Convergence of Digital Libraries, Museums and Archives to Collective Memories

Gerald Jäschke, Matthias Hemmje, Erich Neuhold
GMD – German National Research Center for Information Technology
IPSI – Integrated Publication and Information Systems Institute
Dolivostr. 15, 64293 Darmstadt, Germany
[jaeschke, hemmje, neuhold]@darmstadt.gmd.de

Abstract
This paper describes the effects and resulting information technology support requirements created by the increasing convergence of Digital Libraries, Archives, and Museums towards Collective Memories. It provides an introduction into Collective Memory (CM) / Knowledge Management (KM) models as well as into human computer interaction support for such models based on mechanisms of interactive information visualization out of data stores, dynamically generated during the ongoing dialogue. Finally it outlines the nature of such an interactive, i.e., visually direct-manipulative information visualization dialogue and identifies a conceptual framework for a generic information visualization system, consisting mainly of three components (information model, visualization model, and information visualization function). Implementations of the system model will, first of all, provide technology to effectively produce visualizations of the information space, tightly coupling information objects and their representatives to the user, the visualization objects.

1. Introduction

Berners-Lee [4] writes, “Hope in life comes from the interconnections among all the people in the world. We believe that if we all work for what we think individually is good, then we as a whole will achieve more power, more understanding, more harmony as we continue the journey. We don’t find the individual being subjugated by the whole. We don’t find the needs of the whole being subjugated by the increasing power of an individual. But we might see more understanding in the struggles between these extremes. We don’t expect the system to eventually become perfect. But we feel better and better about it. We find the journey more and more exciting, but we don’t expect it to end.

Should we then feel that we are getting smarter and smarter, more and more in control of nature, as we evolve? Not really. Just better connected – connected into better shape. The experience of seeing the Web take off by the grassroots effort of thousands gives me tremendous hope that if we have the individual will, we can collectively make of our world what we want.”

The prime aim of many currently ongoing R&D projects related to Digital Libraries, Archives and Museums is the creation of sustainable Web-based virtual information and knowledge environments enabling citizens and professionals to capture real world multimedia information and knowledge of high quality as housed in – or whose quality is endorsed by – museums, archives and libraries, and to incorporate such information into tailor-made dynamically adaptive virtual exhibitions. Users of the system will be able to enhance the cultural information and knowledge objects they retrieve, as desired, with their own content: the resulting personalized virtual exhibitions, in turn, once made publicly accessible, will be able to be used by other citizens and professionals as the basis of further virtual exhibitions.

Most of these projects are highly innovative because they are performed within a holistic approach:
They emphasize the needs of users and the uses of natural world information outside the formal scientific and scholarly communication system. There is a mass of information about the real world now accessible via the Internet: what portions of this mass would common citizens and professionals really like access to?

They determinedly build upon existing publicly accessible information and knowledge management systems and services concerned with real world information and knowledge: especially those developed and being developed within R&D frameworks. Therefore they usually do not need to reinvent wheels.

They create generic systems and process architectures whose parameters can be set dynamically by the system’s providers and users alike. This is technically challenging; but it is a key logical development, which is ideally
required within these types of Web-based virtual information and knowledge environments;

They use existing advanced state-of-the-art technologies as the basis of further development which can support user-centered information visualisation, interactive navigation, exploration and manipulation as well as collaborative working and general information and knowledge management of multi-media digital cultural knowledge and information objects.

They establish economically and organisationally what would need to be done to ensure that the overall system developed can be sustained in the longer-term – including, as transpires to be optimal, elements of commercial exploitation.

2. The application context

Such R&D projects related to Digital Libraries, Archives, and Museums usually focus on a specific field of real world information – which might seem to those outside of the field to be relatively narrow and uncomplicated. In reality, in common with most areas of ‘real world’ information, the target information model often reveals to be rather complex. Prior to the growth of the Internet and the World Wide Web, people and organizations would make public the results of their studies of ‘the real world’ through paper-based data, information, and knowledge artifacts: journals, books, reports, conference proceedings, magazines, newspapers, and so on. The information made available via these artifacts often would have been peer-reviewed prior to publication to provide some guarantee of its relative worth. Especially within science, commercial publishers came to play a key role, though many scientific societies also became publishers, too. Secondary and tertiary abstracting and indexing journals, bibliographies, catalogues, directories, encyclopedias, and so on became prominent as the volume of primary literature grew.

Because interest in real world data, information, and knowledge objects and interest in data, information, and knowledge about collaborative research activities and processes performed in special research fields extends well beyond and, indeed, pre-dates the widespread availability of formal scientific publications (one thinks, for instance, of Linnaeus), there was – and continues to be – much data, information, and knowledge of relevance made publicly available (one hesitates to write ‘published’) outside the formal communication system. Fortunately, libraries, archives, and museums, as well as collecting and making available to their users, e.g., publications which had been produced within the formal system, have long tried in addition to ensure coverage of this less formal material: manuscripts, expedition reports, field notes, drawings, artworks, and so on. They have also tried to ensure that they are able to secure the resources that will enable them properly to conserve data, information, and knowledge materials, which are inherently unique, or are at least very rare. The result is that libraries, archives, and museums now collectively contain an immense cornucopia of such material – as rich as any in the world: but a cornucopia which for the most part has so far been inaccessible to those who are unable physically to visit the libraries and archives themselves.

Meanwhile, the scientific departments of the museums have for centuries been developing cornucopial, i.e., data, information, and knowledge collections of their own: For example, the biological specimens and their associated data, each representative of particular biological organisms – or species. Readers will be aware that defining what we mean by a ‘species’, naming that species, and establishing its phylogenetic (evolutionary) relationship to other species are extremely taxing activities. There is also the challenge of scale: it has, e.g., been calculated that some 1.8 million different species have been characterized and described in the literature to date. The ‘type’ specimens – the specimens used initially to define new species – of perhaps two-thirds of these 1.8 million are housed within, e.g., European institutions alongside collectively hundreds of millions of ‘non-type’ specimens. Beyond the species/specimens themselves, a vast amount of specialist biological and environmental data has been gathered. In recent years, the development of the techniques of molecular systematics (DNA analyses, etc.) has given the whole field a new imperative.

All this resulting data, information, and knowledge, some formally published and accessible through the normal communication channels, but the majority not, can potentially be made available directly or indirectly to the citizen and professional interested in ‘real world information’ as, e.g., digital surrogates of cultural objects accessible via the Internet. Alongside the traditional publishers, organizations can become their own (electronic) publishers, outside of the formal (paper-based) communication system; and, even just with regard to ‘the natural world’, tens — if not hundreds — of thousands, worldwide have already done so. Meanwhile, the commercial and scientific society publishers themselves have moved into ‘electronic publishing’ via ‘The Web’, joining — and in some cases forming alliances with — broadcasting and news organizations such as the BBC and Reuters, who also purvey information about ‘the real world’ such as films and news. Databases have been made ‘web-accessible’: especially important in this field is the use of GIS techniques to provide, for example, species distribution maps. The overall result is a very wide range of forms of literature within which information concerned with ‘the natural world’ appears, whose media formats can be data, text, graphics, image, moving image, voice, or ‘multi-media’ combinations of these. On top of this
already complex mélange of natural world information artifacts and the stakeholders who make public such artifacts — and driven by its very complexity — we have then seen the formation of organizations whose raison d’être is the creation of order out of a potentially uncontrolled diversity (no-one “runs” the Internet): the collaborative network services and portals. Such facilities generally are concerned not just to try to bring some semblance of order to the Web; they also use various manual and automated techniques to try to sift out information of relevance to each of their target audiences.

3. The innovation challenges

3.1 Collaboratories

Apart from the novel combinations and applications of virtual cultural object representations and related cultural content management processes applying metadata based approaches, currently ongoing state of the art research introduces the concept of virtual cultural information and knowledge environments to implement collaboratories in educational and non-professional communities.

Collaboratories as “centers without walls” [18] “in which the nation’s researchers can perform their research without regard to geographical location – interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries” [13] have been put to the test in natural and computer sciences and in distance learning and adult education in the USA since the early 1990s [9].

In other research fields, for example, in arts and humanities, there exist so far only pre-studies of an editorial Web collaboratory, for example for the effort to establish a European cultural collaboratory with the help of email publishing.

The technical span of existing approaches to establish and support the concept of collaboratories varies considerably. In the simplest case, a working group shares data on a special topic and manages a mailing list or bulletin board. More elaborated examples of collaboratories comprise remote access to scientific instruments for shared analysis and experiments, such as astronomical telescopes or particle accelerators; or enlarge the interactive and communicative features of the collaboratory by Audio/Video conferencing or a more structured information and knowledge interchange such as annotation environments like CoNote [6].

Museums and other archiving institutions maintain very effective – but formal and impersonal – collaboratories and connections. Tacit expert knowledge often is made explicit only by face-to-face communication on the occasion of conferences, workshops and professional meetings. Web-based collaboratories interrelate the data collection, the work processes involved in accessing and interpreting materials and the fluctuating community of concerned experts and professionals and more importantly of interested non-experts and non-professionals. The challenge is to enable user groups differing in professional and technical status to work collaboratively and cooperatively to the mutual gain of all.

3.2. Collective memories (CMs) and knowledge management (KM)

The emerging knowledge economy is changing the existing industries and scientific research activities. The knowledge movement in institutional thinking is considered as to be very important in enhancing the competence of institutions and organizations, especially for those, such as Libraries, Archives and Museums where information and knowledge are the primary basis for business development.

In the past 30 years, information management systems have been popular in many institutions and enterprises. Working processes and administration in such institutional organizations are almost controlled by some kind information management systems. However, there is great distinction between knowledge and information. Information consists largely of data organized, and categorized into patterns to create meaning; while knowledge is information put to productive use, enabling correct action and help people to produce further knowledge and by this means innovation. The primary function in a KM system is to rapidly build and utilize the collective knowledge from both inside and outside of the organization and helps staffs in an organization to update their knowledge and do innovation.

The theory and methodologies on KM have been intensely studied in recently years. The researches focus on the human learning cycle, the architecture of KM systems, the knowledge flows management, knowledge repositories and libraries, corporate memory and knowledge pump, knowledge cartography and the communities of staffs.

Due to the globalization, many institutional organizations now work, compete, and co-operate on a worldwide scale. The members of a professional, semi-professional or amateur work group may work in different places, even in different countries. Therefore, the research on distributed KM becomes more and more important. Distributed KM can help information and knowledge workers in institutional organizations to reproduce their core competencies and institutional identity regardless of geographical distance and linguistic and cultural differences of the information and knowledge markets in which they operate at the same time.
Although both central and distributed KM have been studied, however, most representation of KM and CM are prescriptive and working process-oriented, often many definitions are given for one concept, it makes some ambiguous in concept, functions even purposes.

3.3. The knowledge life-cycle, KM models and architectural infrastructures

The knowledge life cycle hinges on the distinction between “tacit knowledge” and “explicit knowledge”. Explicit knowledge is formal knowledge that can be packaged as information. It can be found in the documents of an institutional organization: reports, articles, manuals, pictures, artifacts, multimedia documents as well as software. Tacit knowledge is personal knowledge embedded in individual experience and is shared and exchanged through direct, face-to-face contact. The main function of KM systems is to turn information into knowledge through the interpretation of incoming highly non-standardized information for the purposes of problem solving and decision-making. It is also about content creation: the generation of new knowledge to propel the innovation process of institutional organizations. A KM system can help maximize knowledge development, storage, distribution and combination per se in every knowledge cycle.

Corporate memory is a very important concept in KM. Recently, G.V.H Heijst, R.V.D. Spek and E. Kruizinga summarized the former work given by Wiig 1993, Nonaka and Takeuchi 1995, van der Spek and Spijkervet 1996, pointed out that a CM should be a system with following characteristics:

- It supports new knowledge developing in an organization by tracing the activities of the organizations, such as, recording failures and successes in activities. It is of the ability to collect and store knowledge.
- It can help in securing new and storing knowledge in an organization. It can make it easy to let individual knowledge be accessible to others in the organization who need that knowledge. Furthermore, this knowledge must be available at the right time and place. Knowledge stored in corporate memories is persistent over time and can be retrieved easily.
- It supports a special kind of knowledge distributing ability and offers a facility for deciding who should be informed about a particular new piece of knowledge.
- It helps people to combining available knowledge in automatic or semi-automatic ways.

Clearly, a CM is not only a repository in a KM infrastructure, it consist of many subsystems, such as, the file document systems, meta-knowledge management systems, auto and semi auto learning systems, knowledge bases and information databases.

The knowledge pump is the most complex type of CM architecture. In theory, this model ensures that the knowledge developed in the organization is fully exploited to improve the performance of the organization. The management of the organization enforces an efficient functioning of the lessons learned cycle. In technique, it provides online for existing intranet and extranet-based communities. This takes the form of a technology that channels the flow and use of knowledge in an organization, connecting document repositories, people and process, and leveraging formal and informal organizational charts and structure. It helps communities, defined by their common interests and practices, more effectively and more efficiently share knowledge. The knowledge pump takes a document-centric view of the knowledge generation by relating the learning-cycle directly to the reading and writing documents.

The knowledge life cycle is typically divided in to six stages. First information and knowledge workers gain experience and therefore personal expertise. Then the individual experiences are shared among co-workers. This may lead to a kind of group learning as well as collaborative group knowledge production. The result is that others can now apply the knowledge, i.e., the lessons learned by one individual user. In fact, each individual has to decide if to accept the provided knowledge or not. If users accept the lessons, they will store them in CM. The CM furthermore should offer a mechanism that can reward to communicate lessons learned to others. The information and knowledge workers who accepted the lessons, will apply them in their future works and provide their feedback to the CM.

Infrastructures and architectures for CMs can be classified as problem-solving models and workflow-processing models. In the problem-solving models, the learning process is the top-down learning type. With top-down learning, a particular knowledge area is recognized as promising, and that deliberate action is undertaken to acquire that knowledge. The knowledge is collected and distributed along with the problem solving processes. In this model the CM is mainly in charge of the communications among actors or activities, maintains the dependences of sub-goals decomposed from the main goal, arranges the common resources to be shared by activities in right time and place, and keeps the knowledge in the shared memories consistent.

In workflow-processing oriented CM models, which is also often related to a process-oriented learning model, the learning process is the bottom-up learning type, where the bottom-up learning refers to the process where a staff (either on the management level or on the “work floor”) learns something which might be useful and then this “lesson learned” is distributed through the organization. With the term lesson learned we refer to any positive or negative experience or insight that can be used to improve
the performance of the organization in the future. In this learning process knowledge is collected, combined, generated and distributed according to the order of activities happened in the organization and the file system is organized in deferent level.

Both models are of the basic characteristics of CM mentioned in section 2, while the main difference is that they are organized according to two different learning processing, i.e., the up-down type and the down-up type. If independent software agents carry out the independent components in a CM, the CM is often called “Agent oriented corporate memory”.

The problem-solving model is very efficient for a specific activity of unstructured collaborative knowledge production for problem solving while the workflow-processing model adapts better to the well structured knowledge production and information working process in institutional organizations. The technique given in both models constitutes the main technique in establishing a CM system. However, we think both of the models lack generality for CM because a general model for CM should be independent with a concrete problem and an organization, it should depend on the basic knowledge life cycle only; furthermore the retrieval function and the contribution function for information and knowledge workers in a knowledge life cycle should be added and emphasized in the concept and architecture of CMs in the future. A CM is real useful only if there is sufficient knowledge for the CM to distill and there is a strong power to support CM to work. We believe the power in a CM should come from the enthusiasm of staffs to participate the knowledge invention processes, and the enthusiasm should come from the fact that information and knowledge workers feel that the CM can really help them in their work and they can get the timely encouragement from the organization after they did a contribution in a collaborative environment.

3.4. Virtual information and knowledge environments

Building on its existing expertise in the area of distributed collaborative digital libraries, multimedia archives, information retrieval, filtering, linking, enrichment, personalization, and information visualization GMD-IPSI is currently aiming at supporting CMs by means of supporting its collaborative content and knowledge management functions, mainly about historical and cultural, i.e., in our case Digital Library, Archive, and Museum contents, collections, and the related knowledge portal and management applications by focusing on research and development enabling the efficient and flexible implementation of content and knowledge management portals and virtual information environments.

Content and knowledge portals are web based distributed information systems which enable the corporate as well as the personal and cooperative acquisition, management, access, distribution and usage of information and personalized content and knowledge.

Virtual information environments are user interface front-ends for naive users to Internet-based information systems. They apply real world as well as abstract metaphors and information visualization layouts. In using these means a visually direct-manipulative and sometimes also collaborative information dialogue paradigm is supported at the user interface level.

Within the content and knowledge acquisition and management functionality of content and knowledge portals, content based indexing (including, e.g., feature extraction and similarity computations) of all acquired documents is ensured by means of the underlying digital library infrastructure. Furthermore user-, task-, domain-, situation-, location-, and goal-oriented metadata are acquired. They relate to the personalized working context of users who provide content and knowledge which are incrementally added to the collections managed by the underlying digital library. Finally, document-oriented and structural metadata are acquired to build an additional information retrieval basis.

Within the content and knowledge access, distribution and usage functionality of content and knowledge portals, methods of information retrieval, information filtering, linking, and personalization as well as automated enrichment are applied to the collections described above. By these means, naive users are provided with mechanisms supporting content-, meta-data-, structural-, and context-oriented searching. Furthermore advanced methods of managing the information needs of users and their information need and knowledge profiles enable the implementation of collaborative information agents which autonomously search collections as well as dynamic information streams for information that could satisfy the information needs of users, enrich their knowledge profiles, and thus support a dynamic interaction with and usage of the newly acquired material by the user.

Building on such content and knowledge management infrastructures, the virtual information environments introduced above will support the cognitively efficient access to and management of content and knowledge collections as well as information need and knowledge profiles of naive users.

Within GMD-IPSI research divisions DELITE and OASYS, content and knowledge management infrastructures are built on the basis of full fledged commercial object-relational database management systems as well as on basis of a suite of XML/XQL processing tools. The ML/XQL tool suite provides a comprehensive, lightweight solution for querying and storing large, distributed XML documents. It is based on
two major components, a persistent implementation of the W3C DOM (Document Object Model) API and a web-aware query engine for the XQL language. The lightweight implementation requires only an off-the-shelf Java runtime environment. Having a remarkably small footprint for memory, storage, and CPU, it is well suited for distributed applications, which require query able document/data repositories. Likewise, it can serve as a solid base for CD-ROM applications. With its integrated support for HTTP and URL, the XML/XQL suite also enables the generation of scalable, multi-user server applications built on Javasoft's Servlet technology. A particular strength of XML/XQL server applications is the support for seamless integration of existing web information services into value added information products.

Virtual information environments are built on the basis of DELITE's multi-user information visualization (IVIS) framework, which enables the dynamic construction of virtual information environments. At the same time, DELITE's IVIS framework enables the distribution and sharing of virtual information environments to and by different users on different sites. In this research context, DELITE applies methods of e.g., dynamic 3D information visualization to provide naive users with a visually direct-manipulative user-interface to information or knowledge management systems. Virtual information environments can therefore, e.g., dynamically visualize retrieval and filtering results and their relation to the users' information needs. They can furthermore automatically adapt to changes in the collections of the underlying digital library and knowledge management systems. Furthermore, user information and knowledge profiles can be visualized and interactively navigated and explored within virtual information environments. This functionality is achieved at run-time by means of dynamic layout and scene construction components, by means of automatic metaphor definition mechanisms as well as by means of informational, personalized and interactive level-of-detail and tight-coupling mechanisms.

From the users' point of view, DELITE's ongoing research is aiming at establishing an integrated framework for the efficient generation of digital libraries, archives-, and museum-oriented content and knowledge management portals and virtual information environments based on multimedia capable archives and digital libraries. Application areas for teaching and training, e-commerce, as well as for advanced event oriented information, communication, production, publication, distribution and archiving technologies can be built on this framework in the future. From the research point of view, more intelligent, autonomous and collaborative content and knowledge management environments and information processing technologies enabling efficient multimedia information integration (mainly based on XML and MPEG), storage (mainly based on object relational database technology), indexing, retrieval, filtering, exploitation, and reuse as well as dynamically shared, scalable, and interaction-secured multi-user information environments based on next generation scene description and visually direct-manipulative interaction standards will be designed, specified, prototypically developed and experimentally evaluated.

**Excuse: Economics and organization**

Virtual information and knowledge environments for individual use serve human as well as collaborative institutional entities to exchange information and knowledge. The core economic and organizational question is how to make such a system, and the virtual CM/KM network it creates, sustainable, with investment in content and infrastructure guaranteed for the longer-term by following a community-oriented approach. The Internet has created and stimulated growth in the last few years of several interesting cases of knowledge exchange systems: perhaps the most famous example being the network of Linux developers. (Note that we use network here in the sense of interacting agents, and not just with reference to the communication network that makes exchange possible.) Less famous, but of similar type are newsgroups, virtual communities, and electronic journals. The latter are sometimes maintained completely by researchers, without traditional publishers being involved. We need to structure the virtual information and knowledge environments so that they function as sustainable networks of information exchanging entities. This will not exclude the network requiring subsidiaries by governments at a state or European level.

**3.5. Supporting the information and knowledge dialogues in a natural way**

Supporting information and knowledge retrieval dialogues between naïve users and information and knowledge management system systems supporting a CM is a non-trivial task. Visualization is one promising approach to improve human-computer interaction. In today's European landscape of research, the application of visualizations in user interfaces for information retrieval in digital libraries and multimedia archives focuses on five mainstream aspects: Result visualization [7], multimedia access visualization [15], indexing of visual objects collections [1], visual user interface metaphors [14], and visual user interface architectures and frameworks [12].

One challenge within the aspect of result visualization is to support users to easily identify the documents relevant to their interest by appropriate output design of computerized database searches, for example by displaying additional internal document information, thus shortening
search time and improving the users’ feeling of ease and satisfaction [7]. As for the multimedia access visualization aspect, the underlying technologies used to capture, compress, store and transmit multimedia data are now so far advanced that the development of digital multimedia archives has become an achievable target, allowing focusing the efforts on the human-computer interaction. For example, content-based browsing of digital video typically employs keyframes, each representing the content of one coherent camera shot. Considering the many possible useful ways of keyframe-based browsing of video, a single all-encompassing interface would be impossible and several variations are needed, categorized by different underlying dimensions: layeredness, provision of temporal orientation, and spatial vs. temporal visualization [15]. Due to their nature, automatic recognition and indexing of visual object collections, for example Arabic scripts and calligraphy, is difficult and yet not completely solved. Through visual feedback and visual interaction users can enhance the indexing result of automatic extraction techniques like, e.g., pattern recognition and feature extraction [1]. Visual user interface metaphors are well established to provide users with a familiar interaction environment. For example, the Visual Book study explored the importance of the visual component of the book metaphor for the production of more effective electronic books, while the WEB Book study applied the findings to the production of books for publication on the World Wide Web [14]. To tackle the aspect of visual user interface architectures and frameworks, AQUA [12] embeds several common and advanced query techniques into a query interface which supports iterative query refinement. Although all these works tackle most interesting aspects of applying visualization, i.e. the visual communication medium in order to improve the human-computer interaction to interface digital libraries, archives and museums, there are only few works that investigate on the constructive aspects of a tight coupling between information objects in underlying collections and the visual objects in an image, i.e. of the visual communication medium itself. But only if we succeed in coupling these two worlds on the basis of a generic, automatic mechanism, visual user interfaces for collective information and knowledge collections forming CMs can become cognitively more effective and, at the same time, their technical production and maintenance more efficient.

4. Generic information visualization model

Throughout the upcoming chapters a generic information visualization model is introduced and all its concepts will be illustrated by an example from the domain of congress information systems, the so-called “scientific program” of a congress. Its application data schema is displayed in figure 1. A congress’s scientific program comprises the following entities: Abstracts accepted by the committee are grouped thematically and assigned to sessions in which they will be presented. Each abstract is written by a specific person, the author. Let first name, family name and nationality be the only attributes of interest here. Title and text body characterize the abstract they belong to. In case an abstract gets presented as a poster, the poster’s graphical content has to be added. Sessions of the congress take place on a specific day within a specific period of time and eventually get financial support from a sponsor. Besides, sessions are differentiated by their type — like poster session, discussion or plenum. Each session carries a title, anyway.

For later modeling this abstract conception already contains unique identifiers.

4.1. Information dialogue

From today’s research point of view the computer-supported information process can be characterized as an interactive dialogue with the computer system, in which the selection, perception, and processing of appropriate information is the goal [16]. Information requests by the user and information offers by the system form the sequence of activities within this information dialogue, accompanied by additional clarification and rating processes [2][3][11]. User and computer activities alternate with each other. Each single dialogue step has to be seen in context with the history of preceding steps. In doing so, users modify the formulation of their information requests step by step. Accordingly, the content of a series of retrieval results may change until it finally matches the user’s information needs [2][3]. In the ideal case each new information dialogue step improves the information state of the user with regards to his information needs.

Information objects define the content of an information source. Information objects comprise both
informational concepts as well as informational relations. Their specification ensues in terms of descriptors chosen by the information source’s author. In case of relational databases this will be the database schema.

Data sources enable access to an information source’s content. Requests are directed to the data source.

The information objects of an information source span its information space. In analogy to the topology of the real world’s places, a set of information objects can be regarded as a set of places inside an information space. Moving to another place is an informational navigation and therefore an activity within the information dialogue’s scope deployed by users to change their conceptual position within an information space.

The whole process of such an information dialogue is called exploration. The degree of information space exploration calculates as the share of visited or different visited information space objects (in spatial terms called “information places”), respectively.

The set of information objects opened by means of exploration defines the context set of informational navigation. As a consequence, exploration and navigation create the context set of an information dialogue.

By visiting distinct places within an information space during exploration, the user follows an information oriented information path. These information paths consist of a sequence of navigation actions on top of the context set created by exploration activities.

The information dialogue aims at users perceiving and processing appropriate information provided by the system. The interest set describes the information objects of an information space that have the potential to fulfill the user’s information needs. Possibly there will be information objects in the interest set that are not accessible via the data source. Seen the other way round, in most of cases only a small subset of information objects contained in a data source are relevant to the user’s information needs. If this so-called relevance set is empty the data source cannot contribute to satisfying the user’s information needs. By means of navigating and exploring the context set the user tries to isolate the core set of information objects with highest relevance from the data source’s probably huge relevance set.

By means of navigation and exploration the user changes his conceptual location inside an information space. This focus is characterized as the specific element of the context set which is currently in the center of the user’s attention.

In order to plan their further information actions, users depend on viewing information-object details like, e.g., attributes or complete document contents. This inspection of an information object within the context set is defined as accessing data and attributes which are components of this context set element.

4.2. Interactive information visualization

Gathering information. The user’s demand for information in order to satisfy his information needs is the starting point of the interactive information visualization process (figure 2). His request induces the system to perform an information retrieval on the data source, leading to a result considered as relevant. The result’s information objects then get rated and integrated into the context for later reuse. The rating and integration of the exploration result set explicitly evokes a change of the context set.

Visualizing information. Within the next step, the information visualization creates a visual representation of the context set. In doing so, information objects of the context set get mapped onto visualization objects by means of an information visualization function. This process of combining information and representation produces information visualization objects, or constructs. The collection of information visualization objects forms the structured information visualization context — the visible pendant of the information dialogue’s context.

Visual objects are descriptions of visually perceptible pictorial presentations of an object. The visualization model describes a set of visualization objects and the rules according to which the visualization objects can be
arranged into scenes. Information visualization objects are descriptions of visually perceptible pictorial presentations of information objects. The description of the structural combination of all information visualization objects denotes the information visualization context.

Finally, computer graphical display methods create an image or a sequence of images, respectively, presenting the information visualization context. Only after having performed this last rendering step can users perceive the information from the resulting image by means of their visual perceptive system.

4.3. Information model

Within the exploratory information dialogue, the context set is the starting point for the information visualization. Its content and structure determine the visualization’s appearance and consequently also the informational concepts and informational coherence presented to the user. The information model offers a notation for describing the context set.

A poster exhibition. Take the following request to the congress’s scientific program (figure 1) as an example:

![Poster exhibition query tree](image-url)

Figure 3. Poster exhibition query tree
1. Retrieve all days on which events of the scientific program take place and show their dates.
2. For each day, list all poster sessions that occur on that day and show their title, start and ending time, the location, and the financially supportive third party (sponsor), if any.
3. For each poster session, display all poster contributions presented within this session.
   a. For each poster contribution, show it’s title, body, and if available, the poster graphics of the underlying scientific abstract.
   b. For each poster contribution, show the author’s first and family name, and his origin.

A nested relational data model. In its form and expressive power the information model is equivalent to the nested relational data model [17]. Roughly speaking, such nested relations can be considered a set of tuples whose attributes themselves can again be nested relations [8].

The nested relational data model extends the relational data model. First normal form (1NF) for relational data models requires attributes to be atomic. The nested relational data model repeals this first normal form restriction, enabling relations inside of relations. This nesting results in complex tuples with a hierarchical structure — also known as NON1NF or N1NF.

Any query schema has exactly one unique root node marking the entry point for exploration.

A view onto data sources. The structures of the information model and the accessed data source need not necessarily match. In this sense, the information model can be considered a specific view onto the informational concepts and coherence of the underlying data source. It is the information system’s responsibility to map between the information schemes of the information model and data source during all retrieval, rating and integration processes. Without considering the mapping rules here we will now point out the outstanding benefits gained by separating these two structures by means of the example.

In both schemata identical concepts can be recognized. Abstract, author and poster session inside the query schema are mapped onto the scientific program’s abstract, person and session, respectively. In doing so, for the concept of person renaming occurs, emphasizing the person’s particular role as author in this context. Moreover, the query schema introduces a new concept called day, exploited for dividing the set of poster sessions into groups. The data source attribute date gets promoted to a query schema record node. Instead of all sessions of the scientific program the system will display only those of type poster session. This constraint corresponds to the selection as introduced in the relational algebra. Since all poster sessions are of one and the same type, this attribute is omitted for display purposes. This approach corresponds to the projection as introduced in the relational algebra. Within the extended entity relationship (EER) model relationships between concepts are bijective. An abstract is written by a person, the author. The other way round, the author lists a couple of abstracts within his bibliography. Using the nested relational data model the query schema commits to one perspective.
Sets, bags and lists. Set-of nodes provide a set of homogeneous elements. Typically, they summarize a subset of the content space — the day entry with date “1999/08/10” from the example can represent all poster sessions occurring at that date. Because of this characteristic, record nodes are employed to support user navigation. Adding new day entries for every poster session scheduled for August 10th, 1999, would be counter-productive; the date statements too abstract. Consequently, the query must be capable of restricting the result, potentially a bag, to a
result set in the mathematical sense. However, bags must not be excluded.

The query schema defines the structure in terms of which the user approaches the context set. To support the user in his efforts to orientate, bringing the result set elements into order is wise.

**Key attributes.** Attributes helping the user to orientate inside information space are called key attributes. In contrast to the key attribute constraint within the relational data model, information model key attributes do not unequivocally identify entities. Record lists are ordered by their key attribute.

### 4.4. Visualization model

The process of information visualization creates a visual representation of the informational context set. Information objects of the context set are mapped onto visualization objects and by these means the information visualization context is built up constructively. Whereas the information model provides a notation to describe the context set — the information, the visualization model offers a notation to describe potential visual representations — the visualization. This comprises both the visualization objects and the visualization schemata shaped by these objects and their structural relations.

**An exhibition tower.** In the following, an example for the presentation of visual objects and their structure is introduced by means of the so-called “exhibition tower”, defining a visual metaphor for visualizing the poster sessions of a congress (figure 5):

Within the exhibition tower, an elevator connects the different floors of the building. Inside the elevator cabin, a line of labeled buttons directs the elevator to the right floor. These buttons symbolize the building’s floors. On every level — immediately in sight after leaving the elevator cabin — a sign displays information about this floor. A corridor is directly attached: it leads to four halls on that floor level at maximum. Big doors separate corridor and halls. Signs on the door inform the visitor about what to expect behind. The hall, eventually furnished with wallpaper, hosts a banner, a pillar, analog clocks on the wall and stands. The remaining elements open themselves in a similar pattern.

**A box of bricks.** Visualization objects are the building blocks deployed for composing potentially very complex scenes for information visualization purposes. The following visualization objects characteristics are essential for proper visualization of information spaces:

- Visualization objects must be capable of displaying the data types appearing in the context set.
- Visualization objects must be capable of copying accumulations of information objects within the context set.

Even though visualization objects are visual objects, their classification does not follow their visual properties but ensues by means of the role they have been assigned for representing information objects and interacting with the user.

**Visualizer objects.** The visualization model (figure 6) introduces the node types data-type, accessory, decorator, and jumper as well as the container nodes aggregator and classifier.

- **Accessory** objects are simple objects that can display atomic data of specific type.
- **Data-type** objects specify the type of atomic data an accessory object can display.
- **Classifier** objects comprise a set of homogeneous objects. The number of subordinate objects cannot be assigned at will. Upper (maximum) and lower (minimum) boundaries limit their quantity to a range credible in the eyes of the human observer. Moreover, the typical number of occurrences (average) can be specified.
- **Aggregator** objects bundle a — possibly empty — set of objects of different kinds.
- **Decorator** objects contribute to the representation of information only in an indirect way. Not intended for the display of single information-objects or their interrelations, decorator objects are rather decorative elements in the scene to enrich the overall visual appearance of the information visualization context. The concrete properties of a decorator instance are not determined by the context set but may be chosen randomly. Placing decorator objects makes locations inside the virtual scene distinguishable and therefore supports navigation within the interactive information visualization.

In addition to the above-mentioned elements **jumper** objects enable the user to shift between visualization schemata. Changing the visualization schema implies switching to another virtual scene.
Visualization schema. The visualization schema could be described as a collection of visualization objects that typically appear in a common environment — as is the case in our example of buildings, floors, and corridors. What is more, the visualization schema imposes a structure on the objects; it can be considered a blueprint for potential combinations of visualization objects.

Decoupling objects and their combination conforms to the everyday experience that objects not only occur once in one specific surrounding but in diverse places. Nodes within the visualization schema are not the visualization objects themselves. Instead, they refer to visualization object type. As a benefit, visualization object definitions can be shared across visualization schemata.

Geometric construction. This paper deals with the conceptual and structural characteristics of visualization objects exclusively. The role they have been assigned for representing information objects and interacting with the user drives their classification. The rules for admissible combinations of objects are based upon this classification. Consequently, these pre-given structures subordinate the geometrical arrangement of visualization objects (figure 7). Our group will report about this geometrical construction in further papers to come.

Manual world design. The underlying concepts of visualization objects’ geometry, classification and structure offer starting points for automatic virtual world setup. However, the analysis of broad experiments regarding requirements and expectations towards a museum authoring tool resulted in the insight that there is no use in automatically placing objects inside dynamically generated worlds — the author of this virtual world would change it, anyway [10]. Since this wish for individual design seems to also hold true for generic information visualization systems, this approach is considered less promising. Instead, the creation of visualization schemata is viewed as manual world design. Object selection and arrangement lies in the hands of a world designer.

4.5. Information visualization function

Information visualization is the process of bringing information into a visual form. The information model and the visualization model provide for notations to describe context set and a potential set of its visual representations, respectively. In this connection, the process of information visualization can be understood as a mapping between information schema nodes and visualization schema objects. The information visualization function — or simply function in short — establishes this mapping.

Structural sound metaphors. To make the resulting visualization an effective one the following guidelines must be obeyed:

- The structural composition of the query result on the basis of the part-of relationship must be retained.
- For each data item inside the information schema an object from the visualization schema must be assigned.
- For the navigation the user must have all necessary information at hand.

Put into a formal notation, metaphors that fulfill these requirements are also known as structurally sound metaphors.

A poster exhibition inside the exhibition tower. As an example, the previously introduced poster exhibition and exhibition tower are brought together. Together they build the metaphor of the poster exhibition inside the exhibition tower (figure 8).

The poster exhibition takes place inside the exhibition tower. Each floor collects all sessions of one specific day. Using the elevator, visitors reach the different platforms. They choose their destination by pressing on one of the buttons inside the elevator cabin. All the buttons are labeled by the date the floors represent. As users arrive on the selected floor, a sign again reminds them of the date dealt with on this platform. The remaining elements open themselves in a similar pattern.

Information interaction. The mapping between information and visualization objects exclusively highlight the aspect of visualization, going from information objects to visualization objects. But in the scope of interactive information dialogues the information system should not be designed as a one-way street. The way back, going from visualization objects to information objects and manipulating them, is still a field of further research investigation. Results within this area of our research will be presented in publications to come.
Figure 8. Poster exhibition at exhibition tower metaphor

4.6. System model

The presented information model, the visualization model, and the information visualization function form the foundation for the information visualization. Within this process, starting from the information schema, information objects are retrieved from data sources and, according to the rules of the information visualization function, get merged with visualization objects to result in
the visualization context. Figure 9 shows a system model modularizing this conception.

The core of the system model is the global repository, whose elements are the query repository, the virtual world repository, and the metaphor repository.

The query repository hosts a collection of information schemata and keeps them ready on call. The initial idea of such a repository is that there is a collection of useful standard queries or views, respectively, for every data source, which should be offered to different user groups.

To be able to map information objects onto visualization objects, having such a visualization objects repository at hand is a precondition. The repository contains both the virtual information objects and schemata as blueprints for potential virtual worlds.

Not every mapping of information objects onto visualization objects is valid or leads to a reasonable representation of the information space. Therefore, the metaphor repository, akin to the query repository, makes available a collection of useful metaphors. Each metaphor refers to exactly one query schema and one visualization schema. The user’s choice for one metaphor simultaneously fixes a request and a visual presentation. As a matter of principle, query schemata and visualization schemata are arbitrarily combinable, as long as they obey the conditions for structurally sound metaphors.

![Figure 9](image)

Figure 9. Generic information visualization system model

The scene constructor is perhaps the most important module of the system model presented here. It executes the process of information visualization and generates the virtual realities for exploration — the visualization context — the render engine will finally display graphically. Such three-dimensional scenes feature objects which occlude one another due to their structure. Following this observation, the scene construction can be performed stepwise, subdivided into sections, reducing both data source and rendering workload at the same time. With this method, the scene constructor generates the virtual world adaptively in response to user information requests.

The system administration creates and maintains the global repository. Tools for visualization object creation and their combination to visualization schemata (in the virtual world repository), tools for formulating query schemata (in the query repository), and tools for creating metaphors (in the metaphor repository) are part of the set of system administration modules.

Application data denotes the data source whose content space is to be visualized. Typically, these data sources already exist. Queries performed by the scene constructor retrieve the information objects from the data source. Please note that neither the system model nor the generic information visualization model restrict the physical arrangement or data format of the data sources.

Similar to the data sources, the render engine is considered a module external to the system model. Its task is to produce graphical depictions of the virtual world. The system model does not describe the rendering method to be deployed. Since the information visualization objects inside the virtual scene serve as the interface for human-computer interaction during exploration, the rendering component also has to accept user actions and forward them to the scene constructor. In this sense, the rendering engine may also be referred to as browser.

5. Conclusions and outlook

In this paper, we have described the effects and resulting information technology support requirements created by the increasing convergence of Digital Libraries, Archives, and Museums towards Collective Memories. We have provided an introduction into CM/KM models as well as into human computer interaction support for such models based on mechanisms of interactive information visualization. Finally we have outlined the nature of such an interactive, i.e., visually direct-manipulative information visualization dialogue and identified a conceptual framework for a generic information visualization system, consisting mainly of three components (information model, visualization model, and information visualization function). Implementations of the system model will, first of all, provide technology to effectively produce visualizations of the information space, tightly coupling information objects and their representatives to the user, the visualization objects. Current results only cover the visualization aspect, going from information to visualization. Future work will deal with manipulating information objects by means of their visualization object assigned. Secondly, the geometrical aspects of combining visualization objects on the basis of a classification by means of their role in the information visualization process requires further investigation.
References