

WIDE
SEMANTIC WEB-BASED INFORMATION MANAGEMENT AND KNOWLEDGE SHARING FOR INNOVATIVE PRODUCT DESIGN AND ENGINEERING (KA3)
Integrates, using Semantic Web technologies, proprietary in-house databases, off-line and on-line catalogues, and the World Wide Web to support the information and knowledge sharing needs of industrial designers and product engineers.
http://www.cefriel.it/topics/research/index.xml?tid=27

WISPER
WORLDWIDE INTELLIGENT SEMANTIC PATENT EXTRACTION & RETRIEVAL (KA3)
Automatic semantic mark-up of structured and multi-lingual digital content (patents), in support of searching and visualizing search results.

WonderWeb
ONTOLOGY INFRASTRUCTURE FOR THE SEMANTIC WEB (FET)
Analyzes requirements for large-scale deployment of ontologies: ontology languages, semantic integration, migration, reconciliation and sharing of ontologies, foundational ontologies, tool support (for editing, integrating and extracting ontologies), ontology server architectures and services such as persistent storage and reasoning support.
http://wonderweb.semanticweb.org

B A selection of FP5-IST action lines pertaining to *Knowledge Technologies*

KEY ACTION II - NEW METHODS OF WORK AND ELECTRONIC COMMERCE

- Corporate knowledge management
- Knowledge Management for eCommerce and eWork
- Technology Building Blocks for Trust and Security

KEY ACTION III - MULTIMEDIA CONTENT AND TOOLS

- Authoring and design systems
- Content management and personalisation
- Media representation and access: new models and standards
- Access to digital collections of cultural and scientific content
- Content-processing for domestic and mobile multimedia platforms
- Information visualisation
- Semantic Web Technologies

KEY ACTION IV - ESSENTIAL TECHNOLOGIES AND INFRASTRUCTURES

- Engineering of intelligent services
- Methods and tools for intelligence and knowledge sharing
- Information management methods

FUTURE AND EMERGING TECHNOLOGIES

- Open domain (FET OPEN)
- Universal information ecosystems
- The disappearing computer
- Global computing: co-operation of autonomous and mobile entities in dynamic environments

CROSS PROGRAMME ACTIONS

- CPA9: GRID Technologies and their applications

References

[Cordis_IST 2002] http://www.cordis.lu/ist/

[Cordis_RTD 2002] http://www.cordis.lu/rtd2002/background.htm

[DeRoure et. al. 2001] http://www.semanticgrid.org/v1.9/semgrid.pdf; David De Roure, Nicholas Jennings, and Nigel Shadbolt: Research Agenda for the Semantic Grid: A Future e-Science Infrastructure, Report commissioned for EPSRC/DTI Core e-Science Programme, December 2001

[Fensel, Bussler 2002] http://www.cs.vu.nl/ dieter/ftp/paper/wsmf.pdf; D. Fensel and C. Bussler: The Web Service Modeling Framework WSMF, Amsterdam 2002

[Globus 2002] http://www.globus.org/; Globus project developing fundamental technologies needed to build computational grids.

[Grid 1998] Ian Foster and Carl Kesselman (Eds): The Grid - Blueprint for a New Computing Infrastructure; July 1998; http://www.mkp.com/books_catalog/catalog.asp?ISBN=1−55860−475−8

[Gruber 1993] T. R. Gruber. A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220, 1993

[Hyper-G 1995] Udo Flohr: Hyper-G Organizes the Web,Byte Magazine, Nov 1995; http://www.byte.com/art/9511/sec5/art4.htm

[MEMEX 1945] http://www.cs.brown.edu/memex/; The Memex and Beyond web site is a major research, educational, and collaborative web site integrating the historical record of and current research in hypermedia. The name honors the 1945 publication of Vannevar Bush's article *As We May Think* in which he proposed a hypertext engine called the Memex. (Maintained at Brown University)

[Microcosm 1997] http://www.mmrg.ecs.soton.ac.uk/projects/microcosm.html; Wendy Hall: The History of the Microcosm Project

[Miller 1998] Eric Miller: An Introduction to the Resource Description Framework; D-Lib Magazine; May 1998; http://www.dlib.org/dlib/may98/miller/05miller.html

[Negroponte 1996] http://archives.obs-us.com/obs/english/books/nn/bdcont.htm; Nicholas Negroponte: Being Digital (Selected Bits); New York 1996

[W3C_Data 1999] http://www.w3.org/1999/04/WebData

[W3C_Dybvik 1993] http://www.w3.org/People/howcome/p/telektronikk−4−93/Dybvik_P_E.html; Paper by Per E Dybvik on the differences between Internet services and those of telecom administrations, and the culture clash between telecommunications and computer markets.

[W3C_Sem 1998] http://www.w3.org/DesignIssues/Semantic.html

[W3C_SWA 2001] http://www.w3.org/2001/sw/Activity

[W3C_TBL 1989] http://www.w3.org/History/1989/proposal.html;
Information Management: A Proposal; by Tim Berners-Lee, CERN; March 1989, May 1990; see also: http://www.w3.org/People/Berners-Lee/Weaving/, Weaving the Web; by Tim Berners-Lee with Mark Fischetti; San Francisco 1999

[XANADU 1960] http://xanadu.com.au/xanadu;
"the original hypertext and interactive multimedia system", under continuous development since 1960