

# **Personalisation and Recommender Systems in Digital Libraries**

## **Joint NSF-EU DELOS Working Group Report**

**May 2003**

Jamie Callan, Carnegie Mellon University, USA  
Alan Smeaton, Dublin City University, Ireland

Micheline Beaulieu, University of Sheffield, UK  
Pia Borlund, Royal School of Library and Information Science, Denmark  
Peter Brusilovsky, University of Pittsburgh, USA  
Matthew Chalmers, Glasgow University, UK  
Clifford Lynch, Coalition for Networked Information, USA  
John Riedl, University of Minnesota, USA  
Barry Smyth, University College Dublin, Ireland  
Umberto Straccia, ISTI-CNR, Italy  
Elaine Toms, University of Toronto, Canada

Digital libraries are collections of information that have associated services delivered to user communities using a variety of technologies. The collections of information can be scientific, business or personal data, and can be represented as digital text, image, audio, video, or other media. This information can be digitised paper or born digital material and the services offered on such information can be varied, ranging from content operations to rights management and can be offered to individuals or user communities. Internet access has resulted in digital libraries that are increasingly used by diverse communities for diverse purposes, and in which sharing and collaboration have become important social elements.

As digital libraries become commonplace, as their contents and services become more varied, and as their patrons become more experienced with computer technology, people expect more sophisticated services from their digital libraries. A traditional search function is normally an integral part of any digital library, but users' frustrations with this increase as their needs become more complex and as the volume of information managed by digital libraries increases. Digital libraries must move from being passive, with little adaptation to their users, to being more proactive in offering and tailoring information for individuals and communities, and in supporting community efforts to capture, structure, and share knowledge. Digital libraries that are not personalised for individuals and/or communities will be seen as defaulting on their obligation to offer the best service possible. Just as people patronize stores in which they and their preferences are known, and their needs anticipated, so too will they patronize digital libraries that remember them and anticipate their needs.

The *DELOS/NSF Working Group on Personalisation and Recommender Systems for Digital Libraries* was organized to bring together specialists from differing backgrounds who have vested interests in the future development of these systems. The explicit goal of the working group was to identify the key research challenges in this area and to develop a series of recommendations and research priorities.

The Working Group defines personalisation as the ways in which information and services can be tailored to match the unique and specific needs of an individual or a community. This is achieved by adapting presentation, content, and/or services based on a person's task, background, history, device, information needs, location, etc., essentially the user's *context*. Recommender systems are a particular type of personalisation that learn about a person's needs and then proactively identify and recommend information that meets those needs. Recommender systems are especially useful when they identify information a person was previously unaware of. Personalisation can be user-driven which involves a user directly invoking and supporting the personalisation process by providing explicit input. Examples of this include systems like MyYahoo! and MovieLens where the user explicitly initiates actions and provides example information in order to control the personalisation. Personalisation can also be completely automatic, where the system observes some user activity and identifies the input used to tailor some aspect of the system in a personalised way. These two examples of user-driven and automatic personalisation are at the extreme ends of the spectrum and many personalisation tools will have elements of both approaches.

Personalisation systems have had great success in other areas. For example, in the area of targeted advertising we see tailored advertisements in the output pages from almost all web search engines. When using online retail systems such as Amazon.com we are given suggestions for additional complementary services and products. When using a WAP handheld device, the presentation arrangement of menu options is personalised and tailored for different users. Adaptive hypermedia systems demonstrate how personalisation can be used to assist a person in navigating through large, online courseware systems.

If we consider the history of digital libraries we should be conscious that substantial digital libraries have been in place and operating long before the term "digital library" became popular in the early to mid 1990s. These include commercial systems as well as university and government systems but thus far personalisation has had only a limited impact on digital libraries. The initial focus in digital library research was on increasing the availability of digital content, and on creating and rolling out basic digital library services. Much digital library research and development activity to date has concentrated on the complete digitisation process, covering things like metadata standards, interoperability and rights management. Many applications of personalisation in digital libraries such as MyLibrary have thus far focused on applying basic personalisation and rudimentary recommender systems in a reasonably straightforward way and these applications of personalisation do not really add much value to the digital library and certainly do not bring the digital library up to the next level.

Many fields will contribute to the development of personalisation in the area of digital libraries. These include information retrieval, human-computer interaction, computer supported collaborative work, machine learning, user modelling, hypermedia, and information science, to name a few. To date, these fields do not have a great history of collaboration or of working together. This needs to change as in order to realise the potential of digital libraries we need to incorporate personalisation in a major way, and in order to develop research in the area of personalisation we need to bring together the many multi-disciplinary fields which contribute to its development. This report urges such a coming together.

This report is structured in the following way. In this section we have presented some background on our scope and definition of what we mean by personalisation and recommender systems. We now follow with an outline of our vision for the evolution of digital libraries and our perspective on personalisation in such libraries. That is followed by a series of research challenges which we believe need to be met in order to realise the true potential of digital libraries. Finally, we conclude with a series of specific recommendations and research priorities.

## **1. Vision of Personalization and Digital libraries**

A digital library is defined as a set of collections, services, a user community, and supporting technology. Formal large-scale research programs on digital libraries began about a decade ago with the initial NSF/DARPA Digital Libraries research program and a number of digital libraries were created as a result. Much of the research during the initial stages was on digitizing existing sources, creating large-scale collections, technological solutions, and providing simple forms of access. The first generation of digital libraries derived from this research provided a small set of services to relatively well-prepared and knowledgeable user communities.

The emerging generation of digital libraries is more heterogeneous along several dimensions. The collections themselves are becoming more heterogeneous, in terms of their creators, content, media, and communities served. The range of library types is expanding to include long-term personal digital libraries, as well as digital libraries that serve specific organizations, educational needs, and cultural heritage and that vary in their reliability, authority, recency, and quality. The user communities are becoming heterogeneous in terms of their interests, backgrounds, and skill levels, ranging from novices to experts in a specific subject area. The growing diversity of digital libraries, the communities accessing them, and how the information is used requires that the next generation of digital libraries be more effective at providing information that is tailored to a person's background knowledge, skills, tasks, and intended use of the information.

As computers have become common business, educational, and personal tools, long-term personal digital libraries are becoming commonplace. Children begin using computers regularly for education and entertainment by the time they are ten, and will continue to use them, for an increasingly complex set of tasks, throughout their lives. The information that a person accumulates during a lifetime of computer usage is a personal digital library, which makes *everyone* both a user and a creator of digital libraries. People

will want to save their pictures, music, educational and professional materials, personal and other information throughout their lives, but their needs, abilities, and computing platforms will change. People will need personal digital libraries that help integrate information gathered and organized by the ten year old with information gathered and organized by that same person at 20, 30, 40, and beyond. Long-term modeling of a person's evolving interests, preferences, knowledge, goals, and social networks will be required to help people manage their personal digital libraries during a lifetime of use, and this information must transcend specific systems, which will change often. The information that a person acquires during a lifetime, how the person organizes it, the tasks for which it is used, and the people with whom it is shared, paint a detailed picture of a person, but little is known today about how to use this information effectively. Interpreting the trails of a lifetime of computer usage, across the many different tools and resources involved, is an extraordinarily challenging and complex problem.

Digital libraries will also be affected by a trend towards mobile devices that have computational power similar to that of desktop machines, are wirelessly net-connected, and have builtin positioning systems such as GPS. In ubiquitous computing (ubicomp), access to information sources such as digital libraries is not only possible in many locations, but is dynamically adaptive with regard to a person's location as well as the artifacts and interaction devices and other people in those locations. For example, when visiting a museum, one might automatically be presented with information selected from the museum's digital library as one moves through the galleries. The user's motion towards an exhibit would be taken as an implicit cue for retrieval and recommendation of information, based on the exhibit and the previous activity and preferences of the user. Moving out of the museum to explore the surrounding city might trigger access to other information resources in a way responsive to the user's recent museum visit. Ubiomp's adaptation to the shifting context of use, in order to personalize and contextualise information access, has a side effect of binding together multiple digital libraries. Adaptive ubiquitous computing demands the integration of information that is heterogeneous with regard to ownership by or containment in different digital libraries.

The ubiquity of computing and telecommunications devices also means that communication and sharing of information is afforded to a large proportion of the public, and hence that digital libraries become less centralized and controlled. Greater community sharing of and interaction through digital libraries is another trend that must be addressed, and this trend makes its presence felt at both technological and social levels. With regard to technology, sharing and access can be carried out not just through institutional digital libraries but through large and dynamically changing 'peer to peer' networks. Current digital libraries have just begun to address the issues of provenance, subjectivity and consistency that come to the fore here. These issues have both negative and positive aspects. For example, having many such sources of information may lead to complex heterogeneity and inconsistency, but also innovation and personal contribution to the shared information resource. With regard to social aspects of sharing, there is great potential for community building and interpersonal interaction. Both institutional and community digital libraries can serve as meeting places where people can communicate with each other through the documents, annotations and logs they make available to each

other, and through the conversation and discussion around this shared information. Again, there are both negative and positive aspects to consider, balancing invasiveness and privacy with sharing and collaboration. As with all forms of social communication, the same contribution may be considered as useful and novel by one person, and as annoying and offensive by another. Since personalization involves not just the isolated individual, but the individual as a social actor, digital library research will have to be both socially mature and technically innovative as it steps up to play its part in the wider environment of public discourse, community and culture.

### ***1.1 The Future of Personalization in Digital Libraries***

The first generation of digital libraries were created for people whose information needs are well-defined and well-matched to the digital resources they contain. They assume relatively homogenous and possibly well-informed users, and relatively accurate descriptions of information needs. These characteristics limit their impact on wider society.

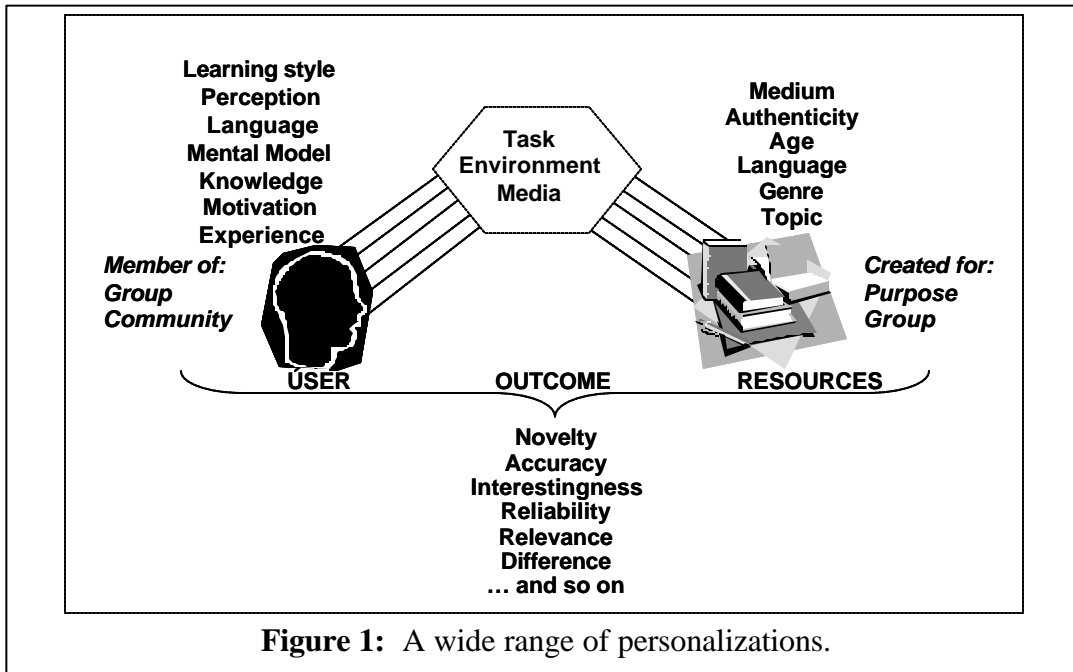
Personalization is required to make an increasingly heterogeneous population of digital libraries accessible to an increasingly heterogeneous population of users. It is no longer realistic to expect every user to adapt to every digital library. If a person must be an anthropologist to use an anthropological digital library, the library is available to only a limited community; if the library can tailor its services and materials for a wider range of users, the impact and utility of the library is magnified greatly. The next generation of digital libraries must support a wide range of personalized services that support the activities of a wide range of users.

Early research on digital library personalization used simple models of user interests to make individual recommendations. Future digital libraries need to feature broad user models, including a person's background, knowledge, tasks, social activity, and preferences. Moreover ubiquitous computing requires digital libraries to adapt to various parameters related with the context of a person's work. Finally, the need to support communities of users requires extending individual user models with group and community models.

### ***1.2 A Wide Range of Personalizations***

Digital libraries can be personalized in many different ways to support many different purposes and types of people, and many types of tasks. As illustrated in Figure 1, the personalization can be based on different types of characteristics such as characteristics of:

- ?? a person as both an individual or member of a group (e.g., knowledge or motivation);
- ?? the resources or information or documents (e.g., genre of the materials, age or authenticity; and/or



?? perceived outcome (e.g., novelty and accuracy), all of which are related to the media or channel used (e.g., PDA, versus a cell phone or computer), the task that is being performed, and the environment in which the user is immersed

This list of characteristics emphasizes short-term personalization and is not comprehensive, but it highlights the relationships between the most important components namely people, resources and perceived outcomes, and serves as a guide to illustrate the rich types of data that are available and need to be manipulated to personalize and/or recommend.

### 1.3 Potential Applications

Digital libraries that support a broader range of information seeking activities, build detailed models of users and user communities, and can tailor information for a wide range of uses will enable new types of software applications designed to support a variety of information seeking, building, and sharing activities.

Information-seeking activities extend well beyond the classic ad-hoc search that is the main access method in the current generation of digital libraries. A few recent examples show that information services can adaptively support diverse information-seeking activities. Writing aids automatically suggest related and supporting materials from personal or external digital libraries. Peer-help systems use information about the tasks and knowledge of individuals to suggest collaborators with specific skills. Adaptive hypermedia systems guide students towards the most relevant items in an educational digital library.

In the future it will be routine for applications to draw upon and integrate materials from multiple digital libraries and to use long-term user histories to help personalize this

material; such systems are beginning to emerge, but the difficulty of integrating material from multiple sources and utilizing long-term user history for effective personalisation makes them expensive to build. Digital libraries that explicitly tailor delivered information for specific uses will simplify such integration as well as support new types of applications. A tutoring system for English as a second language or that provides reading practice designed to address specific reading comprehension problems based on a long-term model of a user's language learning will need digital libraries that can provide materials that satisfy very specific and detailed user requirements. Integrating government digital libraries with citizen discussion groups will support more informed debate about public policy, especially if the evolution of that debate is incorporated into the personalisation. To be effective, government digital libraries will also need to bridge the gap between the language of administrators and bureaucrats and the language of ordinary citizens. Lifelong learning services will take a specific information need, interpret it in the context of a person's user model as it changes over time, and create a personalized learning plan that spans multiple digital libraries. These, however, are just examples and there are many other potential examples of cross-digital library personalisation.

## **2. Research Challenges**

### **2.1 Modeling Users**

To date personalisation has been inhibited by limited user modeling that reflects an over simplistic representation of users and their information seeking behaviour. Current user models draw on a limited set of parameters but people, jobs, and workplaces are much more complex.

More realistic user models should take into account the overall information space – the context – including:

- ?? cognitive abilities, e.g. learning styles, perception;
- ?? individual differences, e.g. experience, education, age, gender;
- ?? individual and group behavior patterns and history;
- ?? subject domains, e.g. engineering, arts, health;
- ?? work tasks, e.g. writing an essay, choosing a movie, planning a holiday;
- ?? work environments, e.g. university, hospital, business office, home; and
- ?? how all of the above change over time.

Information seeking encompasses elements of all the above and for personalisation some or all of these elements will come into play. Furthermore individuals are members of different types of social groups, forming information communities, which adds to the complexity of model building.

Currently both explicit and implicit methods of learning about users have been used including explicit questionnaires and implicit transaction logs, but new techniques for data collecting and analysis need to be developed for building more useful long-term user models. The challenges in building user models are multidimensional. Fundamental

questions need to be addressed, such as:

- ?? what data can and should be collected;
- ?? how can the data be captured;
- ?? how should the data be analysed;
- ?? what parameters need to be set;
- ?? how is anomalous data recognized and filtered out; and
- ?? how is data weighted appropriately over time?

Once these questions have been addressed more meaningful and appropriate user models must be developed to better inform the application of personalisation to information tasks and environments relating to digital libraries.

In addition the user models will need to be flexible and dynamic because the information elements listed above will change in terms of time and space and thus the models will need to evolve accordingly. For example we envisage access to digital libraries to become ubiquitous, mobile, portable and adaptable. Moreover the convergence of an individual's personal information space with the global information space will drive the creation of truly personal digital libraries. Finally, user models must include features of the community aspect of human behavior and preferences, in which memberships in social and work groups can influence a person's needs and requirements.

A particular challenge for personalization research is that long-term user models must encompass a timespan that is defined in terms of a human lifetime. This need defines a type of research and experimental evaluation that has not been done before in Computer Science and related disciplines.

## ***2.2 Making Recommendations & Doing Personalization***

Current personalization and recommendation techniques are based on relatively simple models. Pervasive personalization and recommendations in digital libraries require research on a range of topics that current systems only begin to address, for example making distinctions between ephemeral and persistent characteristics and requirements to support both long-term and short-duration personalization and recommendations. Incremental improvement in existing algorithms will not achieve these goals. Basic research is required on algorithms for personalization and recommendation that go beyond current similarity-based accuracy to address issues such as confidence, privacy, shilling resistance, authority, reputation, trust, novelty, recency, and utility.

Recent research on recommender systems focused on server-based systems that make recommendations based on the activities and preferences of large groups of people. Server-based personalization is natural for building models of user groups and communities, and can be tightly integrated with the content and services a digital library offers. Server-based personalization is also sometimes preferred in commercial environments because it can be used to bind customers to the service and switching services may mean losing one's personalizations or user models. Client-based



personalization is natural for building a detailed model of an individual over a variety of tasks and transactions and over a lifetime of use, and it gives people greater control over how and what personal information is revealed. The training data available at the client differs significantly from what is available at the server, for example requiring the client to understand much more of the semantics of user interaction with various information services and resources. Server-based and client-based personalization use different techniques, rely on different amounts and types of data, and may be studied by different research communities. One of the important challenges in this research area is bridging the gap between these two extremes, to develop portable server-based user models, and hybrid models.

The balance between user-specific and community-based personalization for an individual and a particular resource or task will vary. The first time a person encounters a digital library (“cold start”), personalization can be accomplished by relying on commonalities between the individual’s library-independent model and similar individuals that have interacted with the digital library in the past. Over time, as the individual has more experience with the digital library, the balance between user-specific and community-based personalization will shift and the contribution that the individual makes to the community and the community model will increase.

Sophisticated personalization requires more sophisticated control strategies. Time-sensitive personalization requires an understanding and adaptation to the timespan of an individual’s information needs (short term, long-term), and appropriate convergence mechanisms. Convergence must be balanced by an ability to adapt to an individual’s changing preferences, knowledge and abilities over time.

Traditionally many of these topics have been studied using online experiments in operational digital libraries, because they require interaction with large user communities and detailed information about user preferences or histories. This research methodology is effective, but it is also expensive and a barrier to entry for new research groups. There is a strong need for greatly improved simulation and modeling capabilities, to reduce research costs and enlarge the community of people who can study these topics.

### **2.3 User Interaction**

Unlike the way that user/system interaction is traditionally interpreted, where each party has a fixed role to play, in personalisation the system should promote a more flexible mixed initiative approach which would allow for the integration of human and automated reasoning for more in-depth interactivity. To date most recommender systems have been designed to implement a very simple model of human-machine interaction. Mixed initiative systems adopt more flexible approaches to recommendation and feedback. In addition to communication, interaction is concerned with the presentation and representation of information in all of its forms. The research challenge is how to design and facilitate personalised user interactivity.

Interaction should accommodate the variability in the way users undertake different tasks within a myriad of work situations. Personalisation should allow for systems to adapt to users whilst enabling an appropriate degree of user control. Hence a second research challenge in applying personalisation to user interaction is achieving this balance. Both challenges cannot be met without drawing on robust user models.

## **2.4 Evaluation**

Personalisation raises new evaluation issues and standard traditional approaches are inadequate. First, there is the need to assess personalisation from the perspective of the individual, the individual within a group as well as the group or community as a whole. Second, because of space and time dimensions, longitudinal studies must be conducted; some of these studies will need to span very long timespans, which will require long-term funding commitments. Third, user-centered quantitative and qualitative evaluations will need to be undertaken in both live and laboratory settings depending on the research objectives. Fourth, the design of evaluative studies will need to identify appropriate criteria and metrics for defining success that extend beyond and complement current measures of performance.

A central research challenge in the evaluation of personalisation – particularly in short to medium term horizons – is to build a suitable platform for evaluating personalised information seeking. This would contain rich data sets for training and comparative testing, standard tasks and scenarios, open source software for applying standard algorithms and services for conducting both laboratory and live evaluations.

## **2.5 Social Effects**

As digital libraries become more common and numerous, and as digital library support for social communication and sharing of information becomes routine, digital libraries both affect and are affected by social interaction, and thus must consider the dynamics of social settings.

Personalization means that each person's experience of the digital library will vary from that of others. The library will produce different experiences, and possibly different answers, for each individual, thus reducing the common experiences and shared references that bind the community together, and increasing problems associated with transparency, divergent interpretations, and training. Even without direct social interaction, a social effect can thus arise from personalization. More direct communication, such as recommendation and annotation, also has social effects. Sharing information with others creates possibilities for discovery, reinterpretation and discourse. An individual may contribute to a digital library and its community not only traditionally, as an author, but as a source of recommendations and annotations. Recalling the 'pathfinders' envisioned in Vannevar Bush's *As We May Think*, a digital library user may become increasingly significant to others as his or her personalized interaction with the digital library is made persistent or public. Other people may not wish to take on such a role, preferring to have less information about them made available to others.

A central and complex challenge for digital library research is the balance between privacy and collaboration, an issue that is familiar in other fields. The past treatment of privacy in both traditional academic, public and research libraries and in many experimental and commercial digital library systems has been overly simplistic. Most have chosen to protect privacy at all costs, not even offering users choices that might allow the benefits of sharing information with other users. A few have promoted sharing using privacy methods such as pseudonyms that are easily compromised.

Privacy protection does not mean imposing crude barriers that stop an individual from interacting with people he or she might benefit from. The public has demonstrated repeatedly, in settings as diverse as online commerce services and supermarket value cards, that it is often willing to give up a degree of privacy in exchange for a specific benefit. The challenge for digital library developers and researchers is to protect people's most essential privacy while also ensuring that desirable social effects are supported. Privacy solutions must allow people to shape and control how they present themselves to others, which requires that the solutions be comprehensible and based on informed consent. It also means helping people understand that any sharing of information can bring benefits and losses.

Privacy protection is a societal issue that spans many fields. One example is research in Computer-Supported Collaborative Work on design for privacy in ubiquitous computing environments that offers basic guidelines for control over what information is released about an individual as well as feedback about who has accessed which information. Feedback affords better understanding of how to adjust the controls over one's presentation to others, and how to adjust one's behavior given the available controls. Privacy research in digital libraries will necessarily be influenced by privacy research in other fields.

In summary, the challenge is to look beyond purely technical approaches to privacy and collaboration, and beyond purely social ones. Instead, we should start from the assumption that users will adapt to and control digital library systems just as much as digital libraries adapt to and influence their users, and we should look towards ways to support this larger system of control, feedback and adaptation.

### **3. Recommendations and Priorities for Future Research**

Our recommendations and priorities for future research follow the research challenges identified earlier and are divided into five major areas.

#### **3.1 User Modeling**

In order to enable a greater range of personalisations which range over more heterogeneous data, over short and long-term, and which cover input from multiple digital libraries including personal digital libraries, more needs to be known about users, user communities and their tasks. Greater emphasis needs to be placed on investigating methods for building more robust, flexible and portable models of the complexity of users, tasks and contexts to inform the diverse possibilities for personalisation in digital libraries. This implies a need to support interdisciplinary collaborative research from the different research communities including: HCI, CSWC, IR and others. Targets for this work include being able to develop implicit rather than explicit methods for learning user preferences which form the user models and developing user models that are portable across applications, devices and digital libraries. Perhaps the biggest challenge in this area will involve the development of user models that will drive personalisation and recommender systems, that are rich enough to capture as much of the user's task environment (context, task, situation), history, contribution to communities and individual preferences, as possible while conforming to a person's privacy choices. Such models will need to exploit and use the rich data made available from, for example, personal digital libraries.

#### **3.2 Personalisation**

The development of more sophisticated and complete models of users needs and behaviors opens up an opportunity for the development of more elaborate and sophisticated techniques for personalisation and recommender systems that more accurately capture the demands of the real world and that range over external as well as personal digital libraries. Aspects of personalisation and recommender systems that need to be researched, developed and tested include the differences between ephemeral and persistent needs, long term versus short duration requirements, hybrid client- server personalisation architectures and supporting a balance between community based and user specific personalisation. Basic research is required to go beyond the current similarity-based measurement as the building block for personalisation and recommender systems. Finally, the personalisation process itself should be open and transparent to users, and form part of their model of what the system is supposed to do.

#### **3.3. User Interaction**

In order to provide effective and diverse forms of personalisation the focus must be on the design of the interaction per se as an integral part of the whole system. In much of the research in areas related to digital libraries user interaction has been seen as an

afterthought, or something that is bolted onto a system last.. There is a need to develop multi modal mixed initiative interfaces that draw on a range of user information seeking models, those same models which we have earlier indicated will need to be enriched. The requirement is thus for research to develop theories of interaction which underpin the design of applications and vice versa and which go beyond issues of elicitation, presentation and feedback.

### **3.4 Evaluation**

User-centered evaluation must become an inherent part of system design and the evaluation of new technologies. New methods and evaluation criteria are required to assess personalisation systems in a cost effective way. The imperative is to develop evaluation methodologies, and make standard resources and tools more readily available for system developers. Evaluation based on quantitatively measuring system performance will remain important but should not be as dominant as it is now.

A serious challenge to research is the need for a large existing infrastructure of software, content, and committed users with which to do evaluations. Research progress would be improved considerably by a set of large-scale operational digital libraries in which any qualified researcher could conduct experiments “in the real world”. Such shared research infrastructure might be viewed as the equivalent of the particle accelerators used for Physics research. In practice a shared research infrastructure might be digital libraries created for the dual purposes of research while also serving some community, or it might be some form of access to existing commercial or non-profit digital libraries. A set of shared digital libraries would dramatically lower the barrier to entry in this research area, and their costs would be amortized across a larger research community.

Research on personal digital libraries, which is nascent, must accelerate quickly. In less than a decade many projections suggest the average home computer user will have sufficient disk space to store full motion video of every moment of a person’s life, from cradle to grave. Some people no longer delete email; soon they won’t delete anything else, either. There is a long list of interesting research to be done on personal digital libraries, but perhaps the biggest challenge will be evaluation. Studies of how a person’s use of a personal digital library evolves over time will need to be *very* long-term very multi-disciplinary.

### **3.5 Social Effects**

Social interaction is a feature of large-scale digital libraries that distinguishes them from most other computing environments. Explicit and implicit recommendations and sharing of information, preferences, and experiences exposes a range of social issues that are rarely faced in Computer Science. The most serious among these is privacy, a problem whose solutions are as much a matter of social policy as technology. A particular challenge is to develop stronger and more varied forms of privacy protection while

supporting the collaboration and sharing of information that has come to characterize many popular digital libraries.

Computers were once viewed as isolating people from people. Now many digital libraries play an important social role in forming and strengthening communities of people. Recommendation systems, which are based on sharing of information, and personalization, which recognizes an individual's specific needs, clearly play an important role in community development. However there has been little study of the social dynamics of such communities, the roles people play within them, how their members interact, and how they evolve over time.

Digital libraries are a forum in which to study a wide range of technology, social, and policy issues at the intersection of Computer Science and the Social Sciences. Progress on the issues described in this report requires collaboration among researchers from a variety of disciplines.