ICT Priority
Strategic Objective ICT 2009.4.3:
Intelligent Information Management

Technical background

Technologies for Information Management in the EU's seventh Framework Programme

Intelligent information Management is a Strategic Objective (SO) of the Information and Communication Technologies (ICT) Priority, a major component of the seventh Framework Programme (FP7).

Excerpts from the ICT FP7 2009-10 Work Programme:

Digital content is today being produced in quantities that are deeply transforming the enterprise and the creative industries. Conditions for production and consumption are also rapidly changing as more and more content is produced by users. Organisations, public and private, are faced with maintaining, managing and exploiting increasing amounts of data and knowledge, in environments that are continually changing. New ways of expressing and representing cultural and scientific content in digital form are creating new opportunities for people to experience and share assets.

Progress in ***knowledge modelling and processing*** has enabled the creation of innovative commercial and community services and is progressively transforming scientific discovery. Semantic web technologies are likewise starting to be used on an industrial scale by information providers and search engines alike to offer more sophisticated services. Conceptualising and producing digital content as a container of rich objects that can be individually selected and manipulated is emerging as a trend.

The purpose of this document is to provide some background information and general principles for the interpretation of work programme for this Strategic Objective. All examples contained in this document are purely meant as illustration of these principles, they don't express a specific preference for any domain of research and development, nor do they imply that proposals in domains that are not discussed are discouraged. In any event, the only authoritative source is the text of the FP7 2009-10 ICT work programme.

I. Background

1. Current status within IST

The **2009-10 work programme of Unit DG INFSO/E2, Technologies for Information Management, aims to refocus some of the knowledge extraction and management research activities** of WP 2007-8, in response to the data growth trends that are readily observable and will inevitably accelerate as the vision of the Internet of Things (which is the focus of strategic objective ICT 2009.1.3) comes to fruition. This SO will call for project proposals under the 5th ICT call. Successful submissions are likely to lead to contracts effective from the second quarter of 2010.

ICT calls for Knowledge and Content RTD in FP7 (calls 1 and 3) yielded 28 projects totalling around €101M (EU funding). The first projects began work in January 2008 and projects form the latest call will continue until the end of 2012. Refer to *http://cordis.europa.eu/fp7/ict/content-knowledge/projects_en.html* for more details regarding the projects and the research topics they address.

It is expected that the fifth FP7 call will lead to equally ambitious endeavours. The SO budget envelope for the 5th call is in the region of €70M.

Even when the objectives of the SO overlap with those of FP7 2007-8 objectives, new proposals should avoid repeating the tasks of **ongoing projects** and aim to progress beyond, or complement with

compelling contributions, topics covered in projects resulting from any of the two FP7 calls that have been devoted to these themes.

Considering the comprehensive portfolio of projects that resulted from FP6 and FP7 calls, significant progress in the field is expected by the middle of the next decade. This means extremely robust and scalable systems used daily by millions of people to support sophisticated and time-sensitive information management tasks.

2. Consultation process and update, leading to Work Programme 2009-10

In the ICT 2007-8 Work programme for FP7, knowledge and content related research themes were presented under the SOs '*Intelligent Content and Semantics*' ICT 2007.4.2 and ICT 2007.4.4. The 2009-10 Work Programme pays due attention to the i2010 policy objective and was designed, based on progress to date in existing FP6 and FP7 projects, an assessment of the considerable changes in social and market conditions and requirements, as well as the progress of ongoing scientific research. The work programme was designed with the help of external experts in a broad and systematic consultation process running for several months between the end of 2007 and the beginning of 2010 and addressing all ICT SOs.

As a result of this process it was decided to target aggressively the development of technologies and infrastructure for **(i) capturing and integrating extremely large amounts of information (ii) integrating information with time sensitive organisational processes such as planning and decision making**. Particular emphasis was placed on

a. **designing and implementing systems designed to be inherently scalable in volume and flexible in their ability to integrate noisy, inconsistent and rapidly changing information of disparate types** (text, images, audio, video, 3D, engineering designs, statistical data, sensor data, formal knowledge representation, databases …) **;**

b. systematic empirical testing of the properties of such systems going from storage requirements to run time performance to usability in ecologically valid conditions to ease of installation and maintenance.

c. privacy, security and integrity being ensured by design and always under the control of the end user and open to verification

3. European perspective

Intelligent information management is fundamental to **data intensive sectors**, which include some of Europe's key industries – manufacturing, including automotive and aerospace, engineering, chemical and process industries, pharmaceuticals, financial, geospatial sectors, media and telecoms as well as other important areas such as scientific Research (particularly in bioinformatics), eGovernment, eHealth etc..

EU funded research aims to leverage Europe's skills and vast wealth of knowledge resources and to encourage European industries to play a full and competitive part in developing novel content products and experiences and knowledge-based services. The added-value in European research is typically in stimulating technologies and systems which are suited to the multi-cultural and multi-lingual European markets, and to the high expectations of European users.

In Europe there is a further need to promote a **critical mass of interdisciplinary research**, to encourage different communities to learn from each other while contributing valuable skill sets to the problems of content and knowledge capture, interpretation and sharing. Collaborative efforts will help to create new alliances in Europe in key emerging areas like automatic multimedia interpretation and encoding based on the psychology of human reasoning and decision making, collaborative and personalised information delivery, trust, accountability and privacy. The participation of **industrial stakeholders** – technology providers, ICT suppliers and integrators, content providers and especially **leading edge users** is vital for testing and benchmarking innovative solutions in realistic settings, i.e. with significant volumes of data and real time performance needs.

II. Research aims and focus for 2009-10

The Intelligent Information Management SO is aimed at **contributing extremely scalable and robust integrated solutions for all domains in which the availability of data now outstrips the availability of the human intelligence needed to exploit them.** Novel techniques and infrastructures are solicited for managing these data volumes. In this scenario, managing typically means identifying and exploiting appropriate approximations for problems whose exact solution is provably intractable. Semantic-based methods are seen as a means for extracting actionable knowledge from massive data sets and providing complex and yet flexible, interoperable services and applications. Particular attention will be given to the problem of integrating heterogeneous data types (e.g. sensor data streams, geospatial information and imagery, financial transactions, news feeds, 3D models, engineering data, information for policy intelligence etc…) into models and analysis that are more useful and informative than the sum of the individual data feeds alone.

The overall objective is to provide the back end intelligence and analysis power to match the flood of data that is being created by ever cheaper methods to collect and transmit information from our physical and virtual environments.

Research will aim to maximise **automation** of the information lifecycle in support of existing activities and novel challenges and opportunities in the respective domains. The intention is to delegate to machines as many tasks as possible so as to allow human creativity and insight to flow and grow unimpeded. The **expected outcome** of a successful proposal is typically a software system or component that has been used and tested for a sufficiently long time by a sufficiently large number of real life users so as to be realistically considered at a pre-product stage in terms of robustness, ease of deployment and maintenance and usability.

1 Approach

This SO is mostly centred around **medium- to long-term research with relatively short-term by-products**. The emphasis is on

a) **scalable and robust enabling technologies and middleware**;

b) **application platforms and architectures** which are portable across domains;

c) **psychological, social and economic foundations** for the technologies proposed

d) **foundations of approximated and probabilistic reasoning**;

e) **semantic mediation and integration of distributed data repositories;**

f) **integrated knowledge infrastructure for organisations and enterprises**

g) **systems that are accountable and privacy preserving by design**

The strategy is to undertake the necessary foundational and component-level research on the one hand, and to test and prove technology and components in representative demonstrators and broadly based test-beds, addressing functionality and usability, robustness and complexity, scalability and ease of customisation on the other. **Emphasis should be placed on the user** (enterprise, institutional), particularly with regard to **hiding complexity, exploiting contextual and social knowledge** and developing intuitive interactive information interfaces. Large-scale, though relatively short-term (commercial exploitation expected in 3-4 year) efforts must demonstrate the effective deployment of content and knowledge technologies in support of data intensive environments. They are expected to use realistic, large-volume data sets from information bound sectors including (e-)business, manufacturing, e-science as well as other economically important fields with high growth potential. Estimates on the volumes of realistic data sets naturally vary from domain to domain and can be established with a modicum of research. By way of example, GIS systems today handle petabytes of data, knowledge repositories manage billions of RDF triples, etc…Successful proposal are expected to operate on data sets that are considered realistic in size by the respective industries or communities.

Test-beds, **benchmarking** and **experimental testing** will play a critical role in proving the technical feasibility and socio-economic / organisational impact of knowledge systems in the provision of adaptable and context-aware content products and in knowledge based web services for planning, problem solving, automated diagnosis and decision-making. The tools and methods developed should likewise chart a credible path towards less complex applications of greater usability for the general public beyond the original circle of specialists.

2 Research themes

The research focus is on five main themes:

a) Capturing tractable information;

b) Delivering pertinent information;

c) Collaboration and decision support;

d) Personal sphere;

e) Impact and S&T leadership;

2.1 Capturing tractable information

Robust and performant technologies to acquire, analyse and categorise extremely large, rapidly evolving and potentially conflicting and incomplete amounts of information. These technologies will extract, correlate and integrate data from diverse sources and formats (multimedia and 3D content; heterogeneous databases; data streams from sensors and scientific equipment; social interactions and networked appliances; information from business processes and software services) while tracing provenance, evaluating trust level and assessing reliability. The scalability, flexibility and performance of such methods and techniques will be demonstrated by rigorous empirical testing over large-scale testbeds.

Instruments: IPs, NoEs, STREPs,

This SO theme addresses the requirements of environments whose processes depend on extremely large and heterogeneous data volumes. This is typically the case in all those environments that have automated their information gathering tasks. For example, DNA sequencing which used to be a laborious task carried out manually by highly trained scientists is now a task automated by machines of ever increasing speed. As a result, the amount of sequenced DNA has increased enormously (it is by now entirely conceivable that millions of people will have their entire genome sequenced within the next ten years) and forced the development of specialised bioinformatics tools to inspect and analyse it. Similarly, thanks to the automation of aerial photography analysis it is today possible to create in a matter of days a complete 3D model of a large urban area. This in turn creates a need for a layer of interpretation: we can see the 3D buildings or structures but will not know what type of buildings they are and their functions without an additional level of analysis or knowledge integration. These are just two specific examples of a phenomenon that is much more general: when machines become good enough and cheap enough to replace humans in information gathering tasks, a flood of data immediately follows. These data need to be understood and managed. The next generation of machines that is expected (indeed, planned) to start generating enormous amounts of ambient information are devices belonging to the so called Internet of Things, networked sensors or RFID tags as they detected by the corresponding readers. Similarly, we already have many software components that automatically generate data (logs of internet searches, telephone calls, financial transactions, etc…) reflecting activities by humans or other machines. Scientists and developers will no doubt have in mind many other examples in which the automation of one aspect of their activities (data collection, generation) has created an intelligence deficit in other aspects (data analysis, theory development). Any domain in which this phenomenon is observed is a potential domain of activity under this theme. For this reason is absolutely crucial for consortia wishing to submit a proposal under this theme to secure the participation of an organisation committed to making available very large data sets or streams whose interpretation is of clear concern either to the organisation itself or to other consumers of those data.

There are two important additional dimensions to consider: data integration and robustness.

The issue of data integration can be described as follows: while it is entirely possible for there to be environments in which the large volumes of data are all uniform and all require the same, well specified, type of analysis, this theme is also intended to cover the case in which the value of data comes from the comparison and synchronisation of disparate data sets or streams. An example might be an application that updates epidemiological models by cross-referencing streaming reports of anonymised pharmacy purchases and hospital admissions, travel patterns obtained from transport authorities and GPS information from volunteer cell phones. Since it is inevitable that each of the contributors to this composite data stream (pharmacies, hospitals, transportation, and phones) would have independently developed their own formats and protocols, a non-trivial amount of high quality conceptualisation and engineering would have to go into making these streams inter-operable. Proposals operating along similar lines will have to plan accordingly.

The issues of conceptual integrity and robustness come from the physical and/or distributed nature of the data generators: sensors that gather data from their physical environment may be prone to failure or otherwise generate noisy or faulty data. Similarly, environments in which information about a given entity comes from independent sources may generate inconsistent information (e.g. my laptop may think I am in Budapest while my telephone thinks I am in Barcelona). Proposals under this theme will not be allowed to assume that the data they operate over will be clean, complete and consistent. There is also an issue of engineering robustness: data repositories or streams cannot be generally assumed to be stable. They may have different latencies, they may arbitrarily join or leave a network, they may provide a very sparse or a very redundant coverage, etc... Once again, proposals under this theme will be rewarded for describing detailed and credible plans for dealing with these uncertainties.

These guidelines thus exclude or discourage proposals that are working over small or unrealistically clean collection of data, that don't have a clear plan on how to obtain the volumes of data required (this includes both technical and privacy and intellectual property issues) and that don't give detailed and credible description of the software infrastructure to be developed and of the plans to keep it exercised at capacity for as much as possible of the life of the proposed project.

Foundational research in this theme will typically address formal models for integrating information sources, for managing noise and uncertainty and for implementing approximations to solutions that would be otherwise intractable. Whenever such approximations cannot be proved to be performant by analytic means, data-driven approaches ensuring replicability and benchmarking will be strongly favoured.

Component level research will address methods and tools aimed at implementing the research findings described above into integrated information management platforms allowing end users (i.e. competent professionals that were **not** part of the consortium's development team) to perform operations such as importing data from available sources, weaving them into the desired composites, monitoring them, etc… Where appropriate, work should build on and re-use existing technology, improving and recoding where necessary, with a view to **integrating project results within larger systems and applications.**

Priority will be given to open architectures or alternative approaches ensuring seamless **interworking between components or across sub-systems.** In the interests of maximising the impact of the research programme as well as driving the adoption of the technology, consortia must commit to making project results (architectures, tools and interfaces) widely available – including as open source, via appropriate licensing agreements – as part of a well-defined **dissemination and exploitation plan**.

2.2 Delivering pertinent information

Usable and customisable systems to improve the efficiency of the information lifecycle, starting from proactive diagnoses of information gaps and triggering goal-dependent search, acquisition, structuring and aggregation of relevant local, remote and streaming resources. Managing this information and making it actionable requires large-scale reasoning resulting in effective ranking, profiling and interpretation as well as versioning for time-dependent

compliance and justification. Such systems will support the navigation, manipulation and consumption of digital information by means of adaptive user-information interactions based on the state of the art in the psychology of human perception and attention. The effectiveness of such systems will be validated with appropriately-sized groups or communities of representative users.

Instruments: IPs, STREPs

The large volumes of data discussed in the previous theme are being collected for a present or foreseeable purpose. The second theme of this strategic objective is how these purposes can be understood and best supported. If the first theme was about the digestion of data volumes, the second is about its foraging, the assessment of one's current information state and the planning of how to modify it to accomplish a given goal. Many current information systems can expose their content in response to a request (e.g. return relevant documents in response to a keyword search) but few in the domain of unstructured data can identify and diagnose information gaps (they can tell you what they see but seldom what they should be seeing but they are not). To use the example of sensors from the previous theme, somebody interested in security from chemical contaminants may come to realise that for the purposes of a certain type of analysis, certain areas of a building should be covered by appropriate sensors but currently are not and arrange for some human or robotic intervention to remedy the situation. Or that one's financial position is influenced by data from a certain geographic area that are currently not being gathered and analysed.

The dual objective here then is to allow users to record their information goals (and their dependencies) in a format that can be mapped onto the underlying data resources and to allow the information agents so obtained to proactively seek and bring to the user's attention the information desired once it is found potentially aggregating it in the form of models of various levels of complexity. A prototypical example of this would be the identification and aggregation of complex events (e.g. a market manipulation attempt or opportunity; the unfolding of a civil emergency; etc…) from the time sensitive analysis of heterogeneous lower events (specific trading sequences and/or new announcements).

Just as in the previous theme it was crucial for successful proposals to obtain the commitment of organisations with large data volumes, in this theme it will be crucial for proposals to obtain the commitment of organisations with credible user groups. A credible user in this scenario is an expert who has a well motivated real-life task to carry out and needs to carry it out at a desired level of performance (precision, completeness …) within a desired time window which can often be rather short if not real time (as in the case of security, civil emergencies...)

The resulting systems will have to take into account and model not only the information state of the digital system but also the goals and capacity for attention and analysis of the human user. A system providing a spreadsheet with thousands of entries to a user who needs to make a decision in seconds would fail this test; an alternative system opting for a visual summary of the same thousands of entries might on the contrary succeed.

Proposals will be also rewarded for their ability to support additional analyses of the domain of interest. In many data rich environments users are not just interested in a statement of fact but also on the evidence supporting it; not only on the state of the system as it exists today, but also in its state at previous times and the reasons for its transition from previous state to the current one (e.g. what was the version of a given document at a given time in the past and why did it change the way it did? what was the structure of a given computer network at a given time and why did it change the way it did?).

This guidelines thus exclude or strongly discourage proposals where the evaluation of the systems developed is carried out with respect to artificial or technology driven tasks and with a small population of users not clearly representative of the end users for which the systems are built.

Foundational research will address the definition, evolution and monitoring of information needs, ideally exploiting findings from the psychology of human perception, memory and reasoning (particularly under uncertainty, a task at which humans are notoriously poor). Findings from these activities will form the basis for workflows that proactively assemble digital information and summarise or visualise it for distribution to comply with constraints coming from the time-scale of the activities to be supported.

Component- and system-level research will help to bring existing and emerging technologies to maturity, integrating them to create **digital information workflows**, and leading to scaleable platforms to manage, search, share, personalise, present and exploit complex information products that cross the boundaries between data types and sources. Work will include new strategies and innovative combinations of components which challenge, complement or improve established approaches.

While this research theme makes room for foundational research, it is primarily geared towards **integrative scenarios (test-beds)** intended to address issues such as run time performance, reliability, scalability and user acceptance, as well as data- and service- level interworking between data sources and target consumer populations. The overall aim is to stimulate **cross-disciplinary partnerships** leading to **large scale experimentation** demonstrating technical and organisational feasibility as well as cost effectiveness and replicability.

Clearly, this emphasis on integrative scenarios means that **several or all research themes of the SO can be covered by the most ambitious projects.** The desired result is robust, scaleable solutions which are portable across key application domains in data intensive industries.

2.3 Collaboration and decision support

Efficient and dependable problem solving and decision support systems for critical, information-bound domains in which our ability to share and exploit information is outstripped by the rate of its growth in size and complexity. Intended beneficiaries include organisations with complex business processes and access control policies; scientific communities collaborating on challenging projects and building very large datasets; teams of professional creators working on complex designs or multimedia materials; and web communities with sophisticated cooperation needs. The effectiveness of such solutions will be tested against the requirements of the respective groups or communities.

Instruments: IPs, STREPs,

The third research theme addresses the management of large quantities of information as it affects not just individual users but organisations and their business and decision processes.

There are at least three distinct aspects to this theme.

The first one is the fact that the way the individuals in a social group (be it a formal organisation or a spontaneous community) interact with information is, itself, information that can be recorded and exploited to everyone's benefit. Knowing who exchanged what information with whom can be exploited in order to satisfy or even pre-empt future information needs.

The second is that organisations often have collective information needs that are fulfilled in different stages by different people making different contributions. This theme is calling for environments where these collaborations can take place efficiently.

The third theme is that in enterprises and formal organisations, the information interactions within the organisation or with its customers are often regulated by systems of rules that themselves present information management challenges. For example, an enterprise may operate on the basis of thousands of business rules. While the number of such rules is not imposing as the volume of other data types (e.g the number of pages on the web or the number of calls in a large telephone network) what is challenging is the number of possible interactions among them. Large organisations are often subject to sets of rules that cannot all be simultaneously satisfied, typically as a result of different rules having been introduced by different actors at different times and with different purposes in mind. Repairing such conditions when they are detected is in turn a daunting task because the interaction of said rules also makes it difficult to predict the effects of any proposed change. A special case of this problem is the management of information disclosure policies: who has the need/right to access/distribute what type of information.

In this theme as well it will be crucial for proposals to obtain the commitment of organisations with credible user groups. A credible user group in this scenario is a group of people who have a well

motivated task to carry out and need to carry it out at a desired level of performance (precision, completeness …) and accountability (was the task carried out from valid information and for valid reasons? did everybody have access to the all information needed?)

The resulting systems will have to take into account and model not only the information state of the collaboration system but also the time sensitive goals, social dynamics and legal or moral rules of the human users.

These guidelines thus exclude or strongly discourage proposals where the evaluation of the systems developed is carried out with respect to artificial or technology driven tasks and with a small group of users not clearly representative of the groups for which the systems are built.

Foundational research will focus on the content representation, reasoning and planning required to support various types of social network analysis and policies and (business) rules management and analysis. Possible activities also include the psychology and economics of expert cooperative work and decision making.

Component and system-level development will provide usable and easy to maintain/extend systems for supporting collaborative information management and decision support. Benchmarking should address scenarios with a non-trivial number of concurrent users in realistic environments.

2.4 Personal sphere

Intuitive systems that help individuals secure, manage, visualise and interpret their personal information, attention trail and social history so as to enable the provision of personalised and context-dependent information from multiple sources and services. A specific requirement and design principle is that such systems preserve privacy and implement auditable information disclosure policies that are under user control and whose application can be verified at all times. Their usability and rate of uptake will be monitored by means of verifiable quantitative indicators.

Instruments: STREPs

The fourth theme of this strategic objective addresses the issue of personal information management. Until recently, it was only organisation and professionals who had volumes of digital information sufficiently large to require dedicate management tools. Today, in industrialised countries, it is on the contrary quite common for people to start building from a fairly young age digital collections of pictures, music, to exchange daily dozens if not hundreds of electronic messages, to interact with several social networks and web services. All of these activities create an attention trail that could be used to improve the provisioning of many services, from web searches to music recommendation to health care and generally give people insights on their preferences and activities by means of appropriate statistics and visualisations.

Naturally, the promise of better services and experiences based on an analysis of one's past activities directly conflicts with one's desire for privacy. This theme invites research and development activities designed to solve this dilemma by implementing disclosure policies that maximise the benefit to the user while minimising the amount of information that is being disclosed. Since different individuals are likely to give a different assessment of the benefit received and its costs in terms of privacy, proposals in this theme will have to provide rules that individuals can use to remain always in control of how much information is being disclosed and to verify that the rules do, in fact work as intended. Since individuals may specify several such rules depending on context, the systems proposed under this theme will also have to provide very intuitive methods for managing one's privacy profile.

In this theme as well it will be crucial for proposals to obtain the commitment of communities with credible user groups. A credible user group in this scenario is a group of people who agree to participate in an extended experiment in personal information management and to provide usability and satisfaction feedback. Proposals are also required to have credible plans to track to what extent the personal information management systems developed are adopted outside the confines of the project itself.

These guidelines thus exclude or strongly discourage proposals where the evaluation of the systems developed is carried out with respect to artificial or technology driven tasks and with a small group of users not clearly representative of the individuals for which the systems are built. Also discouraged are proposals for personal information management and visualisation capabilities that already exist or can soon be expected to be provided on many successful websites (e.g. tagging, geotagging, voting, etc…)

Foundational research will focus on the logic of information disclosure and privacy, the psychology of its effective use on the part of naïve users and the economics of the trade-offs between privacy and improved services..

Component and system-level development will provide usable and easy to maintain/extend systems for personal information management and analysis. Benchmarking should address scenarios with a non-trivial number of concurrent users in realistic environments. Usability and portability of such systems is also essential.

2.5 Impact and Scientific and Technological leadership

Networks and other initiatives designed to link technology suppliers, integrators and leading user organisations. These actions will help develop a common understanding, including vis-à-vis neighbouring disciplines, and ensure proactive cross-fertilisation between EU projects and other relevant industrial and national activities. They will address barriers hindering a wider deployment of research results, work towards establishing or advancing widely recognised standards, reference architectures and benchmarks, and increase awareness of the potential of the technologies at stake within broader audiences.

Instruments: CSAs, NoEs

The four previous research themes explore various properties of how humans can harness extremely large data volumes to support effective and well value rich processes. In order to increase the chances of success for this research several kinds of activities are needed:

1. encouraging a dialogue between suppliers and consumers to identify current shortcomings of information management and promising development for the future;
2. encouraging a dialog between developers and content creators to identify a roadmap for an effective uptake of the novel techniques described;
3. fostering community building to encourage multi-disciplinary approaches bringing together scientific communities that could contribute complementary skills to the problems at hand.
4. organising meaningful and rigorous evaluations of technology built by other projects

This fourth research theme will then contain opportunities for both research and community activities.

Research will focus on technology evaluation activities based on large, shared datasets and rigorous experimental protocols.

Support activities will focus on cross-community acquisition of requirements and dissemination of results by means of conferences, technology demonstrations and tutorial workshops and seminars.

3 Types of research

As already mentioned, three broad categories of activity are encouraged: *foundational research; component-level research and development; system-level integration and validation.*

All research and development work should address issues such as modelling of user information behaviours and how to **hide complexity** from the non-expert user. Projects should maximise cross-fertilisation between approaches and disciplines, promote **open architectures** and help build **shared infrastructures** for research and training purposes, as well as for technology assessment and benchmarking.

Work must also promote **consistent stacks of standards** for semantic interoperability between Web services, and for data and process descriptions, bringing closer together multimedia coding, metadata and semantic Web standards and protocols.

Broadly-based ***foundational research*** **is expected mainly under research themes 1, 2 and 3** and will aim at advancing new formal models, methods and languages for representation of knowledge, processes and goals for reasoning about them under uncertainty, to make them more powerful, flexible and durable. Methods and techniques should ensure interoperability or mapping between ontologies and other emerging data models (e.g. spatial or 3D models), as well as catering for semi-automatic creation, extensibility and long-term maintenance of ontologies (the ontology lifecycle).

Foundational research will be measured on the basis of its innovation and degree of uptake by the research community, where the metrics of the scientific peer system will apply. Infrastructure research will be measured in terms of its contribution to enabling (a) new, advanced applications and (b) broader collaborative research endeavours.

Component-level research **is also most relevant to research themes 1,2, 3 and 4** and should address the functionality of information management systems by developing a new generation of component technologies needed to acquire and represent, analyse, annotate, (re-)organise, browse, filter, process and visualise objects and resources of all media types.

A long-term, but critical, objective is to achieve the same level of functionality across different data and media types, improving access to information – for humans and machines alike – and the functionality of content and knowledge management applications.

Components should be task-oriented and should break new ground, for example in the media types and formats they address, and be scaleable in their degree of sophistication. Components must be able to fit into broader reference architectures and/or be easily integrated into diverse application scenarios.

STREPS which develop specific components in support of more ambitious efforts undertaken by IPs (cf. section 6. below) would be welcomed. Likewise IPs which develop components to be integrated into scaleable platforms, demonstrators, working systems are also foreseen. In some cases it will be appropriate to exploit components as open-source shareware packages or as development toolkits available to related projects and the wider research community.

Component-level research will be measured in engineering terms by comparison with the state-of-the art. Tools developed will need to perform significantly better than their predecessors (i.e. faster, more exhaustive, more cost-effective, robust and flexible) and have a high potential for wide-scale integration into operational systems.

System-level integration **is intended mainly for research theme 1, 2, 3 and 4** and will aim to tie together methods and components into innovative (i.e. beyond the state of the art) end-to-end systems and services, with a view to demonstrating their technical feasibility, usefulness and impact in real-life environments.

Research will address performance and effectiveness issues, including ease of integration into the working environment, degree of customisation and impact on information flows and work processes. User acceptability and the need for training of users will also be addressed. Where appropriate, projects will tackle the problem of managing a principled migration from ‘legacy’ methods and techniques.

They will stress collaborative methods for content and knowledge acquisition, management and use, smart portals and other approaches for supporting networked organisations and virtual communities of practice. Target applications will have a strong **multi-sectoral potential** and cater for **leading-edge users**.

Regardless of the application sector, system test-beds and demonstrators should target a **specific user profile** – e.g. the executive, the engineer, the scientist, the analyst or the content/media producer. **Proposers should clearly indicate which profile they address.**

Test-beds will prove the successful integration of component technologies into robust, performant and scalable systems in representative domains (finance, engineering, science...), which are readily transferable to other key sectors. Such test-beds should be:

- ambitious, yet realistic, and should address solutions to challenging problems,
- based on highly representative and potentially transferable sectors,
- feature leading-edge user organisations in the consortium to ensure strong user drive and feedback,
- include an in-depth user requirements phase,
- use real or very representative, realistic data for validation purposes,
- be intuitive, operating seamlessly and transparently, and hiding complexity,
- address a complete end-to-end system or process, or a meaningful subset thereof,
- ensure interoperability with widely accepted industry standards to cater for legacy data and existing popular platforms,
- demonstrate clearly technical feasibility and commercial / operational benefits.

System-level research will need to show outcomes beyond the scientific and technical results, typically in its practical impact on the target user groups and wider sectors of industry and society. What direct effect does it have on users? Is user acceptability and usability higher? Is there significant potential economic or social impact? Ease of integration and customisation, enhanced functionality – leading to improved efficiency and productivity gains? Will such a system demonstrate the benefits – and therefore contribute to widespread acceptance and take-up – of advanced information management approaches?

III. Practical issues

Specific Targeted Research Projects (STRPS) are particularly suited to **high-risk endeavours**, breaking new grounds, with high potential rewards. They are also appropriate for component-level research for particular domains.

Integrated Projects (IP) are the preferred instruments for **system-oriented efforts** and medium-term activities. IPs are expected to encompass all stages of the research and development lifecycle, where appropriate cutting across research themes and addressing system-level integration and validation in realistic scenarios.

Networks of Excellence (NOE) are expected to be used for the longer-term activities e.g. related to new research topics where a **critical mass of research** does not yet exist. NoEs should build communities focusing on longer-term, cross-disciplinary research related to knowledge representation and reasoning, and understanding of non-textual information.

Specific Targeted Research Projects should target **single research problems**, especially the development of truly innovative techniques and new approaches to challenging research problems for the medium to long term. **Consortia of 5-8 partners** are likely to comprise partners mainly from academia and public/private research laboratories. Ultimately, however, the skills mix should be dictated by the scientific and technical challenges being addressed.

STRPS are encouraged to make project results widely available (e.g. as open source or via appropriate licensing arrangements) especially to organisations and consortia undertaking further research in the area. This is intended to ensure the optimum exploitation (in the broadest sense) of project results and ensure a solid basis for future work.

STRPS are also suitable for the development of system components and domain specific tools, possibly in support of a complementary IP or STRP – including running projects arising from previous calls. Such STRPS would typically be driven by industry, including SMEs.

In general, **STRP projects are expected to be in the €2-4M range of EU funding,** and to last up to 3 years. Larger STRPs may be accepted where a large, focused effort would clearly have particularly high impact. The likely impact and exploitation prospects should be very clearly elaborated in the proposal.

Proposals conceived as part of a complementary set of proposals should make clear reference to each other and specify clearly the expected benefits and synergies of co-operation and especially how such a co-operation would be implemented and managed in practice.

Networks of Excellence **are expected to be in the range of €3-4 M of EU funding**, with a duration of up to 3 years. The favoured approach is to drive integration of somewhat scattered research communities by focusing on a number of specific, longer-term research problems – i.e. engaging in **joint integrative research.** This instrument is potentially a very powerful one, but probably also presents the greatest management challenges. Therefore, a **maximum consortium size of 10-12 partners**, of which 5-6 are the core partners driving the initiative, coupled with a **task-based programme of work** is recommended to ensure the necessary focus.

NoEs will generally be driven by academic research communities but will also cater for **industry and user requirements.** Experience so far has shown that industry is reluctant to participate fully in NoEs. However, all the most promising NoEs in this SO to date do have an industrial advisory board. It is recommended that NoE proposers adopt this approach. NoEs will be inter-disciplinary and include facilities for exchanging and testing both human and knowledge resources (e.g. ontologies, tools, reference architectures, evaluation metrics and data).

Integrated Projects are clearly the most appropriate instrument for **large scale demonstrators** and test-beds. IPs should take on board several, if not all the research types mentioned, plus supporting activities if need be. **IPs are expected to be in the range of €4-8M of EU funding** over a 3-4 year period, without ruling out the possibility of more ambitious endeavours. Larger IPs would have to show very clearly that they will make a definite impact in their target area.

In general, consortia are not expected to include more than 8 partners, with up to 12 where justified. Smaller scale (i.e. in terms of consortium, duration or funding) IPs can be appropriate in some cases, provided that they are clearly focused on achieving their objectives within a time window which is dictated by outside circumstances e.g. market opportunities.

IPs will bring together a highly complementary mix of researchers with industrial developers as well as SMEs. Typical consortia will include a central core of 4-6 key partners, who are fully committed to achieving and exploiting project objectives. The core partners would probably be involved in all work packages and would demonstrate their commitment by allocating 2-3 full time equivalents (FTE) per year to the project. Additional partners bringing specific skills and complementary expertise or addressing particular issues (technical, exploitation, socio-economic…) are of course welcome.

IPs will necessarily include leading-edge user organisations for user drive and feedback.

Given that many of the challenges surrounding content and knowledge systems are as much related to technology adoption as pure S&T issues, **IPs should cover much more than RTD tasks** and should address user requirements, test and demonstration, exploitation and dissemination, resulting in a functional 'system'. IPs should be active in training both within and outside the consortium (e.g. training integrators and launching users), organising dedicated training days, seminars and conference tutorials.

IPs are expected to be highly proactive with a professional approach to achieving a high level of **public awareness** for the project from the outset. They should actively promote and promulgate reference architectures built around coherent stacks of standards.

IPs should focus heavily on **impact and exploitation** of project results acting as catalysts for the wider uptake and deployment of the technology. This includes proactively making project results available within and outside the consortium – including via open source or by licensing agreements. As a minimum, an IP should make publicly available the proven, system architecture and the outcome of field trials and user feedback.

Specific Supporting Activities (CAs and SAs) will concentrate on:

- **standards** and their take-up: monitoring and critically assessing existing and emerging standards for digital content assembly, exchange and rendering, for hybrid knowledge representation formalisms and semantic integration, providing guidance on the most appropriate set(s) for the purpose of designing and implementing complex systems,

- **convergence**: monitoring, exploiting and where appropriate driving convergence with other research areas such as networking technologies, where there is scope to develop common approaches, standards and architectures,
- **user aspects**: broad user issues including awareness and user/supplier dialogue, usability guides and best practice, training of leading-edge users, data protection and privacy, etc.,
- **market factors:** identifying and understanding drivers and inhibitors for the uptake and deployment of new technologies, market trends and technology foresight, new business and revenue models, opportunities for global co-operation.
- **dissemination:** organising conferences and workshops, supporting publication

Note, however, that opportunities for SSAs will be limited.

Overview

Research themes	IP	STREP	NOE	SSA	CA
a) Capturing tractable information	Yes	Yes	Yes	No	No
b) Delivering pertinent information	Yes	Yes	No	No	No
c) Collaboration and decision support	Yes	Yes	No	No	No
d) Personal sphere	No	Yes	No	No	No
e) Impact and S&T leadership	No	No	Yes	Yes	Yes

1. Maximising the impact – Synergies with other projects

Proposals are expected to **complement and/or extend the existing portfolio of projects** funded under this SO, especially those funded under the first and second ICT-FP7 calls (see http://cordis.europa.eu/fp7/ict/content-knowledge/home_en.html), as well as in other closely related Strategic Objectives. Duplication of resources or functionality is strongly discouraged and for this reason perspective consortia are strongly advised to familiarise themselves with the scope and goals of currently funded projects.

Proposals which intend to cooperate with existing projects to build-on, complement and support ongoing efforts in a genuine effort to improve the overall impact of the research will be welcomed. Similarly, **clusters of complementary proposals** will be considered. Such proposals should explain clearly their expected contribution to sister proposals and detail the implications of such cooperation both in terms of results, impact and resources as well as well as clearly describing the practicalities of managing the cooperation. An important point is that new projects should avoid rebuilding components that had been previously developed by other projects or that can be easily obtained from open source repositories.

In the case of **multiple participation** there should be no doubt that the organisations have the necessary human, technical and financial resources to tackle all work proposed. In any case, each proposal will be considered on its own merits and there is no guarantee that all proposals in a suggested cluster would be retained for funding.

2. Who should be involved?

The SO constituency is broad and multi-disciplinary. The added value lies precisely in **cross-fertilisation** between a number of different areas, including multimedia and audio-visual technologies, natural language and image processing, experimental psychology, knowledge and Web engineering, agent and database technology,

Software houses, system integrators, user organisations active in knowledge-intensive sectors are expected to be involved in projects aimed at applicative platforms and test-beds. Content creators and aggregators, broadcasters, mobile and broadband operators, equipment manufacturers are likewise expected to participate from the content and media sectors.

Private and public information holdings, scientific databases and audio-visual repositories will be expected to provide rich data sources.

New participants to the programme from academia, research, ICT industry and user organisations are especially welcome. Nevertheless, quality of the proposal (idea, implementation, exploitation etc.) and excellence of the participants will remain the key requirements. Each proposal must illustrate clearly the contribution of each partner in the consortium.

3. Practical information

Which Call for proposals?	Fifth ICT Call.
Which area of the Call?	ICT 2009.4.3 "Intelligent Information Management"
IST 2009-10 Work programme	*ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2009-10_en.pdf*
Application forms and guides	*http://cordis.europa.eu/fp7/participate_en.html*
Proposal evaluation procedure and Model contract	*http://cordis.europa.eu/fp7/participate_en.html* *http://cordis.europa.eu/fp7/calls-grant-agreement_en.html#standard_ga*
Ongoing IST projects	*http://cordis.europa.eu/fp7/ict/content-knowledge/projects_en.html*
INFSO.E2 Web site and Call related documentation	*http://cordis.europa.eu/fp7/ict/content-knowledge/home_en.html*
EC staff	General enquiries and pre-proposals to: *infso-e2@ec.europa.eu*

ANNEX

IST Workprogramme 2009-2010, p. 52-54

Objective ICT-2009.4.3: Intelligent Information Management

Target outcomes

a) **Capturing tractable information**: robust and performant technologies to acquire, analyse and categorise extremely large, rapidly evolving and potentially conflicting and incomplete amounts of information. These technologies will extract, correlate and integrate data from diverse sources and formats (multimedia and 3D content; heterogeneous databases; data streams from sensors and scientific equipment; social interactions and networked appliances; information from business processes and software services) while tracing provenance, evaluating trust level and assessing reliability. The scalability, flexibility and performance of such methods and techniques will be demonstrated by rigorous empirical testing over large-scale testbeds.

b) **Delivering pertinent information**: usable and customisable systems to improve the efficiency of the information lifecycle, starting from proactive diagnoses of information gaps and triggering goal-dependent search, acquisition, structuring and aggregation of relevant local, remote and streaming resources. Managing this information and making it actionable requires large-scale reasoning resulting in effective ranking, profiling and interpretation as well as versioning for time-dependent compliance and justification. Such systems will support the navigation, manipulation and consumption of digital information by means of adaptive user-information interactions based on the state of the art in the psychology of human perception and attention. The effectiveness of such systems will be validated with appropriately-sized groups or communities of representative users.

c) **Collaboration and decision support**: efficient and dependable problem solving and decision support systems for critical, information-bound domains in which our ability to share and exploit information is outstripped by the rate of its growth in size and complexity. Intended beneficiaries include organisations with complex business processes and access control policies; scientific communities collaborating on challenging projects and building very large datasets; teams of professional creators working on complex designs or multimedia materials; and web communities with sophisticated cooperation needs. The effectiveness of such solutions will be tested against the requirements of the respective groups or communities.

d) **Personal sphere**: intuitive systems that help individuals secure, manage, visualise and interpret their personal information, attention trail and social history so as to enable the provision of personalised and context-dependent information from multiple sources and services. A specific requirement and design principle is that such systems preserve privacy and implement auditable information disclosure policies that are under user control and whose application can be verified at all times. Their usability and rate of uptake will be monitored by means of verifiable quantitative indicators.

e) **Impact and S&T leadership:** networks and other initiatives designed to link technology suppliers, integrators and leading user organisations. These actions will help develop a

common understanding, including vis-à-vis neighbouring disciplines, and ensure proactive cross-fertilisation between EU projects and other relevant industrial and national activities. They will address barriers hindering a wider deployment of research results, work towards establishing or advancing widely recognised standards, reference architectures and benchmarks, and increase awareness of the potential of the technologies at stake within broader audiences.

Expected impact

- Better leveraging of human skills, improved quality and quantity of output and reduced time and cost allowing users to concentrate on more creative and innovative activities.

- Increased ability to identify and respond appropriately to evolving conditions (e.g. in finance, epidemiology, environmental crises …) faster and more effectively. Reinforced ability to collaboratively evolve large-scale, multi-dimensional models from the integration of independently developed datasets.

- Higher levels of information portability and reuse by creating an ecology of systems and services that are dynamic, interoperable, trustworthy and accountable by design.

- Increased EU competitiveness in the global knowledge economy by fostering standards-based integration and exploitation of information resources and services across domains and organisational boundaries.

- Strengthened EU leadership at every step of the computer-aided information and knowledge management lifecycle, creating the conditions for the rapid deployment of innovative products and applications based on high quality content.

Funding schemes

a): NoE, IP, STREP; b), c): IP, STREP; d): STREP; e): CSA, NoE

Indicative budget distribution

- IP/STREP: EUR 62 million with a minimum of 50% to IPs and a minimum of 30% to STREPs

- NoE/CSA: EUR 8 million

Calls

ICT call 5