



**European Cooperation
in the field of Scientific
and Technical Research
- COST -**

Brussels, 21 November 2012

TD1210

MEMORANDUM OF UNDERSTANDING

Subject : Memorandum of Understanding for the implementation of a European Concerted Research Action designated as COST Action TD1210: Analyzing the dynamics of information and knowledge landscapes - KNOWeSCAPE

Delegations will find attached the Memorandum of Understanding for COST Action as approved by the COST Committee of Senior Officials (CSO) at its 186th meeting on 20 - 21 November 2012.

MEMORANDUM OF UNDERSTANDING
For the implementation of a European Concerted Research Action designated as

COST Action TD1210
ANALYZING THE DYNAMICS OF INFORMATION AND KNOWLEDGE LANDSCAPES
- KNOWESCAPE

The Parties to this Memorandum of Understanding, declaring their common intention to participate in the concerted Action referred to above and described in the technical Annex to the Memorandum, have reached the following understanding:

1. The Action will be carried out in accordance with the provisions of document COST 4154/11 “Rules and Procedures for Implementing COST Actions”, or in any new document amending or replacing it, the contents of which the Parties are fully aware of.
2. The main objective of the Action is to develop interactive knowledge maps for the study and curation of large information spaces and effective navigation through them.
3. The economic dimension of the activities carried out under the Action has been estimated, on the basis of information available during the planning of the Action, at EUR 80 million in 2012 prices.
4. The Memorandum of Understanding will take effect on being accepted by at least five Parties.
5. The Memorandum of Understanding will remain in force for a period of 4 years, calculated from the date of the first meeting of the Management Committee, unless the duration of the Action is modified according to the provisions of Chapter V of the document referred to in Point 1 above.

A. ABSTRACT AND KEYWORDS

There is no escape from the expansion of information, so that structuring and locating meaningful knowledge becomes ever more difficult. This Action will tackle this urgent problem by using the unique networking and capacity-building features provided by the COST framework. For the first time, a platform will be created where information professionals, sociologists, physicists, digital humanities scholars and computer scientists collaborate on problems of data mining and data curation in collections. The main objective of this Action is advancing the analysis of large knowledge spaces and systems that organize and order them. The combination of insights from complexity theory and knowledge organization will improve our understanding of the collective, self-organized nature of human knowledge production and will support the development of new principles and methods of data representation, processing, and archiving. To this end, the knowledge organization in web-based information spaces such as Wikipedia as well as collections from libraries, archives, and museums will be studied. This Action aims to create interactive knowledge maps. Their end users could be scientists working between disciplines and seeking mutual understanding; science policy makers designing funding frameworks; cultural heritage institutions aiming at better access to their collections; and students seeking a first orientation in academia.

A.2 Keywords: Evolving knowledge orders in information spaces, complex networks, information seeking behaviour and retrieval, big data mining and curating, open linked data

B. BACKGROUND

B.1 General background

The question of how to order our knowledge is as old as the systematic acquisition, circulation, and storage of knowledge. Classification systems have been known since ancient times. On the Internet, one finds both classifications and taxonomies designed by information professionals and folksonomies based on social tagging. Nevertheless, a user navigating through large information spaces is still confronted with a text based search interface and a list of hits as outcome. There is still an obvious gap between a physical encounter with, for example, a library's collection and browsing its content through an on-line catalogue. This Action is concerned with the development of new interfaces – *knowledge maps* – to large knowledge spaces based on insights about the composition and evolution of those spaces. The expression “knowledge space” is understood here in

a broad sense encompassing collections of artefacts (and their description) as studied in the humanities; collections of libraries and archives; web-based information spaces such as Wikipedia and collaborative platforms (e.g. Mendeley, CiteUlike), but also bibliographic databases (e.g., Web of Knowledge) and national research information system (e.g. NARCIS – the Dutch National Academic Research and Collaborations Information System). A knowledge map could take different forms: a simple tree map, a sophisticated scientific visualization or an artistic visual representation. In any case, preferably a user would get visual feedback (i) about the size and composition of the collection she or he interrogates and (ii) about where her or his search “lands” on such a map, and so could further adapt the search. This Action takes the problems of knowledge organization back to its roots applying a “non-linear physics” (or complexity) inspired perspective. For the scientific foundation of knowledge maps this Action will combine well known techniques of classification and indexing - professionalized in libraries, archives, museums and specific collections – with knowledge representation techniques from the computer sciences, and methods of complex pattern recognition from statistical physics.

Within Europe many research groups focus on problems of knowledge organization, representation, management and data curation. However, each group tends to focus on special topics. Despite the similarity of their tasks, these communities do not share information with each other on a regular basis. The time is ripe for COST to facilitate a new network that spans these disparate communities. At this stage, the development of a “common research language”, exploration of methodologies and identification of common research questions are vital to help future collaborative research projects. Without question, it is a high risk project with pioneering characteristics that meets the criteria surrounding COST Actions. This Action will thus pave the way to a new, comprehensive alliance in the area of massive data curation and large scale information processing. Five different communities will be called in to solve the puzzle of navigating effectively through and across different large scale knowledge spaces: **Information professionals** (from libraries and archives but also from information sciences), **Sociologists of knowledge**, Scholars in **Digital Humanities** fields, **Physicists**, and **Computer scientists**. Achievements of the Action will foster scientific progress through newly emerging synergies based on shared concrete research topics.

B.2 Current state of knowledge

From libraries to the web: A vast amount of recorded knowledge (public, corporate and private) has now migrated to the Internet. Sophisticated textual retrieval methods are available, but not readily applicable to all information searching scenarios. Research in information seeking behaviour

has shown furthermore that users need assistance in choosing the correct searching expression and in determining the exact context in which to look for information. Subject metadata, especially language-independent indexing systems such as classifications have been often implemented as pivots in cross-collection, cross-domain and cross-language information searching. Initiatives as The European Library and Europeana (formerly Council of Content Providers and Aggregators) aim for integration and easy access to Europe's cultural and scientific heritage. Semantic Web activities with SKOS (Simple Knowledge Organization System standard) and Linked Data initiatives (e.g. W3C Library Linked Data Incubator Group) and research on Web 2.0 technologies especially in social tagging (e.g. the 'Enhanced Tagging for Discovery' project, funded by JISC -Joint Information Systems Committee in UK) contribute to the efforts to create a global web of knowledge. Several questions remain, however, what is the best way of presenting and visualizing a knowledge space and which kind of interfaces will enable the full use of a more complex knowledge presentation. **From science maps to interactive knowledge maps:** For scientific information, scientometrics – the quantitative study of science – has applied algorithms and tools to render and display large scale networks of scientific publications. Science maps are used to display the position of groups or institutions in the scientific universe, and the portfolio of funding programmes. Research policy and evaluation are natural areas of application. Models of science dynamics from sociology, mathematics and physics contribute to a better understanding of these maps. There are already some attempts towards information retrieval systems based on models of science dynamics. Research projects such as VIVO (an open source semantic web platform, funded by a National Institutes of Health grant, in USA) develop web ontologies for all parts of the science system (institutions, persons, publications, projects). However, science maps are mainly based on citation databases. Their extension to other information resources as well as their use in interfaces to collections remains to be explored. **From fundamental research to infrastructures:** Physicists, working on complex networks, have developed alternative approaches to knowledge organization by extracting patterns from emerging networks of digitized information. But connections to traditional knowledge orders are rarely discussed, which also hampers their diffusion into information retrieval. The same holds for insights from the sociology of science about human behaviour in information seeking and processing.

In contrast to earlier attempts, this Action will provide a platform for connecting and linking primary information sources of three kinds (1) so-called legacy collections, i.e. recorded information as preserved in cultural heritage institutions, (2) knowledge as it emerges within scientific projects (research data, publications, research information) and (3) knowledge spaces created in the collaborative Web-based communities. While there are many projects, networks and

consortia addressing the problem of information and knowledge management in specific communities, this Action will create a cross-community platform, coordinating and focusing research around the topics “classification” (knowledge orders) and “knowledge maps” (navigating through large knowledge spaces).

B.3 Reasons for the Action

This Action will contribute to **scientific advance** in the information sciences with a systematic study of the evolution of knowledge spaces and ordering systems. It will bring new challenges to complexity science by the investigation of dynamic principles behind the emergence of knowledge orders, a highly non-linear problem. It will contribute to a scientific foundation for data curation by comparative research at a general level about organization principles applied to systems as different as large library collections, Wikipedia, or Research information systems. The main **objective** of this Action is the development of knowledge maps for the study and curation of large information spaces and an effective navigation through them. This Action responds to **societal needs** for an effective curation and organization of vast amounts of data. Much emphasis is given by EU bodies to the evaluation of research. Understanding landscapes of knowledge can add value to common practices and evaluation tools, so responding to the growing need of the detection of innovative sectors, and providing tools that may result in fostering European competitiveness. COST represents the most suitable framework due to the interdisciplinary nature of these issues and the need to put together and coordinate the required competencies. **Expected results** are (i) cross-domain publications accessible beyond specialized audiences; (ii) demonstrators for knowledge maps; and (iii) guidelines on how to implement them. This will be achieved by the following **means** (i) Coordinating the expertise from several countries, establishing formal collaboration between the partners due to workshops and short term visits; (ii) Providing means for intensifying such collaboration and creating and extending an international open network under the objectives of the Action by inviting other established experts to join the Action and, in particular, by capacity building and education of Early Stage Researchers (ESRs) through meetings and training schools; (iii) Enabling diffusion of methods and tools in hands-on workshops; (iv) Organizing outreach events with information professionals, and other user groups.

How the research topic includes two or more COST Domains? The complex character of information spaces produced by humans cannot be properly understood without analysing the mechanisms of their generation. Therefore, this Action is rooted in complexity science, computational sociology, and socio-physics (ISCH, MPNS). It can also not be studied without

mechanisms of information retrieval and the underlying information and computer technology (ICT). New tools for navigation will be developed taking into account the dynamic, flexible, and distributed character of today's information and knowledge creation (MPNS, ISCH). These tools are based on recent advances in the computer sciences and information sciences (ICT, ISCH).

How the research requires the reciprocal interaction between these Domains? Traces of human knowledge production, preserved by information specialists (librarians, archivists, scholars in humanities fields, museum curators, contributors of Web spaces) provide the empirical base of this Action. The methodological basis for pattern recognition comes from complexity and network science. The toolbox of statistical physics concepts and methods forms the technical core. New insights resulting from evolutionary, dynamic approach to knowledge systems will be combined with new methods to index, order, and retrieve digital information from large knowledge spaces. In the analysis of knowledge ordering systems and data from large collections, mathematical models of collective knowledge production as developed in computational sociology, computational philosophy and socio-physics will be used. New navigation tools will be developed in collaboration with knowledge-domain specialists in the social sciences and humanities. Semantic web approaches will be used to enhance access to collections and to also link between their different knowledge organizations.

Why within the scientific approach, the reference to common theoretical concepts and methods as well as to their common evolution is needed for all involved Domains? This Action will be founded on the paradigm of self-organization– a concept that has penetrated philosophy, sociology as well as computer science and physics since the 1950s. While present in those disciplines, this paradigm has not been applied to the problem of knowledge creation in a shared, mutual effort. This Action will focus on the co-evolution of objects that knowledge is inscribed in (such as books, articles, art objects, images) and the systems in which these are indexed, ordered, and described. The emphasis is on understanding the principles behind the creation of systems of knowledge organization and their impact on knowledge creation. Specific to this Action is the application of mathematical, dynamic and non-linear models, simulation, animation and visualizations techniques.

B.4 Complementarity with other research programmes

There are several European research projects and Networks which have been or still are directly related to this Action, for example the COST Action MP0801 Physics of Competition and Conflicts; LarKC (Large Knowledge Collider), LATC (Linked Open Data Around the Clock),

CyberMotions, FuturICT (an ICT flagship proposal), ACUMEN (Academic Careers Understood through Measurement and Norms – FP7 project about scientific careers), QLectives (Quality Collectives – FP7 FET project building cooperative socially intelligent ICT systems) , OpenAIREPlus (2nd Generation of Open Access Infrastructure for Research in Europe – FP7), MULTIPLEX (Foundational Research on MULTilevel comPLEX networks and systems)(FP7 project) and EINS (NoE Internet Science). Several members of these are also intended members of the KNOWeSCAPE Action. Therefore, it will be possible to have organic links with these domain specific projects. Dissemination activities will profit from the membership of key members in Research Infrastructure Networks for the Social Sciences and Humanities, such as DARIAH (Digital Research Infrastructure for the Arts and Humanities) and CLARIN (Common Language Resources and Technology Infrastructure). Links are in place with national and international programmes in Digital Humanities and with the Open Annotation Collaboration Consortium (USA). Further links are intended with the Europeana Foundation, National Libraries and Museums.

C. OBJECTIVES AND BENEFITS

C.1 Aim

The aim of the Action is the development of interactive knowledge maps for the study and curation of large information spaces and effective navigation through them.

C.2 Objectives

Scientific Objectives

O1: Compare and analyse large information spaces and their knowledge-ordering or classification systems; find generic dynamic principles for their emergence; and important dimensions to be displayed – **Analysis and models of knowledge spaces**

O2: Design knowledge maps and new Information Retrieval models enabling cross-system, cross-collection and cross-domain searches in large heterogeneous information spaces – **Maps as navigation tools through knowledge spaces**

O3: Develop knowledge maps demonstrators and design ways for their implementation using new data representations (Open Linked Data) – **Guidelines for implementing knowledge maps**

Measurable Objectives

(i) Establish a series of cross-domain, small scale workshops to identify similar problems, developing a common language, and coordinate so far uncorrelated parallel efforts with similar

goals leading to an increase in efficiency in the use of current funding (measure: domain diversity of participants, cross-domain references in publications in peer reviewed journals) (ii) Increase the collaborative work between the partners of the Action through short term research visits (measure: domain diversity between receiving and sending institution, joint publications by Action members) (iii) Mutual co-organisation of satellite meetings at key conference series (measure: domain diversity of venues) (iv) Collaboration with information professionals at cultural heritage institutions (library, archives, museums, institutes with specific collections) (measure: visits, hearings, participation at Action workshops) (v) Development of demonstrators such as visual interfaces to collections, and implementation of new principles of information retrieval (measure: number of interactive maps and guidelines)

C.3 How networking within the Action will yield the objectives?

This Action aims at the coordination of research about large information and knowledge spaces as well as knowledge exchange in the field of information science at the European scale. To attain this goal, this Action will:

- (i) create an effectively coordinated scientific forum for experts from library, information and documentation sciences, and researchers in fields such as archaeology, history, musicology, literature and art history; computer sciences (semantic web); and physics
- (ii) promote and encourage collaboration between young researchers in those fields
- (iii) cultivate a coherent framework for future research in the new area of Knowledge Dynamics

C.4 Potential impact of the Action

The “information flood” can be seen as a threat or as a reservoir for knowledge production. This Action will use methods of mathematical modelling (simulation, agent-based models) to explore scenarios to find knowledge in an effective way. **Scientific impact and benefits for the networking partners:** New concepts and methods will be developed. Concepts and methods will be disseminated across domains. Enhancement of research will emerge through synergy effects. The innovativeness of domain-specific and cross-specific research will increase due to the diffusion of ideas and methods. The Action will facilitate efficient sharing and development of expertise and effort that have otherwise been rather fragmented. **Societal impact:** Creation of knowledge maps as

possible new interfaces to collections, as basis for information retrieval, and as support to data curation enhances access to and permanence of digital content at scientific and cultural heritage institutions. This way, the Action will address societal needs for information and data management. Insights into mechanisms of spontaneously emerging linked information spaces and new practices of knowledge creation will inform the current debate about appropriate infrastructures; open and permanent access to information; linked data and the semantic web. For information professionals this Action provides links to fundamental research. The exploration of new principles contributes in the long-run to better services for the knowledge treasures under their curation. Early involvement of future applicants accelerates knowledge diffusion from fundamental to applied research to public use. **Benefits for early stage researchers:** In a scientific age labelled by words such as e-science, web science, team science or big-data-science, the ideal portfolio for future researchers combines in-depth knowledge in specific domains with the ability to orient globally in a rapidly growing complex and scattered science landscape. It requires the combination of mathematical and computational skills (or at least basic understanding of them) with other research methods and approaches. It also requires excellent communication skills, the ability to identify right partners to fill gaps in competences, and personal skills to understand and moderate scientific controversies at heart of innovative processes. This Action will provide a scientific environment where understanding of other scientific cultures and shared seeking for solutions across epistemic boundaries are natural ingredients. It fosters the growth of the above-mentioned skills in a safe, protected environment, where learning through failures is possible. This Action will ensure that ESRs of both genders are adequately trained in cross-domain problem solving, while at the same time providing excellent domain specific knowledge from the strong science base and science management experience of its more senior experts.

C.5 Target groups/end users

Knowledge maps are of high societal and scientific relevance. They are needed to allow students to find the right entry point in the wealth of academic knowledge. They can inform interdisciplinary discourse between experts by presenting different epistemic traditions. They are useful for science policy to determine measures and target for investment. **Information provider/professionals** form one target group. Librarians, documentalists at archives, and researchers involved in research and development in cultural heritage institutions have triggered the formation of this Action and actively contributed to its preparation. For them the research of this Action will support better

access, re-use and better curation of their data/information/knowledge objects. Eventually, this will be of profit for the **general public** using these services. In **Academia**, Early Stage Researchers and Experts from this and other COST Actions are another target group. The results of this Action will enable them to continue further research in a yet to be established area, but also in their capacity as first test-users of knowledge maps. **Science policy makers** form a further user group. Knowledge maps can inform experts prior to engaging in new alliances; they can also inform science policy bodies about the existing knowledge and science landscape.

D. SCIENTIFIC PROGRAMME

D.1 Scientific focus

This Action will carry out the following five tasks to achieve the objectives described in Section C:

O1: Analysis and models of knowledge spaces, O2: Maps as navigation tools through knowledge spaces, and O3: Guidelines for implementing knowledge maps.

Task 1: Selection of knowledge spaces to be mapped

The notion of a “knowledge space” is fairly abstract and understood here in a broad sense. Empirically, this Action relies on collections of artefacts (and their description) as studied in the humanities; library, archival and museum collections and national bibliographies; but also web-based information spaces such as Wikipedia, social networks (e.g. Facebook), and collaborative platforms (e.g. Mendeley, CiteUlike). The Action will also examine the scientific citation space in commercial abstracting and indexing bibliographic databases such as Current Contents, ERIC, Medline, SPIN, CAS etc. available through service providers such as ISI Web of Science, Scopus, PubMed. The knowledge space of the national research information systems, and databases of funding institutions will also be studied. The first task to be coordinated by the Action is to make a representative selection from the large variety of knowledge spaces and to coordinate studies across different research groups. One important issue to address in Task 1 is the overlap in research interest when it comes to different information systems and their relation to a particular part of the science system. The main difficulty in this task will be to identify a selection of knowledge spaces that will best serve the aim in designing generic tools and methods that can be applied in collections of different scales and contents.

Task 2: Data representations to support data mining across knowledge spaces

The fact that increasingly knowledge spaces are available on-line does not automatically imply that information about them can be easily harvested. Classification systems are usually designed for the

information needs of specific communities of users. How to map them in order to create links between different sources of knowledge has been addressed extensively in vocabulary mapping projects such as Multilingual Access to Subjects (MACS), Renardus, Semantic Interoperability to Access Cultural Heritage (STICHT), Multilingual Subject Access to Catalogues (MSAC) and terminology services projects such as OCLC Metadata Switch and High Level Thesaurus Project (HILT). Recently, the W3C linked data activities within the Semantic Web initiative (a continuation of the SKOS standard development), and promotes publishing and sharing of knowledge organization systems on the web as well as their mapping. While under pressure to organize the ever-increasing amount of digital information there is a danger overlooking data, information, and knowledge that have now become part of our scientific and cultural heritage. Combining traditional and recent ways of knowledge organization this Action will address the question how to define an appropriate generic level on which knowledge organization systems can be linked.

Task 3: Identify dimensions of knowledge maps – modelling and analysing knowledge spaces

To design a knowledge map requires the identification of characteristic attributes and forms of representation. The disciplinary composition of a collection can be represented by treemaps for instance. Geographic spreading can be displayed on georeferences. But, finding alternative reference systems (similar to a parameter or variable space) with dimensions useful to be represented on a map requires pattern extraction based on similarities of knowledge objects (e.g., book, article, author, text piece, institution) as expressed on their description (text), authorship, form and media, age of production and other metadata. Usual top-down approaches in knowledge organization need to be combined with bottom-up approaches to mining relational information among knowledge objects. This Action takes the problems of knowledge organization back to its roots applying a “non-linear physics” (or complexity) inspired perspective. Collections are seen as large-scale systems composed from knowledge objects. Knowledge ordering systems are seen as attributes or coordinates. It seeks universal solutions on a generic level.

Analysis and non-linear models from physics and computational sociology are combined to (i) identify dimension of knowledge spaces in simulations; (ii) extract promising knowledge orders for mapping from real data analysis; and (iii) test the use of knowledge maps in human information navigation. This way the current *ad hoc* diffusion of concepts of complex systems, self-organization, and non-linear dynamics for the analysis of information processes is put on a systematic and reliable knowledge base.

Task 4: A typology of knowledge maps

Ideally, for one collection different knowledge maps would be available between which a user would be able to choose. Another feature could be knowledge maps which adapt to the search

behaviour of the user. To be able to implement such features, one first needs to have a variety of different representations to choose from. In the case of large amounts of information, algorithms have to be developed that allow the display of information. This Action will produce a catalogue of different knowledge maps including requirements for input data format, computational requirements and tools, and possible uses based on insights and experiences of scientific visualization specialists.

Task 5: Implementation of knowledge maps

This Action will develop methods to display and analyse composition, structure, and evolution of knowledge spaces and to identify mechanisms ruling its evolution. The Action will support the application of existing tools such as Gephi, and Sci2 (both are visualization tools for networks), but also the diffusion of new visualization tools. The understanding of the connection between legacy knowledge systems and the current and evolving knowledge space is also likely to improve cross-system, cross-collections and cross-domain navigation. It would go beyond the capacity of this Action to already implement knowledge maps in living on-line interfaces. To integrate new interfaces into living services is a complicated, multi-step process in itself. This Action aims to develop ‘proof of concept’ demonstrators. By documenting the research process and by communicating closely with information professionals and other stakeholders this Action will develop workflows and guidelines for future implementation of knowledge maps.

D.2 Scientific work plan methods and means

To achieve the scientific objectives and to realize the tasks outlined above, the Action will be divided into four Working Groups (WGs). All WGs take responsibility for the whole research process, but do have one main scientific objective, which is further specified.

WG 1: Phenomenology of knowledge spaces

Main scientific objective: New data mining and pattern recognition techniques for large scale information spaces and knowledge orders applied to them (such as collections of libraries, archives and web spaces as Wikipedia, research information systems, bibliographic databases) (O1,2)

Specific objectives, means and methods

- Analyse the co-evolution of knowledge order systems and the collections where they are applied
- Explore models of knowledge revision to map different knowledge order systems (ontology, thesaurus mapping)
- Apply methods of community detection to collaborative knowledge and information spaces

- Compare historical and current socio-economic databases (e.g., census data)
- Analyse knowledge spaces of the current academic systems (libraries, repositories, archives, research information systems)
- Perform comparative temporal analysis of the evolution of recommendation systems in academia, economy and art (e.g. CiteULike, DeviantArt)
- Analyse the structure of the emerging semantic web through time (RDF graphs)
- Identify main characteristics of new forms of scientific communication (including data reviews, micro-publications etc.)
- Analyse the behavioural patterns that determine the evolution of the knowledge order system, using edit logs, revision histories, and discussion forums associated to the structured content.
- Investigate the relation between behavioural patterns of content generation and the specific nature of the produced content.

WG 2: Theory of knowledge spaces

Main scientific objective: Non-linear, dynamic models of the cognitive, discursive and social mechanisms for the emergence of knowledge spaces and knowledge orders (O1,2)

Specific objectives, means and methods

- Develop models of trust and reputation as the basis of recommender systems for knowledge spaces
- Develop and simulate models of the cognitive and social dynamics in collaborative web platforms (e.g., Wikipedia, Mendeley)
- Apply stochastic models of scientific organizations and institutions
- Develop models of argumentation, controversies, competence recombination around scientific innovation
- Identify dimensions for the representation of structure and evolution of knowledge maps

WG 3: Visual analytics of knowledge spaces – knowledge maps

Main scientific objective: Visualization principles for knowledge maps (O2,3)

Specific objectives, means and methods

- Analyse and visualize complex universal knowledge classification systems of specific knowledge domains (e.g. bibliographic classifications such as the Universal Decimal Classification and subject category systems in databases) and large scale knowledge spaces
- Map the principal dimensions of information spaces as well as their occupation by a collection of objects
- Explore clustering-based visualizations
- Integrate different science and knowledge maps

- Outline the best practices of knowledge mapping
- Create a typology of knowledge maps

WG 4: Data curation and navigation based on knowledge maps

Main scientific objective: Guidelines to implement new navigation and retrieval strategies with the aim to increase data access, analysis and re-use (O1,3)

Specific objectives, means and methods

- Design workflows to integrate large knowledge spaces using Open Linked Data principles
- Develop principles for a useful selection of and/or possible combination between different ordering systems
- Analyse the influence of the underlying technological infrastructure of digital collections on sustainable access
- Identify standards of data representation useful for visual analytics
- Explore the use of visual analytics for collection policies and data curation
- Develop interactive interfaces for visually enhanced collection browsing
- Create new Information Retrieval models for cross-system, cross-collection and cross-domain searches in large heterogeneous information spaces
- Design demonstrator cases
- Provide guidelines for implementation

Provisions for the inclusion of new participants and unforeseen activities: The listed tasks and specific objectives of the Working Groups are flexible and liable to alteration should new participants and unforeseen activities become available during the implementation of the Action. They provide a reservoir from which the Action will make a selection. In order to achieve its objectives the Action will not only set key research areas and establish collaboration between the research groups of the consortium to coordinate research, but it will continuously seek new academic partners and stakeholders to join the consortium.

E. ORGANISATION

E.1 Coordination and organisation

The organisation of this Action will conform to the "Rules and Procedures for implementing COST Actions". The Action will be coordinated by a Management Committee (MC). A Steering Group (StG) is elected at the kick-off and first MC meeting consisting of the Action Chair, Vice Chair and Scientific Coordinators for the WGs. The StG is supported by two Specific Tasks Coordinators: a

Dissemination Coordinator (responsible for website(s); monitoring publication strategy) and an Outreach Coordinator (monitors the outreach and implementation policy). The StG proposes the MC a Gender mainstreaming policy (monitored by the Vice Chair) and a policy for Early-Stage Researcher (monitored together with the STMS policy by the Chair). Functions of Scientific Coordinators of WG's and Specific Task Coordinators can be combined. The Annual Conference (including the annual MC meeting) permits sharing experiences, assess key advances and fosters scientific directions. This Action organizes an international meeting at the end of its duration. Connections to international experts from USA and non-COST countries in Eastern Europe are in place.

The Steering Group (supported by Specific Tasks Coordinators) manages the day-to-day operations of the Action, monitors milestones and prepares documents for the annual MC meetings, including a financial plan to adequately share the resources within the Action budget. The Annual MC meetings take place at workshops or conferences organised by the Action. The MC ensures strong interactions between WGs to ensure cross-fertilisation and stimulate cooperation with related COST Actions and networks. An outreach policy will be developed by the StG/MC. **Milestones** are the MC meetings (annual conference), WG meetings (WG specific workshops, one per WG and year), Satellite Events to Major International Conferences, Hands-On Workshops (Outreach Events), and STSMs. Their aim is support newly emerging collaborations, to facilitate their presentation and discussion, and to ensure significant results are published in international peer reviewed journals. Presentations on potential funding opportunities will be used to actively promote the formation of consortia to bid for such funding. Guidelines and workflows concerning how to implement interactive knowledge maps at cultural heritage institutions are a central aim. This will be ensured by close collaboration with and outreach events at cultural heritage institutions. The maintenance of a wiki-like structure for knowledge exchange at a web site; and organization of hands-on workshops are means of dissemination.

E.2 Working Groups

This Action will achieve its scientific objectives via specific objectives, means and methods as identified within four Working Groups. These are: **WG1 Phenomenology of knowledge spaces, WG2 Theory of knowledge spaces, WG3 Visual analytics of knowledge spaces – knowledge maps, WG4 Data curation and navigation based on knowledge maps**. Each WG will have at least one Scientific Coordinator, who will report to the MC and support the daily

monitoring and annual reporting of the Action. If appropriate a Deputy Scientific Coordinator will also be installed. In that case Scientific Coordinator and its deputy should come from different communities involved in this Action. Working Groups develop the plans to address the research tasks (D1) and specific objectives (D2). They present their plans to the MC and seek synergy with other WG's. Synergy can be facilitated by mutual organization of meetings; mutual preparation of future proposals; or organization of (hands-on) workshops. Concerning gender balance and Early Stage Researchers, STSM policy, dissemination and outreach WG Scientific Coordinators inform and consult the corresponding Coordinators as well as the Chair and Vice Chair.

E.3 Liaison and interaction with other research programmes

This Action is located at an interface of fundamental research and the building of research infrastructures. Members of this Action are also members of large scale European Research Infrastructures such as DARIAH (Digital Research Infrastructure for the Arts and Humanities), CLARIN (Common Language Resources and Technology Infrastructure), HERANET (Humanities in the European Research Area), CESSDA (Council of European Social Science Data Archives), DASISH (Data Service Infrastructure for the Social Sciences and Humanities), Holocaust Research Infrastructure (EHRI) and the ESF network NeDiMAH (Network for Digital Methods in the Arts and Humanities). These links will be used to disseminate results of the Action, to consult experts outside the Action, to collect requirements from practitioners and to seek demonstration cases for interactive knowledge maps for science and cultural heritage collections. Existing contacts to the Data Documentation Initiative, The Alliance for Permanent Access and the W3C consortium will be used in relation to the standards and workflows for autoingest/autoexchange of data that will be used. Relevant existing contacts to large scale research information systems include EUROCRIS (European current research information systems); the International Society for Scientometrics, Informetrics and Bibliometrics; The International Association for Social Science Information Services and Technology, and VIVOweb. The Action will present results at satellite meetings, workshops and sessions at major conferences of all participating domains, including at least one appearance at "The European Conference for Complex Systems", "Theory and Practice of Digital Libraries", "Digital Humanities Conference", and "International Semantic Web Conference". This Action will explore overlap with other COST Actions related to Cultural Heritage (TD0902, TD1201) and Information Technologies (IC1203, IS1205, IS1004, IS0904, IC0903, IC0901) and similar projects, such as CENDARI - Collaborative European Digital Archive.

E.4 Gender balance and involvement of early-stage researchers

This COST Action will respect an appropriate gender balance in all its activities and the Management Committee will place this as a standard item on all its MC agendas. The Action will also be committed to considerably involve early-stage researchers. This item will also be placed as a standard item on all MC agendas.

The initially implemented policies will be reviewed annually and if needed adapted. As policy concerning gender balance, the Action aims for a quota corresponding to the share of female researchers in the MC as default. This quota concerns coordinating positions in the Action, participation at workshops and in STSMs. Deviations from this will need to be justified in a specific and evidence-based manner. This policy will be monitored by the Vice Chair. Affirmative action will be taken to enable young researchers of both genders attend network activities of the Action. STSMs will be an important means towards capacity building for ESRs. 50% of STSM's will be dedicated to ESRs. When STSMs are used by non-ESRs, the use of such a mission for ESRs at the receiving institution will be an additional evaluation criterion. This way the Action aims to ensure the flow of knowledge across different generations of scientists. 30% of the STSMs should be related to contacts with stakeholders such as libraries, archives, national networks of information professionals. The STSM policy of the Action will also aim to achieve a balance between involved disciplines and countries, and will be monitored by the Chair.

F. TIMETABLE

Y	Milestones	Deliverables
1	Kick off meeting and WG meetings (Establish work plan and budget plan); <i>Website in function</i> (M6); WG workshops; Satellite meeting; Outreach event; STSM's	STSM/WG reports; <i>First draft of wish list of stakeholders</i> ; Annual Report; Joint Publications (JP); Bibliography

2	WG workshops/STSMs/Outreach event; Satellite Meeting; <i>Training school or Hands-on workshop</i> ; Annual Meeting (MC incl. WG's) joint with International Conference	STSM/WG reports; <i>Knowledge maps (demos)</i> ; Annual Report; JP; Bibliography (updated)
3	WG workshops/STSMs/Outreach event; Satellite Meeting; Annual Meeting (MC incl. WG's) joint with Int. Conf.	STSM/WG reports; <i>Knowledge maps and interfaces as demos</i> ; Annual Report; JP; Bibliography (updated)
4	WG workshops/STSMs; Sat. Meeting; <i>Stakeholder meeting with Information Professionals</i> ; Closing Annual Meeting (MC incl. WG's) joint with Int. Conf.	STSM/WG reports; <i>Guidelines for knowledge maps and their implementation</i> ; Annual + Evaluation report; JP; Bibliography (finalized); <i>Knowledge map of the Action</i>

G. ECONOMIC DIMENSION

The following 20 COST countries have actively participated in the preparation of the Action or otherwise indicated their interest: AT, BE, BG, DE, EL, ES, FI, FR, HR, IE, IL, IT, LU, NL, PL, RS, SE, SI, TR, UK. On the basis of national estimates, the economic dimension of the activities to be carried out under the Action has been estimated at 80 Million € for the total duration of the Action. This estimate is valid under the assumption that all the countries mentioned above but no other countries will participate in the Action. Any departure from this will change the total cost accordingly.

H. DISSEMINATION PLAN

H.1 Who?

This Action aims for new tools for information navigation and retrieval that will enhance the societal relevance of collections provided by libraries, archives and cultural heritage institutions.

Four broad interest groups will be targeted for dissemination: (1) **Information providers**; practitioners and professionals at libraries, archives and other collections (traditional and on-line), (2) **Academia**; research institutes; Higher Education bodies; Early-Stage Researchers; and other researchers working in the field including other COST Actions and potential graduate students; and other networks such as: standardisation bodies for scientific information, W3C consortium, Open Linked Data, and Open access initiatives (see also networks in E3), (3) **Policy makers** and funding agencies at European; national and regional level designing policies for research and research infrastructures, (4) the **General public** using the information services of cultural and scientific heritage institutions.

H.2 What?

Information providers: inviting information providers to workshops; presenting and discussing knowledge maps (outreach event); organizing hands-on workshops devoted to this community (outreach event); visiting their institutions if possible (outreach); involving them in the compilation of a "wish list" from stakeholders with key technological needs to guide joint proposal submissions in the area of research infrastructures (through face-2-face meetings or email communications); exploring implementation strategies with them; organizing expert hearings at libraries, archives and institutes that have specialized collections.

Academia: publishing articles in peer reviewed journals (preferably joint publications of Action members in new collaborative settings); giving guest lectures; organizing satellite meetings at major conferences; providing a growing bibliography at Mendeley.

Policy makers: presenting and discussing knowledge maps in the context of research evaluation and technology assessment; organizing events to discuss their own usual practice of knowledge organization and feeding this back to the research done by the Action members.

General public: Providing information in general publication channels as magazines for a wider academic and lay audience; organizing exhibitions with knowledge maps at public events.

H.3 How?

Dissemination (academic)

The academic dissemination addresses **academia** (experts, ESRs) and **information professionals** in their capacity as researchers. This Action aims to produce a significant number of important peer-reviewed journal publications. The production of edited volumes or special issues is expected and will be encouraged and a final common publication will be prepared and published at the end of the Action.

Networking and capacity building activities - Dissemination will be coordinated by the WGs and comprise the organization of: **(1)** The various training opportunities such as training schools, STSMs, hands-on workshops addressing information professionals and early career scientists at selected centres participating in the Action. **(2)** Lecture tours by prominent researchers from the participating countries. **(3)** One small scale workshop per WG per year, a satellite meeting at an international conference each year and the annual conference **(4)** Joint workshops and training schools with other COST Actions on interdisciplinary areas. **(5)** Fostering links with other relevant EU programmes. **(6)** Encouragement of publication of scientific collaboration within the Action in special issues of scientific journals. The WG coordinators will inform and consult the Dissemination Coordinator in all such activities.

The **Annual Meeting** will be a forum for presentation and discussion of results, giving an overview to participants. This Action aims to ensure that the audience includes a good representation of external experts from cultural heritage institutions. In this way the annual conference will also be a strong dissemination activity. This Action stimulates a secure feedback at the various meetings, and especially the Annual Meeting, supporting the Management Committee plan future meetings and schools as well as agree priorities for short term missions.

The **Dissemination Coordinator** is responsible for monitoring the scientific publication strategy in terms of cross-research group collaboration, domains, and appropriate publication channels. The Dissemination manager will also ensure the updating the bibliography, and of the website.

Action specific web site - A dedicated interactive web site will be established that will contain information about partner groups, research activities, conferences, workshops, symposia, list of potential host groups for short visits and trainings, forthcoming activities, exchange of code and datasets. If needed, wiki's will be established for collective documentation and writing. If possible, datasets to be used in the Action will be deposited at a certified archive and made open access after the termination of the Action at least. This Action will oblige open source principles for code and developed tools using established public archives and licensing standards.

Bibliography: The web service "Mendeley" will be used to document scientific progress and make pdfs of publications available and to document the bibliography of the Action.

Dissemination (beyond academia) - outreach

Outreach addresses: **information professionals** in their capacity as information service providers, **policy makers, academia** and the **general public** in their function as possible users of knowledge maps.

The **Outreach coordinator** coordinates and monitors the outreach activities of the Action initiated by WG coordinators and the MC. Specific outreach events are **(1)** visits to scientific and cultural heritage institutions (archives, libraries) together with hearings of information professionals working there, **(2)** inviting information professionals to Action events, and **(3)** arranging lectures addressing the students of Schools of Information and Library Science (also known as iSchools). During the course of the Action a "*wish list*" based on input from the end users (scholars, ESRs, information professionals, general public) with key technological needs will be composed and used to guide joint research, proposal submissions and technology transfer. Other concrete outcomes include **Guidelines** for future implementation together with **Demonstrators** of knowledge maps, whose development will be monitored and included as regular agenda items at MC meetings.

Participants will be encouraged to publish their work in the general press and take up invitations for media exposure where this is appropriate. Similarly the publication of non-technical articles for the lay community will be encouraged.