Developing Personalisation for the Information Environment
Project 1

Personalisation and Web 2.0

Final Report

Prepared by

Mark van Harmelen
Personal Learning Environments Ltd
and
School of Computer Science, University of Manchester
Executive Summary

Introduction

The Developing Personalisation for the Information Environment (DPIE) consists of two projects that are investigating ways of increasing the usability and utility of the Information Environment (IE) and some of its constituent services. Both projects are concerned with the topic of end user personalisation for users of IE services, but with these differences:

- DPIE 1 is also concerned with personalisation and Web 2.0 services, and their integration with the IE.
- DPIE 2 is concerned with the ways in which infrastructure established to support the UK Access Management Federation could support personalisation of JISC services, and any potential privacy and legal barriers to use.

DPIE 1 approached integration of personalisation and Web 2.0 services with the IE from two complementary perspectives:

- How IE services may gain a presence in Web 2.0 environments, particularly web start pages and social networking systems. This turned out to be a topic that Edina and Mimas were already addressing, and it was therefore de-emphasised in the project.
- How personalisation and Web 2.0 functionality might be integrated into the IE to inter-work with existing IE facilities and services. The project produced various reports on these topics, including an Edina and Mimas service survey using a survey tool. The more important of the project reports and documents may be found at misc.jisc.ac.uk/dpie1/deliverables. The project also extended its work to proposing an extension to the IE architecture, and prototyped some of those extensions in a publicly accessible demonstrator.

Areas of investigation

Several overlapping topics form DPIE 1’s focus. They are the IE and its intersection with Web 2.0, Library 2.0, Education / Learning / e-Learning 2.0, and personalisation.

The project emphasis was on enabling new use of the IE by exploiting Web 2.0, social software and personalisation technologies in order to provide a more personalised and collaborative use of parts of the IE, particularly those currently concerned with the supply of information to users. This is particularly pressing in the light of contemporaneous web developments.

Strategic opportunities

A major strategic opportunity for the IE is to enhance the IE to enable:

- **Architectures of participation**, which are socio-technical systems where the efforts of many users benefits individual users. At a technical level, architectures of participation are supported by social software.
- **Support for communities** of learners, teachers and researchers.
- **Web 2.0 as a platform** for learning, teaching and research.
- The transformation of bibliographic, serial and archival data in Edina, Mimas and other services into **social objects**[^8]: objects around which conversations and discussions can form as a result of user engagement with each other over the object’s contents. Flickr makes photographs available as social objects; similarly bibliographic, similarly serial and archival data may be made available as social objects.
- Provision of other **Library 2.0 facilities** around repositories and services which supply bibliographic, serial and archival services.

[^8]: [Social Objects](http://example.com/socialobjects)
Any adoption of these will result in a transformation in how IE users employ the IE in their learning, teaching and research activities.

A second major strategic opportunity is to deliver

- *Personalised information and service functionality* that is relevant, timely, engaging, and easy to consume and use, and that satisfies the information and task-support needs of the end users.

The benefit of successful delivery of these is the provision of facilities which provide greater effectiveness in the use of the IE, assisting the work of learners, teachers and researchers who use the IE.

The question of whether these services would be used by IE users is open, but the benefits of adoption are large. Therefore service implementations should be of exceptionally high quality and usability in order to maximise uptake (e.g. provision of a superb user experience, provision of personalisation that truly assists). There is a possibility of minimising the JISC's subsequent delivery costs by provision of services that can be rolled-out both in the IE and, in part, by educational institutions which choose to do so.

The project makes various detailed recommendations to the JISC in the areas of IE policy, IE strategy and operationalisation, IE architecture and components, funding, and in relation to Edina and Mimas. These recommendations may be found throughout the project's final report, and are collated in chapter 10 of that document (starting on page 46).

We believe it is possible to achieve both relatively quick and longer-term gains in useful IE functionality in the areas of Web 2.0 and personalisation within the IE.

**A plan for IE investment**

Bearing in mind that the project makes more comprehensive recommendations for funding IE activities, we recommend a minimal programme of investment:

- Try out social objects.
- Provide within the IE social networking services to support communities for learning, teaching, and research, such that the services are integrated with and augment IE resources and institutional library systems. Whatever is provided, it is unlikely to be an existing commercial product (e.g. by utilising Facebook applications).
- Augment the IE with synchronous and asynchronous communications facilities that may be used on their own and reused as components in other IE systems, including social networking systems. Wherever possible, make use of facilities that have already been developed and widely adopted.
- Build in personalisation data gathering services that may be used by end-user services to automatically store personalisation data for users. User opt-in and, possibly, opt-out is required.
- Integrate any personalised recommendation technology that is available in the near time frame, and that integrates into the IE.
- Experiment with personalisation for search, recommendation and the delivery of personalised functionality, both by making any easy gains in personalisation (e.g. from segmentation) that may be possible, and by engaging in longer term developments (e.g. by using personalisation data that is automatically collected, and by personalisation delivered by the use of semantic technologies).
Contents

Executive Summary.................................................................................. iii

Acknowledgements............................................................................... i

Contents........................................................................................................ v

Figures........................................................................................................ vii

1 Introduction .............................................................................................. 1

2 Project deliverables ................................................................................ 2

3 Background.............................................................................................. 3

3.1 Introduction ........................................................................................... 3

3.2 Customisation and personalisation ...................................................... 3

3.3 Web 2.0 and its use in the IE ................................................................. 3

3.4 Web 2.0 and media and technology convergence ............................ 5

3.5 Web-based activities ............................................................................ 6

4 IE and Web 2.0 integration .................................................................... 8

4.1 Contextualising integration activities ............................................... 8

4.1.1 Community ...................................................................................... 8

4.1.2 Learning and research ................................................................. 10

4.1.3 Work ............................................................................................. 11

4.2 Library 2.0 ........................................................................................ 12

5 Related projects .................................................................................... 14

5.1 DPIE 2 ................................................................................................ 14

5.1.1 Introduction ................................................................................... 14

5.1.2 Personalisation ............................................................................ 14

5.1.3 Outputs of the scenario work ...................................................... 16

5.2 The Researcher of the Future Project ............................................. 17

5.2.1 The overview level ...................................................................... 17

5.2.2 User and usage data ................................................................. 17

5.2.3 Users’ levels of information literacy ....................................... 18

5.2.4 Barriers to Use ........................................................................ 18

5.2.5 Semantic Technology ............................................................... 19

5.2.6 Design ......................................................................................... 19

5.3 Related projects ............................................................................... 20

6 Search, recommendation and personalisation .................................... 21

6.1 Introduction ....................................................................................... 21

6.2 Personalisation in commercial systems ....................................... 22

6.2.1 Segmentation ............................................................................ 22

6.2.2 Behavioural targeting ............................................................... 22

6.3 Attention and attention data .......................................................... 24

6.4 Mass recommendation .................................................................... 26

6.5 Personalised search ........................................................................ 26

6.6 Implications for the IE ................................................................. 27

7 Service survey ....................................................................................... 29

7.1 Survey assessment tool ................................................................. 29

7.2 Survey of Edina and Mimas constituent services ......................... 29

7.3 Findings and implications for the IE ............................................. 29
Figures

Figure 1: Web-based activities ................................................................. 6
Figure 2: A digital taxonomy based on Bloom’s Revised Taxonomy .................. 7
Figure 3: Situating Web 2.0 integration .................................................. 8
Figure 4: Most recent six-month growth in Ning page views ................................ 9
Figure 5: FeedForward’s depiction of design to delivery and subsequent use ........ 12
Figure 6: Behavioural targeting sales conversion process ............................. 22
Figure 7: IE architectural extensions ...................................................... 34
Figure 8: Composing social objects ....................................................... 36
Figure 9: Non-exclusive architectural choices: Many loosely coupled specialist systems and more integrated services ...................................................... 38
Figure 10: Compositor search results ..................................................... 40
Figure 11: Compositor generated bibliographic resource ............................. 41
Figure 12: Attention store data ............................................................... 42
Figure 13: Demonstrator architecture .................................................... 42
Figure 14: Demonstrator screens and transitions ....................................... 42
Figure 15: Compositor integration with a learning environment via social bookmark use ................................................................. 43
Figure 16: A user’s profile page .............................................................. 44
Acknowledgements

The DPIE 1 Project team was, for various parts of the work plan, composed of

- Santiago Chumbe (Heriot-Watt University)
- Hilary Dexter (University of Manchester)
- Roddy Macleod (Heriot-Watt University)
- Baptiste Manson (Inovia-Team)
- Malcolm Moffat (ex Heriot-Watt University)
- Mathieu Perrin (Inovia-Team)
- Eric Raffin (Inovia-Team)
- Mark van Harmelen (Personal Learning Environments Ltd and University of Manchester)
- Dave Workman (Independent Consultant)

Tom Franklin (Franklin Consulting) helped review this report.

Balviar Notay, our JISC Programme Manager, assisted us in substantial ways.
1 Introduction

The Development of Personalisation of the Information Environment (DPIE) consists of two projects that are investigating ways of increasing the usability and utility of the Information Environment (IE) and some of its constituent services.

Both projects are concerned with the topic of end user personalisation for users of IE services, but with these differences:

- DPIE 1 is also concerned with personalisation and Web 2.0 services, and their integration with the IE.
- DPIE 2 is concerned with the ways in which infrastructure established to support the UK Access Management Federation could support personalisation of JISC services, and any potential privacy and legal barriers to use.

DPIE 1 approaches integration of personalisation and Web 2.0 services with the IE from two complementary perspectives:

- How IE services may gain a presence in Web 2.0 environments, particularly web start pages, and social networking systems.
- How personalisation and Web 2.0 functionality might be integrated into the IE to inter-work with existing IE facilities and services.
- In addition DPIE 1 has produced a questionnaire based survey tool to reveal personalisation and Web 2.0 characteristics of IE services, and has, with service help, used that tool to survey the Archives Hub, Copac, Intute, Jorum, Suncat, and Zetoc.

The DPIE 1 Project recommends extensions to the IE. Uptake is not guaranteed, but, because of the potential payoffs in increasing IE utility and uptake, the risk of low uptake can be justified.

There are legal aspects to storing personal data for personalisation purposes. The JISC should consult appropriate legal authorities to determine the legal implications. What is recommended here is absolute transparency on storage and use of personalisation data in order to build reputation and ensure uptake.

**Recommendation 01:** It is recommended that any IE system holding data for personalisation purposes, must conform to U.K. and European law and it is recommended that the JISC should consult appropriate legal authorities to determine the legal implications for storing data for personalisation purposes.

**Recommendation 02:** It is recommended that any IE system (a) provide exceptionally secure storage of personalisation data in order to mitigate concerns about insecurity of personal data, and (b) deal with users and their personal data in a totally transparent way in order to build JISC and service reputation and thereby mitigate concerns about use of personalisation data.

In electronic copies of this document, the reference numbers that appear inside square brackets not only refer to the reference list at the end of the document, but are also, for web-based resources, hyperlinks that may be used to view the resource.
2 Project deliverables

In summary the project has:

- Performed desk research and produced reports on various topics and technologies relevant to personalisation and the integration of the IE with Web 2.0 faculties.
- Constructed a questionnaire-based survey and surveyed services.
- Designed an enhancement to the IE architecture to integrate Web 2.0 services and end user facilities with the IE, and prototyped that architecture.

Some of the activities of DPIE 1 were affected by the discovery that one of the major strands of project activity (service presence in Web 2.0 environments) is a largely solved problem. As a result of this the project had sufficient resources to implement an architectural proof of concept demonstrator.¹

Project deliverables, all available through misc.jisc.ac.uk/dpie1/deliverables are:

- A general discussion of topics in the scope of DPIE 1, in this final report.
- A set of recommendations to the JISC, in this final report.
- Recommendations for demonstrators in the context of a potential future ITT.
- A survey tool for IE services to use to provide data to assess their customisation and personalisation facilities, their status in the Information Environment Service Registry, and their approach to Web 2.0.
- Survey results for six Mimas and Edina services: Archives Hub, Copac, Intute, Jorum, Suncat, and Zetoc.
- Reports on technology areas relevant to the integration of the IE and Web 2.0, on:
  - Web 2.0 technologies for use in the IE for teaching and research, together with various live links.
  - Mechanisms (widgets and plugin applications that conform to published APIs) to allow a service presence in Web 2.0 environments.
  - Personalised recommendation ², which includes approaches to personalised search.
  - Semantic technologies for personalisation.
- Architectural extensions to the IE to integrate Web 2.0 and allied services for end-user use.
- The implementation of an architectural proof of concept demonstrator.
- A del.icio.us archive of about 200 web based resources [del.icio.us/tag/jisc-dpie] relating to this study.

¹ This was discussed with JISC representatives and the DPIE 1 Programme Manager late 2007, and a refocusing was agreed on. A change of plan was decided on with the Programme Manager assent on 31 January 2008.
² Personalised recommendation services produce recommendations for users on the basis of personalisation data; in this way they supply what is sometimes called ‘push’ technology.
3 Background

3.1 Introduction

In this chapter we discuss various topics that form the background to the DPIE 1 Project, and that set the field for subsequent discussion. The material here varies from definitions of customisation and personalisation, to an examination of the contemporary Web 2.0 environment, the actions that web users undertake, more specific factors impinging on the study, the similarities between learning and research, and the kinds of resource discovery tools that are emerging as part of the Library 2.0 movement.

3.2 Customisation and personalisation

We use with the definitions of customization and personalisation defined in Smith, Schmoller and Ferguson’s JISC report on Personalisation in presentation services [31].

That report: distinguishes between:

- **Customisation** where the users make explicit choices to customise their own experience of a computer system.
- **Adaptive Personalisation** where the availability of options, interface, access or functionality [and provision of other personalised results] is based upon knowledge about users gained from tracking user activity and/or other sources of user information. In this report we simply refer to adaptive personalisation as personalisation. 4

We consider that the purpose of personalisation is to deliver to end-users information and services that are relevant, timely, engaging, and easy to consume and/or use, and that satisfy the information and task-support needs of the end users.

3.3 Web 2.0 and its use in the IE

Franklin and van Harmelen describe Web 2.0 in the JISC study on Web 2.0 Content Sharing for Learning and Teaching in Higher Education [12]:

Web 2.0 encompasses a variety of different meanings that include an increased emphasis on user generated content, data and content sharing and collaborative effort, together with the use of various kinds of social software, new ways of interacting with web-based applications, and the use of the web as a platform for generating, re-purposing and consuming content.

With Web 2.0 the web becomes a platform for social software that enables groups of users to socialise, collaborate, and work with each other. This change of use is largely based on the use of existing web mechanisms to share content, the use of web protocol based interfaces to web applications that allow flexibility in reusing data, and the adoption of communications protocols and technologies that allow direct user-to-user content.

One way of characterising Web 2.0 and its use is by contrasting the former web (“Web 1.0”) with Web 2.0. In Web 1.0 a few content authors provided content for a wide audience of relatively passive readers. However, in Web 2.0 everyday users of the web use the web as a platform to generate, re-purpose, and consume shared content. Users may, by the means of social networking systems, form loose groupings of individuals and more tightly-knit communities that act towards realisation of common goals.

---

3 The square brackets contain our addition to the original definition.
4 We have not as yet had recourse to distinguish between the report’s notions of Adaptive Personalisation based on User Activity (APUA) and Adaptive Personalisation based On Data held elsewhere (APOD). In a service based world we presume much, if not all, personalisation is APOD.
In its best form Web 2.0 gives rise to architectures of participation. These are socio-technical systems where aggregated user effort is used to provide improvements of use by individual users.

Effective facilitation of the growth of architectures of participation is a key to the successful enhancement of the IE functionality with Web 2.0 systems. However, growth can not be guaranteed. Specifically we note that users will not flock to use social networking systems around IE services unless the social networking systems offer significant advantage(s) to their use.

Within the context of Web 2.0, DPIE 1 is concerned with two topics:

- How IE services may gain a presence in Web 2.0 environments, particularly web start pages, and social networking systems.
- How Web 2.0 functionality may be integrated into the IE to inter-work with and enhance existing IE facilities and services.

IE services presence in Web 2.0 environments is now well understood and implemented in several environments by Edina and Mimas services. The mechanism used is the construction of a widget that may be installed by an end user in a Web 2.0 environment. A combination of ease of implementation, a relatively small number of candidate environments, and the growth of widget standards means that widgets can be supplied for popular environments relatively easily. However, data about the deployment of existing widgets is currently unknown.

**Recommendation 03:** It is recommended that, where possible, services should gather usage information from their widgets. Besides providing basic usage data (number of installs and number of searches etc performed from the widget), wherever possible information should be collected to produce information that is useful in determining if further facilities are needed, and how they should be designed.

In the time available, one of us (Mark van Harmelen) conducted a small survey of 47 students to try to confirm if Facebook is students’ social networking service of choice. Students varied between 18 and 56 years of age, but were mostly of typical undergraduate age. Seven students were postgraduate at different levels of progression. The survey included Birkbeck, SOAS and UCL students, with one each at NYU in London, the Courtauld Institute, and QMW. The survey did not include students at new universities; music, art and design oriented institutions; and FE colleges.

40 students, 85% of the total sample, used a social networking service. All of those users used Facebook as their service of choice. Of these 40 students, 37 used only Facebook. MySpace was the second service used by the remaining three students. For each of those three students, the time they spent using MySpace was less or equal to 18% of the time they used both MySpace and Facebook.

The overwhelming result for the sample was that, for students who answered one of two questions by supplying a percentage, Facebook use seemed to appear very popular. The students were asked one of two questions “What percentage of university students in the UK use Facebook?” or “What is the percentage is of university students you know who use Facebook?” These questions were used interchangeably, and respondents were asked to guess if they hesitated, but not pressed to ultimately supply a percentage. Students responded with figures between 50% and 100%, with an average percentage of 87%.

Opinions varied on if Facebook was a suitable medium for academic content (timetables, notifications of assignment due dates, library search, whatever the students might find useful). Positive opinion included, at one extreme “it would be best thing ever!” and “Oh Yeah! I’d get hot dinners delivered by Facebook.” to, from the NYU in London student “limit Facebook, its used for procrastination, a dip in leads to one hour, two hours; total educational downfall.” One first year student had class representatives who made Facebook groups and entered reminders (and perhaps more). Other students used Facebook groups for revision and study group purposes. Several students wished to get notifications of special events, e.g. guest lectures, via Facebook.
Percentages for negative and positive views on academic content in Facebook are not that different:

<table>
<thead>
<tr>
<th>View</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative, intrusion</td>
<td>15</td>
<td>32%</td>
</tr>
<tr>
<td>neutral</td>
<td>6</td>
<td>13%</td>
</tr>
<tr>
<td>positive, welcomed</td>
<td>19</td>
<td>40%</td>
</tr>
<tr>
<td>not sure</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>not elicited</td>
<td>5</td>
<td>11%</td>
</tr>
</tbody>
</table>

Because of the small and unrepresentative sample results, should be interpreted with caution. However, the interviewer’s sense from general discussion around the questions, is that, at least for Russell Group universities, Facebook is overwhelmingly the social networking system of choice. Academic support supplied in Facebook would be supplied as a Facebook application or in a group, and students may choose to load academic applications and/or join groups or not. Privacy is a concern in Facebook, and application designers and authors should ensure that private data is not exposed to other applications or to people who should not see that data.

**Recommendation 04**: Possibly subject to the outcomes of recommendation 20, because Facebook appears to be many students’ social networking system of choice, it is recommended that appropriate IE services should, wherever possible, supply Facebook applications that interface with the services, but not expose private data to other Facebook applications. There should be careful treatment of what data is shown to users’ friends and group members.

Achieving Web 2.0 integration with the IE is the more challenging area for investigation and solution. The area includes these topics:

- Addition of Web 2.0 functionality and social software tools to the IE so as to allow the formation of architectures of participation and leveraging of user behaviour.
- Transformation of bibliographic, serial and archival records into social objects, around which users may form conversations and thereby generate or promulgate knowledge.
- Addition of learning and research environments which support learning, teaching and research, and which are integrated with IE-provided facilities. We expect that these facilities would be Web 2.0 and social networking system based; rather than being VLE based.
- Addition of personalised recommendation and search the IE. While some aspects of these are not strictly Web 2.0 topics, some technical factors result in us placing them in the Web 2.0 area.

These topics require architectural additions to the IE. A possible approach to these is described in chapter 8.

In order to provide a guide to what Web 2.0 facilities might be added to the IE, we performed a broad survey of Web 2.0 services and technologies and services, filtered them for potential applicability within the IE, and then described about 50 Web 2.0 system types and technologies. For each of these we provide a short description, hyperlinks to exemplars and additional resources, sample educational use, and potential use in the IE. These are described at misc.jisc.ac.uk/dpie1/docs/web2.0.html Several of the most promising systems and technologies are factored into architectural plans in chapter 8.

### 3.4 Web 2.0 and media and technology convergence

Although out of scope in this project, the full implications of the use of Web 2.0 in learning, teaching and research will eventually need to be viewed in the light of media and technology convergence.

The following factors drawn from the JISC study on Web 2.0 Content Sharing for Learning and Teaching in Higher Education [12] pertain:

- The contemporaneous growth of Web 2.0 co-occurs with increased media convergence, particularly in respect of broadband communications, telephony and the broadcast media.

---

5 We note the technological (rather than the user-centered approach) of this work, and elsewhere recommend complementary ethnographic studies of learning, teaching and research activities.
- While professionally produced and edited media are likely to persist, we will see the broadcast media increasingly adopting Web 2.0 technologies, with greater audience participation and audience created content. In parallel we will also see an increasing number of [communication] channels funded in very diverse ways.

- The increased bandwidth offered by 3G telephony will encourage a move from the desktop and the desktop browser to mobile devices and browsers. Content will be created, shared and consumed on mobile devices.

- Ubiquitous computing, computing that is always around us, and always on, will change our everyday digital and media environments, mediating the world in new ways.

- Indication of social presence will increase, and will help mediate between people in different ways, potentially changing aspects of how relationships between people are formed and maintained.

The need to stay abreast of developments in technology and technology convergence is underlined by the final report from the Scoping Study for the Pedagogy Strand of the JISC e-Learning Programme. This states:

To see e-learning from the learner's viewpoint, we must see technology in the broadest possible sense including perhaps the mobile phone, e-mail, instant messaging and PDAs.

The extent to which the five factors above will affect the IE is difficult to predict. However, we note the general principle that the IE is part of a computing milieu which will change as these trends are realised. The IE will have to respond to change in order to remain relevant for its users.

### 3.5 Web-based activities

We can consider that when learners, teachers and researchers use the web, they engage in common activities that can be described at a high-level in Figure 1, with, inevitably, overlap between activities.

- planning
- searching
- browsing
- evaluating, judging, filtering
- collecting
- abstracting, analyzing, classifying
- learning
- researching
- sense making, constructing understanding and meaning
- creating, expressing and applying knowledge
- helping, assisting, teaching and counseling
- rating, commenting, cross linking, embellishing
- re-mixing
- storing
- publishing, distributing
- communicating
- consuming
- socializing
- playing
- collaborating
- negotiating

**Figure 1: Web-based activities**

A further categorisation of web based activities is provided by a digital literacy oriented extension to Bloom's Revised Taxonomy.

---

6 This list was proposed in draft and subsequently edited by Mark van Harmelen from potential refinements offered by the Teaching and Learning Online group.

Figure 2: A digital taxonomy based on Bloom’s Revised Taxonomy

Many of the activities above can in some way be supported by combinations of technology and systems identified in our survey for their potential use in the IE. Nevertheless, there are still large advances to be made in Internet and IE support of the above activities for learners, teachers and researchers. What is needed is a strategy for the next stages of the JISC’s work in developing aspects of the IE. This strategy needs to consider activities above for teaching, learning and research, within a framework that acknowledges constant influences for change on the IE. Chapter 10 considers possible strategy in the context of personalisation, Web 2.0, Library 2.0 and search and recommendation.
4 IE and Web 2.0 integration

It is worthwhile to consider the end aims that have to be addressed when adding Web 2.0 facilities to the IE. Our view is that we wish to enhance the IE’s utility, usability and use. We particularly consider uses of the IE for information discovery in learning, teaching, research. To some smaller extent we are also interested in using the social networking facilities that we propose to support the activities of IE administration and development teams. We consider this aspect to be assumed in our discussions of community support.

4.1 Contextualising integration activities

One can only focus productively on Web 2.0 integration by considering how people interact and what activities they perform. We note similarities between learning and research, and propose the following inter-related areas as a means of providing context to the endeavour of updating the IE with Web 2.0 facilities to support learning and research, and through learning, teaching.

![Figure 3: Situating Web 2.0 integration](image)

Together, these three areas contextualise the introduction of Web 2.0 facilities and personalisation into the IE.

4.1.1 Community

Community has been a topic of sociological study for the last century. One could somewhat loosely define a community to be a group of individuals organized into a unit, where the individuals manifest, to a greater or lesser extent, a unifying trait or common interest. But we consider that a unifying trait or shared interest is not sufficient to define a community and that individuals in a community must have ties to each. Weak ties “are generally infrequently maintained, non-intimate connections, for example, between co-workers who share no joint tasks or friendship relations. Strong ties include combinations of intimacy, self-disclosure, provision of reciprocal services, frequent contact, and kinship, as between close friends or colleagues.” [Haythornthwaite, 13]. While both kinds of ties have advantages to communities, we assume that groups which only exhibit weak ties are just groups, and that for a group to be a community it needs both weak and strong ties between members.

Both learning and research are community activities. Learning is socially based (see section 4.1.2) and occurs inside communities, and often community members help each other to learn. Teachers too may be part of a learning community. We take as the norm that most productive learning occurs in communities8. Research, which can sometimes be a solitary activity, most often proceeds in a social context where researchers interact with each other and build communities to pursue their research aims.

---

8 Even individual learning is socially based and takes place in the context of a community because learning can only take place with socialisation, a common language, and socially constructed meaning.
Given the social nature of learning, teaching and research, community support via web services is a topic of national import.

Communities and less cohesive groups (which, in turn, may become communities) are already widespread on the Internet. While a downturn in U.S. social networking was reported in the last few months of last year [1], this is viewed as having been transitory. [25].

While these reports are about the large U.S. social networking sites, Ning [ww.ning.com/] is a much smaller site that allows users to set up their own social networking sites, customising them with features selected from pre-supplied options. Ning is experiencing a rise in use, and, although Ning may simply be increasingly popular with people trying out facilities and abandoning them, we believe this rise is an indicator of a growing trend that we see to set up smaller specialised social networking service based communities.

![Figure 4: Most recent six-month growth in Ning page views](image)

**Recommendation 05**: Because of the centrality of communities in learning, teaching and research, it is recommended that the JISC adopt as part of IE policy provision of tools to support community formation, growth and maintenance.

It is important to consider community formation in web-based spaces. We believe that the simplistic view of “build a social networking system and they will come” is mistaken. For a community to form there has to be real advantage in being a member of that (future) community.

Thus Derek Powazak [26] discusses the notion that communities will form around pre-supplied “great content”, and internet communities form around sites where high quality information can be obtained. The members then also add content to the site.

Jyri Engstrom [7] discusses one of the aids to community formation in social networking systems: Social objects are objects around which users can base conversations, and thereby start to strengthen links between themselves. One example is provided by photographs uploaded to Fickr.

**Recommendation 06**: It is recommended that the JISC establish a strand of activities to address issues of community formation and maintenance in academic settings, and the delivery of a variety of systems to support these. In any programme that is established, system development must be accompanied by evaluation of real use, and results must be collated and developed by a synthesis team.

Caroline Haythornthwaite [16] discusses a social networking perspective on how communities form: Membership of a social networking system provides latent links between the members. Communication via public channels transforms latent links into weak links. Communication on private channels transforms weak links into strong links that bind communities together. This finding has been almost totally ignored in community support software.

**Recommendation 07**: It is recommended that the JISC provide reusable synchronous and asynchronous recordable Web 2.0 communications tools for video and voice communication. This to support both meetings and communities. A suitable starting point may be provided by the FlashMeeting Project [fm.ea-tel.eu].

Sometimes there is no need for a community to form in order for architecture of participation advantages to occur. Thus, for example, mass recommendation systems [section 6.4] need not rely on a community. Equally, aspects of social bookmarking and the ability to find resources that others have tagged in a
certain way does not need a community to form, but is facilitated by community formation. At other times, for more sophisticated needs, a community must form.

4.1.2 Learning and research

Several perspectives inform our approach to learning and research and, by implication, teaching. We note these below; they in turn have implications for pedagogy and teaching practice. Largely, the perspectives also apply to research.

Social constructivism: Negotiation of (the understanding of) meaning and, thus, learning occurs as the result of social activity. We observe increasing numbers of students learning in groups while engaging in peer mentoring and peer teaching. We also see growing numbers of professionals joining and learning in communities such as the JISC Emerge to help each other further their own professional learning and research aims. There are other strands of social constructivism that are particularly learning-related, e.g. the Zone of Proximal Development. We do not consider these here.

Papert’s constructionism: The process of learning is particularly facilitated by learner production of public artefacts as exhibits. The process of production requires that learners must organise, assimilate and accommodate new concepts in order to produce the artefact.

New literacies and learning practices: John Seeley Brown points to the development of new literacies of web-based information navigation, assessment and selection, coupled with an increase in experiential and discovery-based learning through the manipulation of objects in a process of bricolage. These literacies and learning practices are increasingly used by students who have recently started entering FE and HE. However, as both the Researcher of the Future Project and our DPIE 1 Project note, students still lack some key digital literacies and metacognitive skills.

Recommendation 08: The JISC is aware of the need for digital and other literacies to help enable learners, teachers and researchers take full advantage of the Internet in their learning, teaching and research, and is already engaging in the area as shown by, e.g., the Informs Project. As part of this engagement, it is recommended that the JISC fund a programme of work in this area.

Informs can be found on the Intute web site.

Independent learning: This refers to learners organizing and taking responsibility for their learning. Here independence of action in learning (and in research) does not imply lone work; independence refers to self-determination in formulating action. Thus an independent learner, to a preferably greater extent, determines what and how s/he should learn.

We take independence of action as a given for research, although, of course, researchers at an early stage of development may need to learn become independent researchers, in the same way as some learners may need to learn to become independent learners.

We subscribe to a particular view that significant positive change will largely only come to the education system with an increase in the numbers of students learning independently, be those students learning on their own, or in groups of peer-supporting learners. Within this we take the view that the most profound change will come from learners taking responsibility for their own learning. This is in a sense close to a philosophy of ‘people doing it for themselves’ that also applies to Web 2.0.

We observe increasing use of innovative teaching methods in HE that are increasingly supported with Web 2.0 systems. As just one part of this movement, we note how student effort can be leveraged in architectures of participation so that they create their own learning materials.  

---

9 One successful development of this kind is provided by the Royal Veterinary College and other veterinary schools which have together expanded on the RVC’s earlier student-constructed pathology wiki, and have started to engage students in these multiple institutions to create thier own discipline-wide wiki-based learning materials (see WikiVet [w01.rvc.wiki.wf.ulcc.ac.uk]).
Exemplars

Exemplar services that support learning and teaching are provided at misc.jisc.ac.uk/dpie1/docs/web2.0.html This page lists systems and technology that might potentially be used within the IE.

Many of the services documented on the above web page also support research activities.

Exemplars of specialist social research services include

- academici [www.academici.com]
- allResearchers [www.onlineacademicadvisor.com]
- Elsevier’s 2collab [www.2collab.com]
- my Experiment [www.myexperiment.org]
- Pedestal [pedestal.lboro.ac.uk]
- Zotero 2.0 (still under development)

4.1.3 Work

By work, we mean the activities that learners, teachers and researchers engage in while learning, teaching and researching. While we are interested in all activities in pursuit of this work, we are particularly interested in activities that are common across all three areas, because addressing and supporting these common activities will result in widespread gains.

There are strong similarities between learning and research that we can exploit in developing personalisation and Web 2.0 facilities for the IE:

- Learning and research both rely on similar cognitive activities.
- There is therefore a continuum that may enable the use of common tools to support both learning and research.
- There are already many exemplars of tools that support both learning and research. In the context of this report many Web 2.0 systems have uses that span both learning and research.

Fortunately, the IE is well positioned to inter-work with Web 2.0 services.

The JISC funded FeedForward Project provides a highly relevant perspective on the user organisation of discovery to delivery activities in learning and research. Figure 5 below provides a visual means of understanding these activities in both learning and research.

Wilson [33] describes activities and their location related to the cycles in figure 5:

‘Starting from the outer ring: we have an iterative process of setting up and configuring tools and services in connection with the wider web (the cloud).

From there we shift into an inner process of selecting useful sources, pulling in information, and sometimes focussing in on particular groups of sources.

Then inside there we have a more intensive cycle of scanning sources; selecting items of interest; organising them in groups, lists, (topic/mind/freeform) maps and so on; reviewing what we have and making sense and patterns (possibly leading us back into the outer cycles); "writing" (or other creative acts); and publishing, leading us back out to the wider web. “
We are also interested in the nature of work in groups. Haythornthwaite [15] discusses knowledge flow in inter-disciplinary teams which share a common goal. Her results from examination of three projects “show exchange of factual knowledge to be only one of a number of learning relations that support the teams. Other important relations include learning the process of doing something, information about methods, engaging jointly in research, learning about technology, generation of new ideas, socialization into the profession, access to a network of contacts, and administration work.”

**Recommendation 09:** It is recommended that the JISC’s efforts in supporting the work of students, teachers, and researchers should include a programme that funds ethnographic and other sociological studies on user work in learning, teaching and research that reveals design data for future systems, and that draws on past work in Computer Supported Co-operative Work (CSCW) and any other applicable fields.

### 4.2 Library 2.0

While there is dissent on whether Library 2.0 should continue to be used as the name of a movement in the library community, it remains a useful label that represents putting library users at the center of libraries, enhancing the users’ experience of what a library is about, and bringing a greater level of service to them.

To define Library 2.0 for our purposes we simply supply two illustrative definitions. The first is about libraries and their users:

“Library 2.0 is all about library users -- keeping those we have while actively seeking those who do not currently use our services. It’s about embracing those ideas and technologies that can assist libraries in delivering services to these groups, and it’s about participation -- involving users in service creation and evaluation. Library 2.0 is an operating model that allows libraries to respond rapidly to market needs. This does not mean that we abandon our current users or our mission. It is a philosophy of rapid change, flexible organizational structures, new Web 2.0 tools and user participation that will put the library in a much stronger position, ready to efficiently and effectively
meet the needs of a larger user population." Michael Casey via the squidoo page on Library 2.0

And the second is techno-centric:

"the application of interactive, collaborative, and multi-media web-based technologies to web-based library services and collections" [24]

We add that Library 2.0 is inherently concerned with architectures of participation: Technology makes possible user contributions about library holdings, and those contributions benefit all the users of those libraries.

A major part of this project has concerned bibliographic, serial and archival services in the JISC National Data Centres Edina and Mimas. Some of these services are essentially variants on traditional Online Public Access Catalogues (OPACs). This project has as one of its conclusions that it is a worthwhile experiment to transform these OPAC-style records into user-centric content and communication resources (that we call social objects) in a Library 2.0 style.

Traditional OPACs are considered to have problems in a number of areas, including user engagement, usability, search, and flexibility. These problems are all the more clearly delineated because of the widespread use of modern web-based search engines, and Library 2.0’s emphasis on user-centricity and engagement.

OPAC search facilities are relatively basic: Improvements and transformations are needed. Two posts on the American Library Association (ALA) blog list 20 desirable search facilities that are generally missing from OPACs [27 28]. A third ALA post [29] is more discursive, and discusses higher-level topics that are almost always universally true:

- OPAC content is still about books and other physical objects, rather than other forms of content.
- OPACs are citation indexes and not an index into content.
- OPACs do not perform well as both inventory catalogues and discovery tools.
- Libraries replicate data sets, and this is at odds with what is more optimally possible now.11

This and related material on Next Generation Catalogues starts to raise discussion of library catalogues as pure data sources that allow Web 2.0 style remixability for presentation to and use by end users. See, e.g., Lorcan Dempsey [9] on this topic.

What is also clear is that the traditional OPAC style data services just don’t provide the kinds of social facilities that are starting to be offered by Library 2.0 systems.

Importantly, IE services contain key information for the scholarly activities of learning, teaching and research. It is of national importance that the services be enhanced so as to become more usable and useful, and that it is highly encouraging that the JISC and JISC-funded services are starting this process.

---

10 These are relevance ranking, stemming, field weighting, spell-checking, refining original queries, support for popular query operators, boolean operators for search queries, flexible default query processing, in-line query limiters, duplicate detection, sort flexibility, non-roman character sets, faceting [30], advanced search, easily customized search-result pages, human suggestions, search logging and reports of considerable sophistication, well-rounded administrative interfaces, full text search in citation indexes, and provision of these features well.

11 But we note the existence of union catalogues like WordCat.org and Talis Source.

12 A similar approach is recommended in our architectural suggestions to integrate IE services with Web 2.0 facilities and services.
5 Related projects

5.1 DPIE 2

5.1.1 Introduction

The DPIE 2 Project has as its focus personalisation, but mostly within the areas of identity management and authentication as mechanisms to support personalisation. DPIE 2 will end after DPIE 1, and information in this report is based on DPIE 2’s work to date. This includes a landscape study, and use of focus groups and scenario-building exercises to investigate possible futures for the IE. Future work involves developing use cases based on these three areas of work, and producing a final report that includes recommendations and specifications for demonstrators.

The DPIE 1 Project is generally agnostic on different identity and authorisation mechanisms, and presumes that the JISC and JISC funded projects will choose suitable mechanisms that inter-work successfully with the IE and Web 2.0 services within the IE. Factors that will influence choices are the already strong move to federated authentication mechanisms within the IE, and the increasingly common deployment of OpenID authentication on the broader web.

**Recommendation 10:** It is recommended that any access management system supported by the JISC must be capable of working seamlessly with OpenID, to maximise the possibilities for Web 2.0 interoperability.

The DPIE 2 Project was explicitly tasked with investigating the relevance of Federated Access Management for personalisation, and results will be inclined to that technology. 13

We note interim report and scenario output material that is of relevance to our project. We sometimes add material in brackets.

5.1.2 Personalisation

DPIE 2 proposes a “personalisation spectrum” that covers the following kinds of applications:

- **Personal applications** are of two categories: Firstly, applications that include data about a user and the users access to that data. DPIE 2 cites as examples applications that provide access to a user’s own staff or student records, access to their timetable, and an OPAC application that allows access to user compiled data such as might be held in the user area of an OPAC. Secondly, DPIE 2 puts into this category systems that provide information and functionality based on a user’s role, and thus include all role-based workflow and process management systems in this.

- **Portal applications** collect personal information from diverse sources and present that information in a users “personal area”.

- **Customised applications**, where as discussed above, a user explicitly alters or parameterises the behaviour [and appearance] of an application.

- **Adaptively personalised applications** change behaviour [and thus data and services which are supplied to the user] depending on the application behaviour that [attempts to] predict the needs of the user. In the DPIE 1 project we simply refer to this as personalisation, dropping the use of adaptive.

DPIE 2 points out that some systems will fit into more than one of these categories.

DPIE 2 highlights two systems using personalisation:

---

13 See, however, the JISC Review of OpenID Project [21].
Amazon [www.amazon.com], which uses intrinsic and extrinsic product relations to recommend products. Intrinsic product relations are defined by the nature of the product itself. Thus a CD might be by a particular performer, and an intrinsic recommendation would be for another CDs by the same performer. Or, again intrinsically, a wristwatch might result in recommendations of other wristwatches. Extrinsic product relations are to do with user behaviour; recommendations are based on either what they looked at, or what they bought. Thus a user looking at a particular camera may be provided with recommendations as to what other users who looked at the camera also looked at and/or bought.

In DPIE 1 we use the term mass recommendation rather than extrinsic recommendation (see section 6.4).

Pandora [www.pandora.com] uses user actions and music metadata to parameterise a stream of music that is played for the user.

DPIE 2 also mentions, without reference, Elsevier’s approach to developing services based on understanding the workflow of a researcher, and supporting workflow activities. Workflow is mentioned in this report too, but with the proviso that we believe that academic workflows are generally too poorly understood to be held up as a panacea of “workflow support will solve our problems in supplying support for learners, teachers and researchers”. Projects which progress the state of understanding of academic workflows and their exploitation, such as various projects already supported by the JISC, are exceptionally timely and required in order to progress the field.

There are similarities between the DPIE 2 approach to early gains from personalisation and in the Towards Implementation of Library 2.0 and the e-Framework (TILE) Project [7] and our own DPIE 1 Project: There is an emphasis across all three projects on exploiting initial and cheap-to-exploit opportunities for personalisation. DPIE 2 initially proffers these areas for exploitation:

- Automatic form filling, i.e. automatic use of up-to-date personal data for registration and other purposes.
- Gathering management information that may be useful for personalisation purposes.
- Personalisation on the basis of group membership (e.g. course membership). This is similar to TILE’s notions of personalisation on the basis of progression through the educational system and course and module enrolment.
- Push of content to the VLE based on personalisation data. We regard this as important, because VLEs are so commonly embedded in educational institutions. We note the general desirability of suitable technology, including personalised recommendation technology, that involves push to a number of targets, including VLEs.

The DPIE 2 Project notes that their final report will contain a larger set of areas for exploitation, with significantly more supporting discussion.

The DPIE 2 interim report picks up on “the one stop shop” discussed at a DPIE 1 workshop with Edina and Mimas services. Our view in DPIE 1 is that inevitable data liberation via web-based APIs will give rise to many different kinds of combinations of data sources. Considering the advantages of both Web 2.0 remixing, and possibilities of increased usability and task support, we support the view that combinations of JISC funded services should be able to be aggregated by one or more services, including search and recommendation services provided by JISC services, institutions and/or users.

The DPIE 2 Interim Report also mentions that:

There is a stream of thought … that the appropriate place to undertake personalisation is at the institution, which is more strongly connected to the user. This is a continuation of the traditional publishing model, whereby the service providers supply information and this can be consumed by users as required.

More recent work by the JISC TILE Project adds to this perspective and, subject to further investigation, proposes some personalisation facilities outside institutions. These are facilities that might be supplied non-institutional service providers.
No matter what the case eventually decided upon, intra-institutional or extra-institutional personalisation will be supported by bibliographic, serial and archival services that provide APIs for data access.

5.1.3 Outputs of the scenario work

DPIE 2 has performed with expert users a two-stage scenario study that will lead to use cases. Here we consider the axes used in the second stage of the scenario study, and additional factors that were also derived from the expert scenario participants.

Two scenario axes were chosen:

- Fragmentation vs. concentration of the education and research sector
- Open-access vs. closed access availability of high-quality resources

Wildcards and candidate axes identified by the DPIE 2 scenario were as follows, sometimes annotated with our comments after a dash:

- Death of VRE/VLE as a concept
- Government makes significant changes [improvements] to school[-level] education – we believe that changes need to be made here to set the tone for further and higher education, and that extra information literacy skills and metacognitive skills need to be learned at school level. However we note that the current emphasis seems more concerned with acquisition of knowledge, rather than methods that efficiently aid the acquisition of knowledge.
- Changes in user attitude to privacy – these may remove some of our concerns about the need for privacy, or alternately, with increasing data loss by government agencies, may increase user concern
- Increased proliferation of useless online “tools” leading to user fatigue – this is clearly relevant; reinforcing our support for functionality that supports users’ task-based needs, and personalised recommendation systems that supply useful data and functionality, both coupled with very high quality user interfaces.
- Trust in institutions – relevant for those initiations that collect personalisation data
- Accidental loss/disclosure of personal information by FE/HEIs – will lead to an increase in privacy concerns and, probably, a lessening in the amount of personal and attention data (section 6.3) available for personalisation purposes
- User privacy concerns – as above
- Proportion of open access resources – we believe these will become more numerous than closed academic resources in the future
- Degree to which online identity becomes prime, displacing real-world identity and face-to-face interactions – we feel “prime” is a bit strong for a likely future, but “important” is clearly part of the future
- Proliferation of shared services (not limited to administrative services)
- Proportion of user-provided devices and services (e.g. PCs, storage, email)
- Level and type of user-generated-content adoption in education and research – see above with our prediction of open educational resources growing
- Overall availability of funding to the sector
- Adoption of dynamic management at institutions
Prevalence of open standards – may assist adoption, within the proviso that standardisation sometimes delays progress.

We look forward to the outputs from DPIE 2 that amplify this scenario-generated material.

5.2 The Researcher of the Future Project

We have general agreement with some of the outputs of the JISC – British Library Researcher of the Future Project (RoF) [4, 19]. Regarding their overview of the current state of development in library and library information systems, the need to collect detailed usage data (although we would add an ethnographic bias to their approach), concerns about users’ levels of information literacy in critically assessing and selecting web-based information, barriers to use, search engine and semantic technology, and factors relating to the practice of library system design.

Specific comments on RoF outputs appear below.

5.2.1 The overview level

The RoF project is forceful in stating:

“The main message of this report for research libraries is that the future is now, not ten years away, and that they have no option but to understand and design systems around the actual behaviour of today’s virtual scholar.”

The project specifically identifies a focus on users as a determinant of direction for development of library systems, both in terms of an emphasis on supporting outcomes and becoming more attractive to users:

“The library profession desperately needs … to develop a new vision for the 21st century and reverse its declining profile and influence. This should start with effecting that shift from a content-orientation to a user-facing perspective and then on to an outcome focus.”

“Becoming much more e-consumer-friendly and less stodgy and intellectual. Few digital library offerings make any real attempt to connect with the larger digital consumer world: they simply do not chime with people’s experience of Facebook, YouTube, Amazon or even for that matter, ScienceDirect. Why, for example, don’t academic libraries try to emulate personal/social searching guidance offered so successfully by Amazon for many years?”

5.2.2 User and usage data

The RoF Project maintains, and we agree, that there is insufficient data for the design of effective library systems:

“The research literature is inadequate in this area and what serious material there is is eclipsed by anecdotal or unverified claims. The library community needs to invest more in data collection and analysis and to take its examples from commercial leaders (for example, Tesco) that have a much more detailed and insightful understanding of their customer base and preferences. In particular, there is a need for ongoing longitudinal data and intelligence functions to provide a vital early radar warning of oncoming change. Why don’t major national libraries have in-house user studies departments? Without this intelligence, service stereotypes can easily become detached from reality.”

The project proposes both automatic data collection and the establishment of consumer and profiling resources:

“Introducing robust, fit-for-purpose mechanisms for monitoring and evaluating their users and information services). Faced with the prospect that the future scholar will only ever want to use them remotely it is absolutely crucial that libraries have a means of monitoring and evaluating what they do. Furthermore, it is not sufficient to just listen and monitor it is also necessary to change in response to this data. Otherwise libraries will be increasingly marginalized and
anonymized in the virtual information world. No private sector corporation would survive on the basis of failing to invest in consumer profiling, market research and loyalty programmes. No library we are aware of has a department devoted to the evaluation of the user, how can that be?"

However, we wish to be clear that lack of data collection of the kind that we think is being proposed here does not preclude a useful start that can and should be made on the improvement of library/OPAC systems in the IE.

5.2.3 Users’ levels of information literacy

Disconcertingly the RoF Project reveals poor information literacy (seemingly without citing any sources of evidence):

“The picture that emerges from internet research is that most visitors to scholarly sites view only a few pages, many of which do not even contain real content, and in any case do not stop long enough to do any real reading. This is either a symptom of a really worrying malaise - failure at the library terminal - or maybe a sign that a whole new form of online reading behaviour is beginning to emerge, one based on skimming titles, contents pages and abstracts: we call this ‘power browsing’. We urgently need to understand the root causes of this phenomenon.”

RoF comments on lack of literacy resonate with our own informal observations.

We also note the following RoF perspectives:

“At national level, there is a desperate need for a wellfunded programme of educational research and inquiry into the information and digital literacy skills of our young people. If the erratic behaviour we are seeing in digital libraries really is the result of failure at the library terminal, then society has a major problem. Information skills are needed more than ever and at a higher level if people are to really avail themselves of the benefits of an information society.”

“Emerging research findings from the US points to the fact that these skills need to be inculcated during the formative years of childhood: by university or college it is too late to reverse engineer deeply ingrained habits, notably an uncritical trust in branded search engines to deliver quick fixes.”

“Really getting information skills on the agenda because clearly people are having great difficulties navigating and profiting from the virtual scholarly environment. To succeed it will be necessary to lead on outcomes/benefits (better researchers, degrees etc) and work closely with publishers.”

**Recommendation 11**: It is recommended that the JISC should consider raising the possibility of poor information seeking behaviour and metacognitive skills as an issue in education and work, including in continuing adult education and CPD, with the DCSF, HEFCE, and the BERR. This with a view to joint investigation and possible remedial action.

5.2.4 Barriers to Use

On barriers to use:

“Information consumers – of all ages - use digital media voraciously, and not necessarily in the ways that librarians assume. Any barrier to access: be that additional log-ins, payment or hard copy, are too high for most consumers and information behind those barriers will increasingly be ignored.”

We agree with this statement, but note that much personalisation perforce involves a potential barrier to use, namely authentication. It is important to offer users sufficient benefit as a consequence of registration and authentication such that these activities become worthwhile for them.  

\[14\] However, mass recommendation systems (e.g. “customers who ... also ...”) do not require registration and authentication.
Recommendation 12: It is recommended that the JISC funds development of integrated Web 2.0 based environments for learners, teachers and researchers.

The RoF Project makes some observations about search that are of interest. Firstly:

“… [libraries] need to make their sites more highly visible in cyberspace by opening them up to search engines.”

We agree wholeheartedly with this first point.

We have problems, though, with the second point:

“Given current levels of investment by the big corporate search engines, and static or declining library R&D budgets, it would seem that the only effective strategy is for tighter integration of library content with commercial search engines. This is urgent given that the business case for libraries is beginning to look weak to many outside the profession. It also fits with time-poor student experience as they work their way through college or study part-time or at a distance.”

While we would like to see library (catalogue and other) content opened up to commercial search engines, we also take a ‘UK PLC’ position, and assert that it of national importance to own publicly funded search engines that operate across UK and external library content.

The issue of search engine indexing and the development of IE search engines is revisited in chapter 10.

5.2.5 Semantic Technology

The RoF report identifies semantic technology as potentially supporting libraries, but in a limited sense:

“In five years, 2013, there could be substantial developments that might allow a whole generation of undergraduates to begin to experience its [semantic technology’s] potential. This is especially likely to be the case in niche areas, like e-Science, especially biology, creating new opportunities for major research libraries to be involved in completely new forms of activity such as real-time publishing and the sharing of experimental data on the internet.”

They do not explicitly mention the widespread potential for semantic technology applications. 15

5.2.6 Design

RoF was critical of the way library software was designed. RoF recommended a “suck it and see” or continual beta approach where prototypes are developed and tried out with users and modified according to experience. This approach is different to some service practice within the IE, where user interface changes are not thought to be desirable except during year-end academic breaks.

We like the experimental approached espoused by RoF, but note that there are other more considered approaches to designing new computer systems including ethnography for design, and use of rapid prototyping techniques together with formative evaluation and iterative improvements to the prototypes. These can be used in conjunction with an experimental approach.

There is a plea for simplicity:

“Our final message, one which information professionals have exactly the right skills set to address is the need for greater simplicity. We know that younger scholars especially have only a very limited knowledge of the many library-sponsored services that are on offer to them. The problem is one of both raising awareness of this expensive and valuable content and making the interfaces much more standard and easier to use. The cognitive load on any library user (or librarian) in trying to work through such complexity is at present immense.”

15 The DPIE 1 Project provides an introduction to semantically-based personalisation in one of its work package reports, but again, this is only one aspect of the potential of semantic technologies.
We fully agree with the plea, although we would rather use the term usability, instead of simplicity. Usability is defined by ISO as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [33]. In appropriate contexts simplicity is a component of usability.

We would completely remove the reference to standardisation. Standardisation would hinder development of innovative functionality and accompanying user interfaces, and, indeed, hinder the experimental approached espoused above.

5.3 Related projects

The recently announced report for the JISC on Key Concerns Within the Scholarly Communication Process is of interest; particularly with regard to copyright [32], which is a Web 2.0 in the IE issue.

Ongoing projects to watch are the JISC Review of OpenID Project [21], the JISC TILE Project [7], which is specifically concerned with Library 2.0 (a Web 2.0 in the IE issue), and the JISC Gold Dust Project [18], which is building a personalised recommendation system.
6 Search, recommendation and personalisation

6.1 Introduction

A quote which crystallises thought around the importance of search, recommendation and personalisation is from Richard MacManus, interviewing staff from Hakia, a semantic search engine company:

“Riza and Melek told me that the future of search engines will go beyond keyword analysis - search engines will talk back to you and in effect become your search assistant.” [23]

The notion of a search assistant is a powerful vision. Without doubt, search will become increasingly important as the web content increases, and powerful and usable search and discovery systems will be developed as a consequence. IE users would benefit if the IE offered similar search tools, with facilities to fulfil IE user needs.

Searches are user-initiated matches of user-supplied terms against sets of objects. The search algorithm employed ranks and displays the search results on the basis of goodness of match. Personalised search (as opposed to ‘unaided’ search) also includes personalisation data in the matching process.

Recommendation is an automatic process that is only indirectly caused by user actions. Recommendation matches some terms, e.g. user interests that have been expressed in the past, against object data, and rates and selects objects for recommendation to the user.

Recommendation may be:

- Asynchronous, e.g. supply of recommendations from a feed of new library acquisitions or
- Synchronous, e.g. supply of recommendations for a web page that is about to be displayed to the user.

If the recommendation process uses user-supplied expressions of interest, these are changed relatively infrequently. A greater degree of personalisation involves more frequent programmatic changes to personalisation data based on user actions. Because recommendation is based on expressions of interest and/or user actions, it is an inherently personalised process.

We note that, search and recommendation are similar at a high level; they select things for users. At a lower algorithmic level they may even use similar ways of doing this selection. For example, both search and recommendation might supply personalisation on the basis of the past behaviour of:

- A user
- Other users in the user’s social network.
- All users. If there is no use of a match of user supplied terms involved in the algorithm then this is “customers who looked at/bought this also looked at/bought x, y, z”

User behaviour is a consideration in the design of future search systems. Ferrara [10] describes different search behaviours and factors affecting search behaviour: Different search behaviours include alternating between search and browse, minimising the results set, surveying quickly, making immediate judgments, agonizing over the query, pogosticking (alternating between results of uncertain value). Factors that may cause variation in search behaviour are domain expertise, search experience, cognitive type, goal type, mode of seeking, and situational idiosyncrasies.

It may be well advised to try and support different kinds of search-related tasks as being relatively easy to achieve. A far more difficult step to achieve is personalisation where a search system recognises types of search behaviour and automatically provides appropriate support.

Subsequent sections in this chapter consider personalisation in commercial systems (i.e. personalised recommendation), attention and attention data, mass recommendation, personalised search, and, finally, implications for the personalisation of the IE.
6.2 Personalisation in commercial systems

Part of the DPIE 1 Project has been concerned with an examination of the state-of-the art in commercial personalisation to see what can be utilised in the IE. Subsequent sub-sections are concerned with these topics.

The state-of-the-art personalisation is exemplified by some of the commercial practice aimed at increasing sales conversion rates, where a sales conversion is the transformation of a user’s interest into a sale. The most fundamental view of this process is that some user characteristics are being matched against advertisement or product metadata, and that a selection is being made to show to the user.

In order to make the choice of what to show the user, two techniques, segmentation and behavioural targeting, are in widespread use. Of these behavioural targeting is the more sophisticated. These techniques use data about who the user is, and, possibly, past and current user behaviour.

However, some personalisation may not be based on a user’s own information. Thus many e-commerce sites include recommendations derived from all users, for example ‘customers who bought x also bought y’. We call this mass recommendation.

6.2.1 Segmentation

Segmentation is the practice of dividing site visitors into groups (i.e. segments), depending on factors like referring site, search term used in navigating to the site, or browser type. In commercial applications, advertisements are then served on the basis of segment membership.

This approach can help with click-through rates on targeted advertisements and on-site sales conversions. Segments may remain broad, and the degree of personalisation is not as high as provided by visitor-specific behavioural targeting.

6.2.2 Behavioural targeting

A model for behavioural targeting [11] constructed from an analysis of customer behaviour shows how users may progress towards making a purchase:

![Figure 6: Behavioural targeting sales conversion process](image)

---

22
One key to sales conversion is recognising each visitor’s stage of engagement with the sales process, and providing targeted advertisements appropriate to that stage. Thus, for example, a visitor at product discovery stage may respond positively to a special offer on the product being displayed, whereas at an earlier stage the offer may not be a productive use of advertising space on display to that user. This kind of personalisation has resulted in a 600% increase in conversion rates on one ecommerce site [1].

An even greater advantage can be gained from being able to deduce visitor characteristics and demographics – is a visitor who is looking at pages relating to a people carrier a mother with a large family to ferry around? If so there may be many different kinds of advertisements that she might respond to, which might include, e.g. baby wipes, nutritious food, school uniforms, educational software, and family holidays. Further customer identification or further interaction by the customer could select advertisements likely to eventually result in sales conversions.

It has been suggested that this level of individual targeting can only be achieved with customer relationship management (CRM) data from individual e-commerce sites. However, Double Click, a Google company, certainly does well from only repeat site and cross site visit information as the basis for targeted advertisements.

Exemplars

Two exemplars are provided by Yahoo behavioural targeting, and the Guardian Unlimited’s web site. The latter uses technology from Revenue Science Inc.

According to Yahoo:

“Behavioural targeting … allows advertisers to deliver specific targeted ads to consumers interested in a product, when they are close to the point of purchase, by leveraging actual online user behaviour. Even better, because the ad is served to a person based upon relevancy, it can be on a page that's not directly related to the product which expands the opportunity to communicate with potential consumers.” [34]

To implement this a click stream is generated from search terms, click-through on search results and advertisements, and micro-sites visited. This click stream can be used to deduce visitor characteristics, and in turn these characteristics can be used to target advertisements to the user.

Yahoo has 600 large corporate clients who employ Yahoo behavioural targeting.

The Guardian Unlimited [14] uses behavioural targeting to provide site visitors with targeted advertising based on content read or visited, including:

- The sections and pages that individuals visit on Guardian Unlimited, including on a first visit.
- The words that appear on the pages they have read.
- The number of relevant pages they have read.
- How often they read these pages.
- How recently they were reading those pages.
- Their relationship with Guardian Unlimited.
- Registration data such as job title or seniority.
- Frequency of visit

The Guardian Unlimited can deduce the following for some of its site visitors:

- Domain details.
- The industry they work in.
- Their company’s FTSE 350 status.
This will almost certainly be by reverse translation of the visitors’ IP addresses to URLs, then a URL ownership lookup to see if the URL is a corporate entity, and, if so, a subsequent search in online databases to determine industry sector and corporate status.

The Guardian promises a demonstration of behavioural targeting starting from [13]. Quick trial use of the demonstration didn’t seem to provide much in the way of personalisation, but since the Guardian’s relationship with Revenue Science Inc., the supplier of the technology, has existed since May 2005, we assume that the statistics on ROI and conversions are positive, and that the Guardian Unlimited’s behavioural targeting and consequent income is sufficiently effective to justify an ongoing relationship with Revenue Science Inc.

6.3 Attention and attention data

A somewhat different slant on behavioural targeting is provided by attention and attention data. What users attend to, i.e. where they go on the web, how long they spend there, and what they do there can provides attention data characterising the users behaviour and interests. The attention data can be used for personalisation.

With attention and attention data comes the idea of an attention economy [17] where both attention and attention data have worth. This is not new, many web sites provide services and information in return for the opportunity to display advertisements to users and perhaps gain attention to some of those advertisements. In a recent example the Blikx phone network engages in the attention economy: A network subscriber pays no money to use the network, instead he or she receives targeted advertisements (perhaps paying attention to them) by phone. In return for this, the subscriber receives free phone calls and free SMS texts.

Attention Profiling Markup Language (APML) is a language that is in development to provide a standard for harvesting attention data. APML is an XML based language. APML is still in a very early stage of development and use, and may not be adopted broadly.

"APML allows users to share their own personal Attention Profile in much the same way that OPML allows the exchange of reading lists between News Readers. The idea is to compress all forms of Attention Data into a portable file format containing a description of ranked user interests." [www.apml.org].

One use of APML is described by lifestrea.ms [lifestrea.ms]:

"But we don’t calculate your APML only for external purposes. ... From your attention data we calculate what newsfeeds have a high priority for you, what new topics could be of interest for you and what people you should get to know, 'cause they are heading for the same things. We suggest feeds, topics and people (anonymously) to you as soon as we find them in your … live dashboard."

APML, or a variant thereof, might be suitable as a basis for the construction of an attention store to be used in JISC-funded personalisation activities.

Attention data gathered in one place may be available for use elsewhere. However, there are many instances of attention data being gathered in corporate or Web 2.0 company silos. There is some concern about who owns and who can use attention data.

The Attention Trust offers principles [2] that underpin user control of attention data:

- The user has control over his or her attention data and can store it where they wish
- The user can securely move his or her attention data to a different storage place.
- The user can trade his or her attention for something of value supplied by another party.
- The user can decide who to trust with his or her attention data and can see how it is being used.

It seems unlikely that existing commercial attention data silos will become user controlled, but users may start gathering their own attention data via browser extensions, to deploy that data where they will.
Exemplars

Exemplars are supplied by Pandora and Particles, which both dynamically change their behaviour in response to user response (i.e. according to attention data).

Pandora is a service (only accessible for full functionality in the U.S.A.) that creates personalised streaming internet radio stations. Pandora grew out of the Music Genome Project which used volunteer effort to categorise pieces of music according to factors like melody, harmony and rhythm, instrumentation, orchestration, arrangement, lyrics, vocals and vocal harmony. Users select songs to seed their station, and can then optionally approve or reject individual songs that are playing. Pandora responds to this user attention data by fine-tuning the kind of music it plays on the station. Pandora provides an example of the use of rich metadata in the provision of a user service based on attention.

Particles is an alerts platform that displays information to users as alerts whenever suitably relevant information becomes available. The information is typically derived from RSS feeds, but could also include calendar reminders, twitters, etc. Alerts are displayed while users are doing something else, using different display methods according to how relevant the alert is, and how much of the user’s attention it should therefore consume. Incoming information is evaluated for relevance to the user when it becomes available, and, if relevant, Particles then chooses a display method. Personal information used in relevance computation comes from user entered keywords and sources. Particles can also optionally scan the user’s hard drive to periodically auto-detect user interests from the documents, conversations, emails and sites visited by the user. The system also uses item popularity on the Internet and item recency in its calculation of relevance. Different display methods include in a sidebar, in a semi-transparent alert windows that fade in and out, and in a ticker. Users can click on items that interest them and then see the item in a browser. Particles also has a variety of related functionality.

Firefox users can harvest their own clickstream via the use of the Attention Trust’s Firefox plugin, and may use that data to parameterise the operation of currently only two Attention Trust approved services.

The use of attention data is potentially large in the IE, if the requisite attention data can be exploited, possibly together with the considerable amounts of metadata in the IE. An important component of this is the attention store service that is proposed in chapter 8, and mechanisms to automatically gather attention data for those users who choose this option. We make recommendations for personalisation demonstrators that use our attention store in our recommendations to the JISC for further demonstrators.

6.4 Mass recommendation

Mass recommendation is a term invented for the purposes of this report. We use it to refer to recommendations which are made on the basis of previous customer behaviour. Amazon provides an example of this, in its “customers who bought this item also bought”. Implementing this kind of recommendation seems relatively easy. However, to implement this scheme on a more thorough basis than simply using multiple purchases in the same basket involves the maintenance of non-anonymised data over time to gather data on repeated sales to the same individual.

6.5 Personalised search

In many ways (see section 6.1) we view personalised search as being closely related to personalised recommendation, and particularly so in respect of matching algorithms that produce a ranking for items.

Two separate reports, one on personalised recommendation, and one on semantic technologies, are both relevant to personalised search (and are accessible via misc.jiac.ac.uk/dpie1/deliverables.html).
6.6 Implications for the IE

There are parallels to be drawn between commercial personalisation facilities and the future facilities of the IE, where the aim should be to provide each user with the most relevant information and functionality at any given time. How then might the IE take advantage of segmentation and behavioural targeting?

In the application of segmentation to IE personalisation, some obvious differentiators are:

- **Discipline**
- **For students only: Enrolment on particular programme (aka course) modules**
- **Progression level:** The items of information and bibliographic etc information of use to users might depend on their progression level in the educational system. Thus a first year undergraduate might be interested in a quite different references, texts, objects and conversations than a postdoctoral researcher.
- **Location:** This is necessary for the supply of physical objects or services, e.g. a book in a user’s local institutional library, or a “talk to a librarian” facility offered through the web. While some locational issues are addressed by OpenURL resolvers, students, teachers and researchers may work in different locations, e.g. institutional, at home, on public transport, and have different informational needs in each location.

For behavioural targeting initial parallels to start from are:

<table>
<thead>
<tr>
<th>Behavioural targeting</th>
<th>IE equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customers</strong></td>
<td>IE users</td>
</tr>
<tr>
<td><strong>Purchase stage</strong></td>
<td>Information as to ‘information lifecycle stage’ (e.g. information discovery, retrieval, use, etc.). Questions here are:</td>
</tr>
<tr>
<td></td>
<td>- What useful models do we have of the information use lifecycles; creation to curation and discovery to delivery (C2D2)?</td>
</tr>
<tr>
<td></td>
<td>- What recommendations can be usefully supplied to a user who is browsing?</td>
</tr>
<tr>
<td></td>
<td>- What recommendations and search results can be given to a user who is engaging in different kinds of search behaviour discussed in section 6.1?</td>
</tr>
<tr>
<td><strong>Sales conversion</strong></td>
<td>Successful data supply: Successfully supplying the IE user with some useful information, e.g. bibliographic data, or, even more usefully to the IE user, the online resource that the bibliographic data refers to.</td>
</tr>
<tr>
<td></td>
<td>Successful functionality supply: Successfully supplying the IE user with some useful functionality, e.g. if bibliographic data is presented as a social object, the user can interact with that object, and, through the object, with other users.</td>
</tr>
</tbody>
</table>

Behavioural targeting information in the IE might include:

- The user’s ‘lifecycle stage’ – activities the user is undertaking as part of a larger information acquisition task. We don’t know much about these activities, although some information appears in sections 4.1.2 and 6.1.
- Past behaviour and interests as captured from the user’s click stream as attention data: This includes searches made by the user, and click throughs on items in search results, and any repeat visits made to the resources.

---

16 For fear it get lost, we mention that user generated content on the basis of progression level might be selected manually, as customisation at appropriate times. If conversations do develop around social objects, there might be need to dynamically choose what is to be viewed. Thus while it may be of use to undergraduates to participate in all conversations about those objects, there might also be a need for a particular users to see only contributions from researchers. This could be achieved via customisation choices.
• User supplied data about their preferences and interests as stated by the user and/or extracted from documents or web pages supplied by the user (the latter via specification of URLs).
• The user’s social bookmarks.
• Interests declared by the user.
• RSS and Atom feed subscriptions

There are legal constraints in maintaining personalisation data (see recommendation 01). Beyond these constraints, we suggest that the JISC should treat personalisation data carefully (recommendation 02) to build reputation and maximise the opportunities to accrue personalisation data. Thus:
• The default for a new system user must be for no recording of click stream data, nor for any other kind of personalisation data to be deduced or gathered.
• Before users in a sign-up process are signed up for personalisation, they must be made aware that they can destroy their personalisation data at any time and opt out of the scheme at anytime. The default at opt-out time should be to destroy the opting-out user’s personalisation data, unless the user selects to keep it for some reason.
• Users should also be assured of data privacy from other humans and from external systems. This aspect of security requires close attention.

For personalised search, we have two different views to which we subscribe:
• Search is an activity where a user is interested in finding particular information and possibly personalisation may hinder finding new things as personalisation algorithms concentrate on returning items that are similar to the ‘known past’.
• Search on the Internet will improve to such a point that no one will consider using search engines that do not offer significant and meaningful personalisation-based improvements.

Luckily, these differing views are not at odds with each other, or with a phased approach to IE development.

There is evidence that, at its current stage of development personalisation may lead to a narrowing of results, and exclude less obviously related but relevant results. The payoff, and acid test, has to be that the personalisation yields better results than if it were not used. This may in turn determine what kind of personalisation is offered, and for topic-based personalisation, the choice of personalisation method.

**Recommendation 13:** It is recommended that in the shorter-term IE development of search for production use should concentrate on search without the use of sophisticated personalisation algorithms, but with the use of easily obtainable user data that provides the opportunity for personalisation, e.g. discipline, progression level and location, and as much task-specific support as can be supplied. This will provide early (interim) results in increasing search usability. See also recommendation 14.

**Recommendation 14:** It is recommended that the JISC should fund exploratory experimental projects in the provision of personalised search. Some interesting approaches to investigate are the use of resource metadata, the use of user supplied data (their own preferences), attention data, diverse social and Web 2.0 approaches, and semantic approaches that, e.g., consider content users provide as exemplars of areas of interest. See also recommendation 13.

**Recommendation 15:** The JISC-funded Gold Dust project is currently building and testing an adaptive personalisation recommendation system. The intention is to create personal interest profiles for a target user group of 50 volunteers. These profiles will then be mapped to content harvested from various RSS feeds to provide recommendations. It is recommended that the JISC should examine the Gold Dust Project’s (future) outputs to see if they fulfil the requirements for a recommender that could work over services that are provided by the JISC National Data Centres.
7 Service survey

The service survey part of DPIE 1 produced a survey tool using Bristol Surveys. The following JISC services were surveyed: Archives Hub, Copac, Intute, Jorum, SUNCAT, and Zetoc.

7.1 Survey assessment tool

The assessment tool is available online at Bristol surveys. The survey focuses on questions under the following areas:

- Service details and contacts
- Current and future customisation features
- Features available without logging In
- Registration features
- Utilising users usage data
- Alternative interfaces/Access mechanisms
- Web 2.0 Developments

The survey tool web pages are reproduced and available via [misc.jisc.ac.uk/dpie1/deliverables.html](misc.jisc.ac.uk/dpie1/deliverables.html)

Other services may gain access to the survey tool to survey themselves. Please email markvanharmelen@gmail.com.

7.2 Survey of Edina and Mimas constituent services

We surveyed six Mimas and Edina services as requested by JISC. We pre-populated answers in the survey tool, and asked the individual services to correct and add data to the survey as necessary. We later used a smaller questionnaire to gain more information about the SRU search protocol, persistent URLs, and related information.

The services surveyed were Archives Hub, Copac, Intute, Jorum, SUNCAT and Zetoc.

The detailed results of the survey are accessible via [misc.jisc.ac.uk/dpie1/deliverables.html](misc.jisc.ac.uk/dpie1/deliverables.html)

7.3 Findings and implications for the IE

We are currently missing data from Suncat and Jorum for the second phase of our service survey, and some of their data here may be incorrect.

For future Web 2.0 use the surveyed services generally lacked two key features: Suitable APIs for Web 2.0 use, and persistent URLs which would be useful for a variety of purposes, including social bookmarking, construction of social objects, and external search engine use. However, two services, Copac and Intute, can be used now. There is useful Library 2.0 and Web 2.0 functionality to be obtained from these two services.
Search protocols

Search/Retrieve via URL (SRU) is a RESTful search protocol that utilises Common Query Language (CQL) that we recommend for Web 2.0 use.

Query and search facilities that we know of are:

<table>
<thead>
<tr>
<th>service</th>
<th>search protocol</th>
<th>Z39.50</th>
<th>SRU</th>
<th>SRW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archives Hub</td>
<td>yes</td>
<td></td>
<td>later in 2008</td>
<td></td>
</tr>
<tr>
<td>Copac</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Intute</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Jorum</td>
<td></td>
<td>beta version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUNCAT</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zetoc</td>
<td>yes</td>
<td>later in 2008</td>
<td>later in 2008</td>
<td></td>
</tr>
</tbody>
</table>

We note the existence of services which query multiple catalogues and search engines using, inter alia, z39.50. For example, TechXtra queries Copac, Intute and Jorum. The PerX Pilot cross-searches those three plus Google. It is possible to work with the Z39.50 search protocol, but we would prefer to see SRU supplied universally.

Persistent URLs

A current problem is the lack of persistent URLs that uniquely identify resources held in the services.

<table>
<thead>
<tr>
<th>service</th>
<th>persistent URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archives Hub</td>
<td>later in 2008</td>
</tr>
<tr>
<td>Copac</td>
<td>yes</td>
</tr>
<tr>
<td>Intute</td>
<td>no plans to implement work around exists</td>
</tr>
<tr>
<td>Jorum</td>
<td>no</td>
</tr>
<tr>
<td>SUNCAT</td>
<td></td>
</tr>
<tr>
<td>Zetoc</td>
<td>yes</td>
</tr>
</tbody>
</table>

Persistent URLs are needed to enable all-important search engine indexing of content, and to allow for the construction of the social objects recommended in this report.

We observe that lack of persistent URLs prevents service content being indexed by the major search engines. Elsewhere we recommend that all JISC services move to supporting persistent URLs for their content as a matter of urgency.
**Recommendation 16:** It is recommended that Edina and Mimas services that do not supply SRU and/or persistent URLs make plans to supply them as a matter of relative urgency. It is also recommended that all services ensure that their service contents are indexed by the major search engines.

**Duplicate resources**

Duplicated records can appear in bibliographic or serial catalogues for a number of reasons. In a the catalogue for a single holding, there may be duplicate records for the same replicated physical resources (e.g. same edition of a book), or there may be duplicate records that are of import to some users (e.g. for different editions of the same book). The problem of duplicate records becomes larger in union catalogues, such as Copac, where different catalogues are combined. Removing duplicates from a large catalogue is not a trivial matter, firstly some duplicates may be of some value to some users, and secondly, the sheer volume of records in modern union catalogues is a problem, e.g. Copac has 72 million records.

For searches across multiple data sources, duplicates may also occur because different sources supply the same result.

For search, the experience of the JISC PerX Project is that ranking of results is difficult, and PerX was not able to compare and rank results from different Z39.50 targets alongside results using other protocols. The conclusion is that "The reality is that for Z39.50 based federated search services, true de-duplication is virtually impossible and neither can these services perform a relevancy ranking that is globally relevant and uniform for all the searched Z39.50 databases." [5]

We predict similar problems with SRU based searches, albeit with slightly different issues (see [5]).

From observing problems using Z39.50 initiated searches, where duplicates and arbitrary server-based ranking algorithms make it hard to process and rank results, we predict similar problems with SRU based searches, albeit with slightly different issues (see[5]). Some investigation of the problem of duplication has been performed in Edina and Mimas, but more work is needed

**Recommendation 17:** It is recommended that Edina and Mimas should devote, with JISC funding, significant effort to finding if there are any approaches to catalogue de-duplication, and if not, ascertain what is possible with duplicated records. Worldcat's success in de-duplication of their union catalogue should be noted; it is believed that this involves, at some stage, considerable computation on a Beowulf cluster.

**Recommendation 18:** It is recommended that, subject to work performed in recommendation 17, serious consideration should be given to establishing a replication free service containing bibliographic and serial information.

**Restrictions on access**

Not all services have open access. Zetoc and Jorum require authentication, although Jorum is being transformed into JorumOpen [20]. Some services offer additional services to logged-in users.
8 Extensions to the IE

8.1 Introduction

We propose a set of architectural extensions that enable the introduction of personalisation and Web 2.0 features in the IE. These extend the IE and integrate IE services with Web 2.0, Learning 2.0 and Library 2.0 facilities. If these will be adopted by users is an open question. We venture an approach that involves highly usable task-support facilities to maximise potential uptake, and suggest a cautious approach: The introduction of features with highly usable user interfaces that supply supportive functionality will aid uptake. Conversely, either sub-optimal interfaces or sub-optimal task support is highly likely to hinder uptake.

A scenario involving search motivates the introduction of a specialised user agent, the compositor service, to search IE services. As well as searching, the compositor also composes individual results based on service resources, information from other services (e.g. Amazon reviews) and functionality from other services (bookmarking, writing reviews and so on). In this way IE resources are transformed into social objects that enable discourse about the object contents and communication between users.

In itself this is insufficient for user support within the IE. Support for other user activities must be supplied to take advantage of opportunities to maximise IE utility. This goes beyond simple handling of lists of social objects, or the push of social objects into a VLE. The centrality of communities to academic activities has been discussed above (section 4.1.1): Thus, in providing user support within the IE, it is important to (i) support collaborative academic activity with appropriate social networking systems and (ii) supply tools to find, collect and create knowledge. In our view these should be integrated with social objects as proposed here.

Some of the extensions that follow propose an alternate user interface to several JISC National Data Centre services. In this, we assume that constituent data centre services like Copac continue to supply their own stable user interfaces. Furthermore, we are not promoting a ‘one-stop shop’ or that JISC ‘tries to do everything’; rather we simply seek to enhance the usability of parts of the IE.

8.2 Search

As an example scenario, a student, Jane, is searching for information using a search service that allows her to search for bibliographic, serial, archival and repository data.

Previously, Jane’s tutor recommended that Jane use Academic Search, a personalised search service that provides a search over various services including a range of Edina and Mimas services. Academic Search is available for Jane’s favourite social networking service, social:net, and as Jane is familiar with customising her social:net account with add-on applications like MegaWall, it was simple for her to customise her account with the search service.

Consequently, Jane starts her search for information from social:net. The project is on the influence of the Mancunian socio-economic milieu 1840-1870, so she starts by searching for Engels and Manchester, and is presented with a page of search results. She follows up on four of the search results: A book in the British Library, a journal article that is available through her local university library, a resource on the web, and some archival material held in Manchester.

When she clicks though to these results she sees and can add reviews and comments, is able to bookmark, can ask to be told about similar results on a number of criteria, and can signal interest in seeing other users’ changes to this and perhaps similar objects, with notifications appearing in her feed reader. Here Jane is seeing and using social objects.

When Jane added Academic Search to her social:net, she was asked if she wanted personalised results from the service when she searches. Jane read that personalisation requires that information be recorded about her searches and the items she chooses to follow up, that this information is stored by the JISC,
and that she may easily turn off recording and/or delete this information at any time. Jane is not concerned about personalisation information being stored somewhere, and accepted the personalisation option.

When Jane invokes a search, Academic Search uses a web service that searches several JISC IE services through their search APIs. The web service gathers search results, prioritises some of them according to its personalised search algorithm(s), and presents the results. Because Jane previously selected that she wanted personalised search, her search terms are recorded for future use in personalisation.

Various kinds of personalisation could be used, we mention two:

- The search may prioritise resources at the university that Jane studies at, or on the basis of her progression level and course and module enrolments. See section 6.2.1 on segmentation.
- The search may prioritise results based on Jane’s searches and click through behaviour, particularly for repeated retrieval of the same item. See section 6.2.2 on behavioural targeting and 6.3 on attention and attention data.

When Jane clicks through to a results page, say for a Copac record for the book in her local university, the results page is composed of:

- Bibliographic information from Copac.
- Book covers.
- Tags associated with the book. Clicking on a tag displays a page of bookmarked items for that tag.
- Reviews from IE users
- Reviews of the item from around the web, e.g. from Amazon.
- Recommendations for similar and related items, where it may well work best if recommenders are of similar type to users consuming recommendations and search results, e.g., what a researcher recommends is likely to be of most interest to other researchers.
- A bookmark button, and a button for change notification.
- Importantly, there are facilities to gather user content, including the ability to add a review, rate, or perhaps take part in a threaded discussion.

Together this data and functionality provide a social object for Jane’s use.

Each of the constituent parts above comes from a different service and is assembled by and interacted with via the compositor service. This same service is also responsible for invoking the search and displaying search results (see also section 8.3).

There might well be some useful life to information in the proposed system, for example, twenty years down the line one might not want to be storing student personalisation data for students who have long left the academic system and who have displayed no interest in further research or life-long learning. However, choices are not clear cut, e.g. it would be a pity to deny that person access to their undergraduate item collection and tags when they retired and decided to pick up on their earlier learning.

**Recommendation 19:** We recommend that the JISC should consider whether students should retain continued use of any JISC IE environment enhancements such as social networking systems after leaving education. Some alternatives are: No use after end of education or employment in the academic sector, keeping the account for a few years (choices: active or not) in the hope that ex-students re-enter the wider educational/research system, or let use continue indefinitely to create a national continuing education and research support system. Alternately, the JISC could pass social network support to educational institutions and avoid the question completely, but this would lead to a poorer experience for students who move between institutions or who are registered at more than one institution.
8.3 IE architectural extensions

The architectural extensions to the IE are shown below. Existing IE services, other web services and client side components are shown as shaded rectangles. Extensions to the IE architecture, mostly common social services, are shown as rectangles with a white background. An incomplete but illustrative set of connections between components is shown.

In considering extensions to the IE architecture it is paramount avoid to a walled garden where IE components inter-work with themselves, but not with the rest of the web ecosystem. Lack of interoperability will:

- Impact overall usability of IE resources. This already happens to services which do not supply fixed URLs for their resources, and thereby prevent users discovering those resources via Google and other external search engines.
- Prevent reuse use of IE facilities in user or programmer created mash-ups.

A recommendation to this effect appears elsewhere.

Various new components in the architecture allow for delivery of additions to the IE that provide for architectures of participation through (i) the transformation of IE resources into social objects, (ii) the use
of Library 2.0 approaches to bibliographic, serial, and archival data, and (iii) by the introduction of specialised social networks. Components are underlined below:

- Search, Library 2.0 functionality and social objects are delivered by the compositor and the services it employs. The Library 2.0 functionality delivered via the compositor can start to address added search facilities, including facets, search history, and so on.

- One or more specialised social networking services that are specifically tailored for knowledge acquisition and generation activities, that supply learning, teaching and research facilities to individuals, and that support them in these activities when they work on their own and in groups. Examples of this kind of system could be a PLE, or a research collaboration system that is similar to 2collab [www.2colab.com] or myExperiment [www.myexperiment.org]. Ideally these should integrate the use of social objects.

- A specialist social research facility, which we use as an illustrative specialist social networking service, appears in the diagram as a reminder that there may be more than one specialised social networking service.

- The attention store collects search and clickstream derived data from the compositor and possibly from one or more specialised social networking services.

- The attention store should be accompanied by facilities for users to sign up for automatic attention data collection. If signed-up for this users should be able to (i) delete their attention data while staying in the scheme, (ii) choose to opt out of the scheme while, by default, automatically removing their stored attention data during the opt-out process.

- Provision of personalised search and personalised recommendation, services that are used by the compositor. Personalised search and recommendation may use data from the attention store or other sources as part of their personalisation activities. Personalised recommendation may also supply RSS feeds going to other IE and external components, e.g. a specialised social networking service or an aggregator/news reader.

- Provision of suitable plugins to social networking services and start pages. A suitably comprehensive but minimal set of targets is Facebook, iGoogle and Netvibes.

- Provision of a widget hosting mechanism for others’ widgets. At this time the only approach that we would recommend is the implementation of Google’s OpenSocial API.

- Browser extensions may be a possibility at some stage of development.  

8.4 The Compositor

The compositor creates social objects by wrapping web resources, including IE bibliographic, serial, repository and archival data with information from other services, and with facilities for users to contribute to the data around the object and to communicate with each other through facilities offered by the object.

The compositor has been prototyped as part of the DPIE 1 Project, and interested readers are urged to try that demonstration (which is moving to iris.cs.man.ac.uk/compositor).

To show actions of a compositor-like system, imagine a researcher is using search facilities provided by the compositor to find resources pertaining to web-based community formation. He uses the compositor to search for books on design for communities, and then clicks through the search results to obtain a result for Powazek’s ‘Design for Community’.

---

17 We do have recommendations for delaying implementation of personalised search in the chapter 10.
18 Excellent functionality of different kinds is offered by four Firefox extensions, the AttentionTrust, Zotero, Operator +, and Diigo extensions. An IE extension that interoperates with IE services may offer interesting possibilities for user support. Clearly user-perceivable advantages must outweigh users’ installation inertia.
The researcher is then presented with a page (represented below with book detail and a recommendation for a similar volume, but without showing facilities to add user generated content). This page is created by the compositor mashing up data from several sources.  

19 The (now somewhat older) Figure 9 illustrating compositor functionality contains a slight difference to the overall architectural extensions in Figure 8, and shows a review finder. In our demonstrator, and now in Figure 8, review-finding functionality incorporated in the compositor.
To provide social objects the compositor needs functionality that provides for the addition of user generated content to the object. This is achieved with the use of additional services, as in section 8.3 above.

All responses via the page’s buttons for ‘reserve’, ‘add to my collection’, ‘bookmark’, ‘find reviews’, and ‘review’ pass through the compositor. This allows the compositor to record information about user actions in the attention store for future personalisation use — for example, if the researcher reserves the book then this is assumed to be valuable information about some of his interests; if he does this a second time then that is even more valuable information.

Social bookmarking system choice provides an example of customisation (as opposed to personalisation). On clicking the bookmark button the scenario’s researcher will be taken to his preferred social bookmarking service according to his previous customisation of the system.

### 8.5 Implications for the IE

While some improvements are still needed (see section 7.3), the IE is well positioned to inter-work with Web 2.0 services because its services can, in general, inter-work with Web 2.0 services. However, there are some factors to consider; openness in the larger Web ecosystem, service granularity and interoperability, potentially interesting components for the IE, and aspects of the project and service development process. These are discussed in the remainder of this chapter.

#### 8.5.1 Openness in the larger Web ecosystem

It is important to form a view on using commercial Web 2.0 systems ‘out there’ on the Internet. Ownership of data and data security are two concerns in making such a choice. Web 2.0 services may keep user data in silos, or licence the data restrictively. Web 2.0 services may suffer data losses without guarantee of backups, and they may go out of business taking user data with them. Web 2.0 services may consider some academic material as unsuitable content, and delete it from their sites. Some Web 2.0 services may not offer strong enough access control for some academic material. These and related issues, including those to do with data longevity, copyright and IPR are considered in more detail in the JISC study on Web 2.0 Content Sharing for Learning and Teaching in Higher Education [12], and we support recommendations made there.

Considering the import of national education we consider that it is inadvisable to place parts of the educational infrastructure in the hands of Web 2.0 companies when there is an alternative of a carefully managed UK-based service with continuing funding. We are already seeing some Web 2.0 services going out of business, or being taken over by larger companies and then changed/evolved from the original purpose. There is thus an argument for providing Web 2.0 components in the IE.

However, it would be counterproductive to build a wall between the IE and the outside Web 2.0 world. We want IE users to reuse content and facilities they have constructed outside the IE. We want IE users to be able to take advantage of Web 2.0 developments outside the IE. Thus it is desirable to implement wide interoperability such that advantage can be taken of outside developments, albeit with less data security. In this context, we can imagine components from inside and outside the IE being mashed up together, and with, where appropriate, some IE components providing backup for externally-held data.

Where Web 2.0 services are replicated in the IE, they should ideally be not just a replication if they can be a further improved for use in learning, teaching and research activities.

**Recommendation 20:** While the DPIE 1 project has provided a survey of widget implementation mechanisms, is recommended that the JISC commission a small higher-level study on the advantages and disadvantages in interfacing IE components to non-IE social networking systems via widgets. The study should include functionality, privacy, and data security as topics of investigation.
8.5.2 Service granularity and interoperability

One of the many advantages of Web 2.0 is the ability for a user to pick and mix what specialised web services to use. More sophisticated users may get some services to inter-work with each other. On the other hand, integrated Web 2.0 services may contain a social networking substrate and then layer other facilities, including specialist social software facilities, on top of the substrate. These integrated systems are popular with users who do not want to pick and mix. We provided some examples systems that may be used to support research in the exemplars in section 4.1.2).

<table>
<thead>
<tr>
<th>Many loosely-coupled specialist services</th>
<th>Possibly non-exclusive choices</th>
<th>Integrated multi-facility services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can mix and match facilities from different Web 2.0 services</td>
<td>advantages</td>
<td>With the right user interface and facilities an integrated system can meet the needs of less technically astute users</td>
</tr>
<tr>
<td>Each end user needs technical literacy and bricolage skills to build an environment for use</td>
<td>disadvantages</td>
<td>Some kinds of functionality can only be implemented in a unified system</td>
</tr>
<tr>
<td>Maintaining multiple logins is tedious</td>
<td></td>
<td>If this is a closed system (and it need not be closed), then individual users have no ability to ‘meld in’ extra facilities that they may need</td>
</tr>
</tbody>
</table>

Figure 9: Non-exclusive architectural choices: Many loosely coupled specialist systems and more integrated services

It would be unwise to label more-integrated services as systems that ‘attempt to do everything’ and reject them out of hand. If the set of facilities offered by such a service is genuinely useful to its users, then service has real value, and may contribute to the IE.

What is important is that these services inter-work with the rest of the infrastructure; i.e., that public pages are search engine indexable and therefore discoverable, and that the service has a permeable boundary, e.g. users of the broader web may add comments to some pages to help build and spread knowledge, and RSS or Atom feeds allow changes to content in the service to spread outwards to interested Internet users.

**Recommendation 21:** It is strongly recommended that the architecture of any integrated components should be open, in that they are capable of inter-working with other Web 2.0 components.

Development process characteristics that are useful in developing Web 2.0 and personalisation features for the IE include:

- Continual beta development; where “beta” refers to the practice of trying out new features and adopting or discarding them. As used here, beta does not include public delivery of buggy or badly engineered systems.
- Agile project management (e.g. SCRUM) and development practices (e.g. eXtreme Programming)
- Small trial use, formative evaluation and iterative improvement
- Subsequent closed beta testing, or, for a production system, selective trial delivery to about 10% of the users
- Benchmarking performance, and sizing installations
- Improvements to live systems tried on a subset of users before general implementation

**Recommendation 22:** It is recommended that project work in the Web 2.0 and personalisation areas includes beta development, agile management and development methods, small trial use, formative
evaluation and iterative improvement; subsequent closed beta testing; benchmarking performance and sizing installations; and improvements to live systems tried on a subset of users before general implementation.

**Recommendation 23**: It is recommended that the JISC adopt and encourage more of a continual beta attitude towards parts of the IE; trying ideas out experimentally. If introduced features don’t work, they should be discarded or improved till they do work, as appropriate. In this it is recommend that the JISC allow for both experimentation and failure in the pursuit of success. The emphasis on continual beta requires particular project practices as, e.g., in recommendation 22, and may sometimes require a funding model larger than the more usual £50-200K model.
A demonstrator was constructed by team members from Inovia-Team to demonstrate how the architectural extensions proposed in the previous chapter might be implemented.

The demonstrator uses a compositor architecture (from chapter 8) to search Copac using SRU and CQL, and incorporates facilities drawn from several Web 2.0 services. The demonstrator augments search results with book covers obtained from Amazon, and makes available delicious bookmarking, and where available, search inside books via Google Books.

Figure 10: Compositor search results

When individual book data is retrieved by clicking through from a search result page, it is augmented with a cover and reviews from Amazon, if they are available. Bookmarking facilities are again offered, as well as, if available, search inside the book contents via Google Books. This provides a demonstration of mashed-up results. It is a relatively small step to transform these mashed-up records into social objects.
A rudimentary attention store is used in conjunction with clickstream data gathered by the compositor to demonstrate how APMEl is gathered to record a user’s clickstream in an attention store. Here we show clickstream data in the attention store database, where a single user is, for now, identified by his IP address 20

<table>
<thead>
<tr>
<th>id</th>
<th>ip</th>
<th>query</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>127.0.0.1</td>
<td>Query=m:mark van Harmelen</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>11</td>
<td>127.0.0.1</td>
<td>Query=m:mark van Harmelen</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>7</td>
<td>127.0.0.1</td>
<td>Copaid=021326169</td>
<td>moreinfo</td>
</tr>
<tr>
<td>8</td>
<td>127.0.0.1</td>
<td>Query=matt.welsh</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>9</td>
<td>127.0.0.1</td>
<td>Query=matt.welsh</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>10</td>
<td>127.0.0.1</td>
<td>ISBN=0596007604</td>
<td>searchin</td>
</tr>
<tr>
<td>13</td>
<td>011615926</td>
<td>Copaid=021326169</td>
<td>moreinfo</td>
</tr>
<tr>
<td>14</td>
<td>127.0.0.1</td>
<td>Query=Human-computer interaction</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>15</td>
<td>127.0.0.1</td>
<td>Query=Human-computer interaction</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>16</td>
<td>127.0.0.1</td>
<td>Copaid=72006556894</td>
<td>moreinfo</td>
</tr>
<tr>
<td>17</td>
<td>127.0.0.1</td>
<td>Query=Human-computer interaction</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>18</td>
<td>127.0.0.1</td>
<td>Query=Human-computer interaction</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>19</td>
<td>127.0.0.1</td>
<td>Copaid=72006556894</td>
<td>moreinfo</td>
</tr>
<tr>
<td>20</td>
<td>127.0.0.1</td>
<td>Query=Human-computer interaction</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>21</td>
<td>127.0.0.1</td>
<td>Query=Human-computer interaction</td>
<td>simpleSearch</td>
</tr>
<tr>
<td>22</td>
<td>127.0.0.1</td>
<td>Copaid=72006556894</td>
<td>moreinfo</td>
</tr>
</tbody>
</table>

20 The whole system is running on a single local machine, hence the loopback IP address of 127.0.0.1. A fuller treatment, not possible in the time available, would add data about who the user is, rather than using an IP address.
Figure 12 Attention store data

The basic compositor-service communications are shown below:

Figure 13: Demonstrator architecture

Screens and transitions supplied by the compositor are:

Figure 14: Demonstrator screens and transitions
The results may be incorporated into specialist social networking software, in the case of the demonstrator, with the Manchester PLE. A loose coupling with the compositor is supplied via delicious social bookmarking integrated into the learning spaces which are part of the PLE. Figure 15 below provides a basic example: In the figure a user is dragging a social bookmark from the left hand panel (the highlit bookmark third down in the panel) onto a media space. Note that the bookmark for the augmented search result of figure 11 is already in the space at the top of the main area.

Figure 15: Compositor integration with a learning environment via social bookmark use

Besides these facilities, the PLE provides social networking and other social software facilities for users. A user profile screen is shown, and is accompanied by the normal social networking features.

---

21 In passing we note that this second version of delicious integration is not as ascetically pleasing as our first version, which we have already trialed to small-sample student acclaim in an earlier production version of our PLE.

22 Albeit with a somewhat poor description “demo for the final report”
A few other useful Web 2.0 facilities are deployed in the social software supplied by the PLEs. A full discussion of PLE functionality is out of scope here; the main aim of the PLE’s inclusion is to demonstrate one form of compositor to specialised social network coupling via del.icio.us, and to mention that the system supplies spaces which are specialised for use by individual learners and groups of learners.

We foresee other forms of coupling. One simple example is that a user who is a member of a community could add content to a social object that has been imported into a community media space. This new information could then become easily accessible by other members of the community via a clickable item that is automatically placed in a group feed when the content is generated. An example feed appears above, in figure 15, for a single user and his friends.
10 Recommendations

10.1 Introduction

This chapter begins with a discussion of the strategic opportunities offered to the IE now. This is followed by our recommendations, which are collated into categories of policy, strategy and operationalisation, architecture and IE components, funding, and Edina and Mimas.

The area of concern

Five topics comprise the area of concern here: The IE and the e-Framework, and their intersection with Web 2.0, Library 2.0, Education / Learning / e-Learning 2.0, and personalisation. This project did not consider the e-Framework, but notes that work proposed here will lead to new SUMs in the context of the high-level domain mapping exercise.

The emphasis here is on enabling new use of the IE by exploiting Web 2.0, social software and personalisation technologies in order to provide a more personalised, collaborative and participatory use of parts of the IE, particularly those currently concerned with the supply of information to users. This is particularly pressing in the light of contemporaneous web developments, possibilities for better user support, and rising user expectations.

The strategic opportunities

A major strategic opportunity for the IE is to enhance the IE to enable:

- Architectures of participation, which are socio-technical systems where the efforts of many users benefits individual users. At a technical level, architectures of participation are supported by social software.
- Support for communities of learners, teachers and researchers.
- Web 2.0 as a platform for learning, teaching and research.
- The transformation of bibliographic, serial and archival data in Edina, Mimas and other services into social objects; objects around which conversations and discussions can form as a result of user engagement with each other over the object's contents. Flickr makes photographs available as social objects; similarly bibliographic, similarly serial and archival data may be made available as social objects.
- Provision of other Library 2.0 facilities around repositories and services which supply bibliographic, serial and archival services.

Any adoption of these will result in a transformation in how IE users employ the IE in their learning, teaching and research activities.

A second major strategic opportunity is to deliver

- Personalised information and service functionality that is relevant, timely, engaging, and easy to consume and use, and that satisfies the information and task-support needs of the end users.

The benefit of successful delivery of these is the provision of facilities which provide greater effectiveness in the use of the IE, assisting the work of learners, teachers and researchers who use the IE.

The question of whether these services would be used by IE users is open, but the benefits of adoption are large. Therefore service implementations should be high quality in order to maximise uptake (e.g. provision a superb user experience, provision of personalisation that truly assists). There is a possibility of

---

23 “High quality” should not be interpreted in the sense of ‘taking 3-5 years to develop and being very highly engineered to meet needs identified (perhaps) at the start, and possibly then failing to support user needs that have moved on.’ Elsewhere we propose working methods appropriate for the development of contemporaneous web services.
minimising the JISC’s subsequent delivery costs (provision of services that can be rolled-out both in the IE and by educational institutions which choose to do so).

10.2 A plan for IE investment

While the collated recommendations contain various funding recommendations (section 10.3.6), we recommend a minimal investment programme for the IE:

- Try out social objects.
- Provide within the IE social networking services to support communities for learning, teaching, and research, such that the services are integrated with and augment IE resources and institutional library systems. Whatever is provided, it is unlikely to be an existing commercial product (e.g. by utilising Facebook applications).
- Augment the IE with synchronous and asynchronous communications facilities that may be used on their own and reused as components in other IE systems, including social networking systems. Wherever possible, make use of facilities that have already been developed and widely adopted.
- Build in personalisation data gathering services that may be used by end-user services to automatically store personalisation data for users. User opt-in and, possibly, opt-out is required.
- Integrate any personalised recommendation technology that is available in the near time frame, and that integrates into the IE.
- Experiment with personalisation for search, recommendation and the delivery of personalised functionality, both by making any easy gains in personalisation (e.g. from segmentation) that may be possible, and by engaging in longer term developments (e.g. by using personalisation data that is automatically collected, and by personalisation delivered by the use of semantic technologies).

10.3 Collated recommendations

Boxed recommendations are more JIIE specific.

10.3.1 Policy

**Recommendation 01:** It is recommended that any IE system holding data for personalisation purposes, must conform to U.K. and European law and it is recommended that the JISC should consult appropriate legal authorities to determine the legal implications for storing data for personalisation purposes.

**Recommendation 11:** It is recommended that the JISC should consider raising the possibility of poor information seeking behaviour and metacognitive skills as an issue in education and work, including in continuing adult education and CPD, with the DCSF, HEFCE, and the BERR. This with a view to joint investigation and possible remedial action.

**Recommendation 19:** We recommend that the JISC should consider whether students should retain continued use of any JISC IE environment enhancements such as social networking systems after leaving education. Some alternatives are: No use after end of education or employment in the academic sector, keeping the account for a few years (choices: active or not) in the hope that ex-students re-enter the wider educational/research system, or let use continue indefinitely to create a national continuing education and research support system. Alternately, the JISC could pass social network support to educational institutions and avoid the question completely, but this would lead to a poorer experience for students who move between institutions or who are registered at more than one institution.

**Recommendation 05:** Because of the centrality of communities in learning, teaching and research, it is recommended that the JISC adopt as part of IE policy provision of tools to support community formation, growth and maintenance.
10.3.2 Strategy and operationalisation

**Recommendation 13:** It is recommended that in the shorter-term IE development of search for production use should concentrate on search without the use of sophisticated personalisation algorithms, but with the use of easily obtainable user data that provides the opportunity for personalisation, e.g., discipline, progression level and location, and as much task-specific support as can be supplied. This will provide early (interim) results in increasing search usability. See also recommendation 14.

**Recommendation 22:** It is recommended that project work in the Web 2.0 and personalisation areas includes beta development, agile management and development methods, small trial use, formative evaluation and iterative improvement; subsequent closed beta testing; benchmarking performance and sizing installations; and improvements to live systems tried on a subset of users before general implementation.

**Recommendation 23:** It is recommended that the JISC adopt and encourage more of a continual beta attitude towards parts of the IE; trying ideas out experimentally. If introduced features don’t work, they should be discarded or improved till they do work, as appropriate. In this it is recommend that the JISC allow for both experimentation and failure in the pursuit of success. The emphasis on continual beta requires particular project practices as, e.g., in recommendation 22.

10.3.3 Architecture and IE services

**Recommendation 02:** It is recommended that any IE system (a) provide exceptionally secure storage of personalisation data in order to mitigate concerns about insecurity of personal data, and (b) deal with users and their personal data in a totally transparent way in order to build JISC and service reputation and thereby mitigate concerns about use of personalisation data.

**Recommendation 21:** It is strongly recommended that the architecture of any integrated components should be open, in that they are capable of inter-working with other Web 2.0 components.

**Recommendation 07:** It is recommended that the JISC provide reusable synchronous and asynchronous recordable Web 2.0 communications tools for video and voice communication. This to support both meetings and communities. A suitable starting point may be provided by the FlashMeeting Project [fm.ea-tel.eu].

**Recommendation 10:** It is recommended that any access management system supported by the JISC must be capable of working seamlessly with OpenID, to maximise the possibilities for Web 2.0 interoperability.

**Recommendation 15:** The JISC-funded Gold Dust project is currently building and testing an adaptive personalisation recommendation system. The intention is to create personal interest profiles for a target user group of 50 volunteers. These profiles will then be mapped to content harvested from various RSS feeds to provide recommendations. It is recommended that the JISC should examine the Gold Dust Project’s (future) outputs to see if they fulfil the requirements for a recommender that could work over services that are provided by the JISC National Data Centres.

10.3.4 Funding

**Recommendation 06:** It is recommended that the JISC establish a strand of activities to address issues of community formation and maintenance in academic settings, and the delivery of a variety of systems to support these. In any programme that is established, system development must be accompanied by evaluation of real use, and results must be collated and developed by a synthesis team.

**Recommendation 08:** The JISC is aware of the need for digital and other literacies to help enable learners, teachers and researchers take full advantage of the Internet in their learning, teaching and research, and is already engaging in the area as shown by, e.g., the Informs Project. As part of this engagement, it is recommended that the JISC fund a programme of work in this area.
**Recommendation 09:** It is recommended that the JISC’s efforts in supporting the work of students, teachers, and researchers should include a programme that funds ethnographic and other sociological studies on user work in learning, teaching and research that reveals design data for future systems, and that draws on past work in Computer Supported Co-operative Work (CSCW) and any other applicable fields.

**Recommendation 12:** It is recommended that the JISC funds development of integrated Web 2.0 based environments for learners, teachers and researchers.

**Recommendation 14:** It is recommended that the JISC should fund exploratory experimental projects in the provision of personalised search. Some interesting approaches to investigate are the use of resource metadata, the use of user supplied data (their own preferences), attention data, diverse social and Web 2.0 approaches, and semantic approaches that, e.g., consider content users provide as exemplars of areas of interest. See also recommendation 13.

**Recommendation 20:** While the DPIE 1 project has provided a survey of widget implementation mechanisms, is recommended that the JISC commission a small higher-level study on the advantages and disadvantages in interfacing IE components to non-IE social networking systems via widgets. The study should include functionality, privacy, and data security as topics of investigation.

### 10.3.5 Edina and Mimas

**Recommendation 16:** It is recommended that Edina and Mimas services that do not supply SRU and/or persistent URLs make plans to supply them as a matter of relative urgency. It is also recommended that all services ensure that their service contents are indexed by the major search engines.

**Recommendation 17:** It is recommended that Edina and Mimas should devote, with JISC funding, significant effort to finding if there are any approaches to catalogue de-duplication, and if not, ascertain what is possible with duplicated records. Worldcat’s success in de-duplication of their union catalogue should be noted; it is believed that this involves, at some stage, considerable computation on a Beowulf cluster.

**Recommendation 18:** It is recommended that, subject to work performed in recommendation 17, serious consideration should be given to establishing a replication free service containing bibliographic and serial information.

**Recommendation 04:** Possibly subject to the outcomes of recommendation 20, because Facebook appears to be many students’ social networking system of choice, it is recommended that appropriate IE services should, wherever possible, supply Facebook applications that interface with the services, but not expose private data to other Facebook applications. There should be careful treatment of what data is shown to users’ friends and group members.

**Recommendation 03:** It is recommended that, where possible, services should gather usage information from their widgets. Besides providing basic usage data (number of installs and number of searches etc performed from the widget), wherever possible information should be collected to produce information that is useful in determining if further facilities are needed, and how they should be designed.
11 References


12 Appendix: Recommendations for demonstrator projects

This appendix provides the original recommendations for small IE exploration and enhancement projects as formulated by Roddy MacLeod and Mark van Harmelen in March 2008. A subset of these suggestions were selected and revised en route to ITT / call.

The suggested demonstrators are in these areas:

- Bookmarking
- Tag recommendation
- Personalised search and recommendation
- IE registration and information improvements to the JISC Website
12.1 Bookmarking enhancements

Intute and bookmarking services are involved in similar types of activity – identifying useful/quality websites.

<table>
<thead>
<tr>
<th>Intute subject specialists</th>
<th>Social bookmarking service users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify websites</td>
<td>Identify websites</td>
</tr>
<tr>
<td>Classify websites and add keywords</td>
<td>Add their own tabs</td>
</tr>
<tr>
<td>Describe websites</td>
<td>Add notes</td>
</tr>
</tbody>
</table>

Above – someone adds a website to their del.icio.us favourites, and may also add a description of the site in the ‘notes’ field.

Above – Intute already contains a description of the website in question.
Further, some social bookmarking services provide bibliographic citation for some pages that are
bookmarked. E.g. Conotea [http://www.connotea.org/guide#autocollection](http://www.connotea.org/guide#autocollection) provides citations for articles
bookmarked at several different e-journals. Bibsonomy provides scrapers and citation input masks to help
manage citation data, and exports citations in BibTeX, EndNote, RDF or HTML formats.

Two Firefox extensions provide related functionality and may offer further ideas for social bookmarking
mashups with bibliographic, archival and repository data: Zotero [http://www.zotero.org/](http://www.zotero.org/) helps collect,
manage, and cite sources using screen scraping technology. Piggybank
[http://simile.mit.edu/wiki/Piggy_Bank_Collect](http://simile.mit.edu/wiki/Piggy_Bank_Collect) uses semantic data in conjunction with screenscraping
techniques.

JISC might invite project proposals that provide social bookmarking facilities of use to learners, teachers
and researchers. The bookmarking facilities should provide the ability to bookmark via data from
preferably two or more JISC bibliographic/serial/archival/repository sources mashing up service supplied
data user supplied data. The system should demonstrate additional functionality, e.g. construction of
citation information for bookmarked sources, construction of reading lists and their export in convenient
forms, possibly to a learning environment. The bookmarking service might be written from scratch, or
mashed up from existing services.

### 12.2 Tag recommendation

Aspects of this topic might be coupled with a bid in the above area, or the topic might form a stand-alone
bid.

Tags can be used in conjunction with many different kinds of artefacts in social software systems, e.g.
users, uploaded documents and bookmarks. When tagging it is desirable to have recommendations as to
meaningful and relevant tags, thereby helping to consolidate the use of a common (user originated)
vocabulary of tags. See, e.g. delicious [http://www.del.icio.us](http://www.del.icio.us) and

Bids are invited for the construction of a demonstration of tag recommender system. Functionality may
cover just social bookmarking recommendation, or recommendations for a wider variety of tag use.
Consideration should be made of the use of service metadata, and the problem of making
recommendations in less well populated systems, e.g. social bookmarking services that are just starting to
be used.

### 12.3 Personalised search and recommendation engines

Bids are invited for investigation of and demonstration of search and recommendation engine technology
that could be incorporated into future systems.

Bidders should describe their approach to search and recommendation, including

- The application area(s) addressed by the
  
  - Search, which returns matching results from one or more collections of resources
  
  - Recommendation, which returns matching results from
    
    - A collection or collections of resources, and/or
    
    - Time-ordered streams of data (e.g. new acquisitions).

- The sources of personalisation data, e.g.:
  
  - The user’s past behaviour
  
  - The contents documents and web pages that have previously been specified by the user as
    containing interesting material

53
Past behaviour by users in the searching user’s social network (see, e.g. Delver
http://www.delver.com/about.htm)

Everyone’s past behaviour

A brief description of the algorithmic approach to be investigated or adopted, including what would
be searched over, e.g. full bibliographic data including metadata

A description of any other salient features, e.g., natural language input for search queries, spelling
suggestions for search terms where appropriate, advanced search features.

An architectural diagram which shows how, and by what protocols the engine would communicate
with particular JISC services.

The demonstrator(s) will need to be search one or more JISC services and make use of sample user data
to demonstrate realistic use.

12.4 IE registration and information improvements to the JISC website

Any bids for topics in this area need to demonstrate how the resultant work can be easily imported into the
JISC website.

Registration for JISC Services

A registration facility for all JISC services that allow registration from the JISC website, i.e. one form at the
JISC website where users can enter their information, and click on any JISC services for which they want
to register (default would be all relevant JISC services). Users should be able to return to the JISC site to
update registrations and registration information. Besides registration for services like Intute, JORUM and
Zetoc, users should be able to control JISC mailing list memberships from this facility, and acquire JISC
and service RSS feeds for their favourite feed reader.

A starting point could be:
Sample form fields for registration:

<table>
<thead>
<tr>
<th>Name</th>
<th>Email address</th>
<th>password</th>
<th>Athens</th>
<th>Lecturer/student/etc</th>
</tr>
</thead>
</table>

- JORUM ✓
- Zetoc ✓
- Intute ✓

Surface JISC services’ content at JISC website

One idea in this area would be similar in concept to LibWorm http://www.libworm.com/rss/librarianqueries.php LibWorm collects updates from about 1400 (and growing) RSS feeds. The contents of these feeds are then available for searching, and search results can themselves be output as an RSS feed that the user can subscribe to either in his/her favourite aggregator or in LibWorm’s built-in aggregator.

There are at least two different types of RSS feeds from services:

- Service news including from service blogs
- Additions to content at services.

These might be need to be treated differently.

Bidders might consider how to surface JISC related content at the JISC website, e.g. JISC-funded projects’ RSS feeds.