

Developing personalisation for the  
Information Environment (2)  
Report 1: landscape study



## Developing personalisation for the Information Environment (2)

# Report 1: landscape study

CC319D002-1.0

22 February 2008

Cover + 46 pages

**Dr Max Hammond**  
**Matt Shreeve**

curtis+cartwright 

**Curtis+Cartwright Consulting Ltd**

Main Office: Surrey Technology Centre,  
Surrey Research Park, Guildford  
Surrey GU2 7YG

tel: +44 (0)1483 295020  
fax: +44 (0)1483 295021  
email: [postmaster@curtiswright.co.uk](mailto:postmaster@curtiswright.co.uk)  
web: <http://www.curtiswright.co.uk>

Registered in England: number 3707458

Registered address:  
Baker Tilly, The Clock House,  
140 London Road, Guildford,  
Surrey GU1 1UW



## Executive summary

### *General*

- 1 Curtis+Cartwright Consulting Limited has been commissioned to scope how the JISC Information Environment (IE) may make use of personalisation in order to enhance the user experience, in a contract letter dated 2 November 2007.
- 2 This document is the project interim report, and contains the results obtained during the first phase of the work – the landscape study.
- 3 This version of the document (1.0) is the final release version.

### *Background*

- 4 Most major Internet sites use personal information to tailor the way that their sites are presented to users. The JISC provides or funds a range of services which are available online, and wishes to investigate how these services may be personalised for users. Many already offer some degree of customisation (where a user may choose elements of functionality to be displayed, for example), but there may be opportunities to personalise sites adaptively in a way which is transparent to the user. This may be based on information which is already known about the user (*eg* the institution they attend) or on information obtained by tracking user behaviour (*eg* which types of books they search for in a library catalogue).
- 5 The JISC has funded a number of efforts which are of relevance to the development of personalised services. These include a range of projects investigating specific concerns, and more broadly, the IE programme and the e-Framework. These frameworks help promote interoperability within the education and research communities, although neither of these currently has specific advice or specifications regarding the provision of personalised services.

### *Objectives*

- 6 This study will scope how the JISC IE may make use (or should be developed to make use) of personalisation in order to enhance the user experience. It will specifically consider the ways in which infrastructure established to support the UK Access Management Federation (the UK federation) could support adaptive personalisation of JISC services, and the potential privacy and legal barriers to such use.

### *Approach*

- 7 This document is an interim report; it presents the results of the first phase of the project (the landscape study), and serves as a foundation for the remainder of the project. The audience for this document is wide; it includes the JISC, the JISC content services, institutions, advisory services and further.
- 8 This document covers the landscape of personalisation of online resources for education and research. It has been written based on a review of available literature, and on interviews with key stakeholders from JISC services, institutions, JANET(UK) and other interested parties. It is intended to be of interest across a range of stakeholders; it is not a technical discussion of approaches to personalisation, although some sections do discuss how the technology available could be used.

- 9 The remainder of this project consists of two phases:
- **Foresight:** this phase of the study will investigate how the environment surrounding privacy and personalisation may develop in future.
  - **Analysis and reporting:** based on the results of the diagnostic and prognostic elements of this study, a range of use-cases for personalised services will be generated.

### *Terminology*

- 10 This report uses the terminology of previous JISC work in the area: customisation, Adaptive Personalisation based on Data held elsewhere (APOD), and Adaptive Personalisation based on User Activity (APUA). In addition, we distinguish between Personal Applications (which are personal due to their fundamental nature – they contain information about the user), Portals (which aggregate information for a user), customisation and adaptive personalisation. This forms a *personalisation spectrum*.

### *Web 2.0*

- 11 The “Web 2.0” shift has led to a greater demand to make services personal to their users. One of the key concepts is that gathering information about the behaviour of users can be used to make the site more relevant. It is relevant to consider how this is applied in a commercial environment, for example by suggesting alternative products based on the purchasing decisions of other users.

### *Federated access management*

- 12 Federated access management is an approach to authentication and identity management which is rapidly growing in importance worldwide. In the UK, the JISC is leading this movement, and is funding the UK Access Management Federation, managed by JANET(UK) with technical elements provided by SDSS, based at Edina.
- 13 The UK federation has adopted the Shibboleth technology, which is a profile of the Security Assertion Markup Language (SAML). This technology has the capability to deliver any information (known as attributes) held by a user’s Identity Provider (IdP) to a Service Provider (SP), subject to the Attribute Release Policy (ARP) of the IdP.
- 14 In order to transmit attributes, a schema (a structure to contain data) must be agreed. None of the schemata in use fully represent the profile of a learner (which may be required for complex APOD), and there is currently no significant effort to develop such schemata.

### *Current use of personalisation*

- 15 The personalisation currently undertaken by JISC content services is almost exclusively customisation, rather than adaptive personalisation. A range of techniques are used (including saved alerts, saved searches, bookmarking favourite resources *etc*); there is no common approach or infrastructure. Some services require users to enrol separately before having access, in order to comply with the terms of their licenses.
- 16 Services have several current concerns regarding personalisation, including linking identities from Athens and IdPs, and identifying users who currently use IP-based access or consume the content through Really Simple Syndication (RSS) feeds or other machine-to-machine interfaces.

- 17 The nature of personal services demands that the service providers intimately understand user behaviour: the current approach to funding and monitoring JISC services seems unlikely to provide an environment conducive to significant investment in this area (see 6.1.4-6.1.6).
- 18 The large academic publishers are beginning to transform from traditional publishers into information services companies. As part of this, they are very interested in opportunities to personalise their products, and to ensure that they fit well with the workflow of users. The many smaller publishers may not be well-suited to producing personalised services (see 4.2.8 and 4.3.7).
- 19 A number of institutional systems employ some personalisation, key amongst these being Virtual Learning Environments (VLEs), institutional portals, and library catalogue systems. The personalisation undertaken is primarily customisation or portalisation. There are, however, some examples within the library community of developing catalogues using personalisation techniques which are familiar from online commerce sites (*eg* suggesting other books based on other users' borrowing habits).

#### *Legal issues*

- 20 Developing personalised services is – by definition – liable to involve personal information. This information may be collected directly from users, either explicitly for customisation or transparently in APUA, or transferred from another source, as in APOD. Within the UK, any of these operations is liable to fall within the scope of, and must therefore comply with the Data Protection Act 1998 (the DPA). This legislation places restrictions on when and how personal information can be processed (including disclosure to third parties).
- 21 For the purposes of personalisation, current guidance suggests that many transfers of personal attributes from IdP to SP, and much data recorded by SPs which could be of use for APUA is likely to fall within the act (*Nb:* the contents of this report are not, and are not intended to be, legal advice). The DPA does not make transfer of such information impossible, but user consent is likely to be required before such transfers can take place.

#### *Challenges and opportunities*

- 22 The UK federation provides an infrastructure which could support personalised services, subject to the issues regarding data protection legislation, and appropriate schemata (above).
- 23 The challenges and opportunities presented within this report are a preliminary set, which will be developed further by later stages of this project. A discussion of these possibilities is presented at Section 6.
- 24 There are significant challenges to the development of personalised services, including: unclear demand from users or service providers; lack of motivation to significantly develop services; understanding the resources, as well as the users; and nebulous ideas of privacy.
- 25 A number of initial ideas for personalised services have been elicited, including: automatic form-filling or enrolment; improved management information to develop services; group membership defined by a home institution; and linking to a user's VLE.

#### *Summary of landscape study*

- 26 Current use of personalisation by online services for education and research is limited. Many services offer some ability for customisation, but the type and mechanisms of customisation vary. Within the JISC content services there is no common approach. The situation is similar

amongst commercial services, although some large academic publishers are beginning to develop APUA systems.

- 27 There is demand for APUA amongst JISC content services, but there is currently no real demand for APOD. The views of users are currently unknown; the second phase of this project aims to provide further information in this regard.
- 28 The UK federation provides an infrastructure which can support customisation and adaptive personalisation, although it would be necessary to develop appropriate schemata for APOD.
- 29 Several potential opportunities for development have been identified, but there are significant challenges to be addressed before extensively personalised services can be developed. The remainder of this project will consider how to mitigate these challenges, and how to develop the opportunities into proposals for JISC funding.

## Document history

Version	Date	Description of Revision
0.1	11 February 2008	Draft for internal review
0.2	13 February 2008	Draft for review by the JISC
1.0	22 February 2008	Issue version following review by JISC

This page is intentionally blank

## List of contents

<b>Executive summary</b>	<b>1</b>
<b>Document history</b>	<b>5</b>
<b>List of contents</b>	<b>7</b>
<b>List of abbreviations</b>	<b>9</b>
<b>1 Introduction</b>	<b>11</b>
1.1 General	11
1.2 Background	11
1.3 Objectives	11
1.4 Scope	11
1.5 Approach	12
1.6 Timeline	12
1.7 Terminology	13
1.8 Overview of this document	13
<b>2 Background</b>	<b>15</b>
2.1 Introduction	15
2.2 The rise of personalised services	15
2.3 The personalisation spectrum	15
2.4 Web 2.0 and personalisation	16
2.5 Federated access management	17
2.6 User-centric identity	18
2.7 JISC activities	19
<b>3 Technologies and standards</b>	<b>21</b>
3.1 SAML and Shibboleth	21
3.2 Schemata	22
3.3 The e-Framework	24
<b>4 Current examples of personalisation</b>	<b>25</b>
4.1 Introduction	25
4.2 JISC content services	25
4.3 Publishers	27
4.4 Institutional systems	28
<b>5 Legal landscape</b>	<b>31</b>
5.1 Legal issues	31
5.2 UK federation policy and guidance	32
<b>6 Challenges and opportunities</b>	<b>35</b>
6.1 Challenges to personalisation	35
6.2 The potential of the UK federation	37
6.3 Initial opportunities	38
6.4 Summary of landscape study	39
<b>A List of interviewees</b>	<b>41</b>
<b>B An introduction to UK data protection legislation</b>	<b>43</b>

B.1	Introduction	43
B.2	The data protection principles	43
B.3	Conditions for processing	43
B.4	What is personal data?	44
B.5	Sensitive personal data	45

## List of abbreviations

API	Application-Programme Interface
APOD	Adaptive Personalisation based on Data held elsewhere
APUA	Adaptive Personalisation based on User Activity
CM	Core Middleware
CMS	Course Management System
DEST	The Australian Department of Education, Science and Training
DPA	The Data Protection Act 1998
ELF	e-Learning Framework
FE	Further Education
HE	Higher Education
HR	Human Resources
HTTP	Hypertext Transfer Protocol
ICO	Information Commissioner's Office
IdP	Identity Provider
IE	Information Environment
IMS eP	IMS e-Portfolio
ITT	Invitation to Tender
JISC	Joint Information Systems Committee
LIP	Learner Information Profile
M2M	machine-to-machine
OASIS	Organisation for the Advancement of Structured Information Standards
OPAC	Online Public Access Catalogue
RSS	Really Simple Syndication
SAML	Security Mark-up Assertion Language
SCQF	Scottish Credit and Qualifications Framework
SOA	Service Oriented Architecture
SP	Service Provider
SRS	Student Records System
SUM	Service Usage Model
TOC	Table of Contents
UGC	User-Generated Content
VLE	Virtual Learning Environment
XCRI	eXchanging Course Related Information
XML	eXtensible Mark-up Language

This page is intentionally blank

# 1 Introduction

## 1.1 General

- 1.1.1 Curtis+Cartwright Consulting Limited has been commissioned to scope how the JISC Information Environment (IE) may make use of personalisation in order to enhance the user experience, in a contract letter dated 2 November 2007.
- 1.1.2 This document is the project interim report, and contains the results obtained during the first phase of the work – the landscape study.
- 1.1.3 This version of the document (0.2) is issued for review and comment by the JISC prior to broader circulation.

## 1.2 Background

- 1.2.1 Most major Internet sites use personal information to tailor the way that their sites are presented to users. The JISC provides or funds a range of services which are available online, and wishes to investigate how these services may be personalised for users. Many already offer some degree of customisation (where a user may choose elements of functionality to be displayed, for example), but there may be opportunities to personalise sites adaptively in a way which is transparent to the user. This may be based on information which is already known about the user (*eg* the institution they attend) or on information obtained by tracking user behaviour (*eg* which types of books they search for in a library catalogue).

## 1.3 Objectives

- 1.3.1 This study will scope how the JISC IE<sup>1</sup> may make use (or should be developed to make use) of adaptive personalisation in order to enhance the user experience. It will specifically consider the ways in which infrastructure established to support the UK Access Management Federation (the UK federation) could support adaptive personalisation of JISC services, and the potential privacy and legal barriers to such use.
- 1.3.2 This document is an interim report; it presents the results of the first phase of the project (the landscape study), and serves as a foundation for the remainder of the project. The audience for this document is wide; it includes the JISC, the JISC content services, institutions, advisory services and further.

## 1.4 Scope

- 1.4.1 This document covers the landscape of personalisation of online resources for education and research. It has been written based on a review of available literature, and on interviews with key stakeholders from JISC services, institutions, JANET(UK) and other interested parties. It is intended to be of interest across a range of stakeholders; it is not a technical discussion of approaches to personalisation, although some sections do discuss how the technology available could be used.

---

<sup>1</sup> <[http://www.jisc.ac.uk/whatwedo/themes/information\\_environment.aspx](http://www.jisc.ac.uk/whatwedo/themes/information_environment.aspx)> [accessed 12 October 2007].

## 1.5 Approach

1.5.1 For this landscape study, there have been two main data gathering activities:

- A review of:
  - published reports and papers of relevance;
  - published and developing standards of relevance to any area of the project, including attribute transfer, identity management and privacy control;
  - the current and potential use of personalisation by JISC services, and current development activity in the area;
  - current and recent JISC-funded or other tools to support customisation or personalisation of Internet resources.
- Interviews to solicit information and opinion from:
  - JISC national datacentres;
  - institutions considered to be advanced in the development of personalised services;
  - other individuals and organisations of relevance.

1.5.2 A full list of interviewees is included at Annex A.

1.5.3 The remainder of this project consists of two phases:

- **Foresight:** this phase of the study will investigate how the environment surrounding privacy and personalisation may develop in future. There will be two elements: a study to investigate how current HE and FE students regard privacy and self-disclosure online, and the development of some scenarios to consider alternative possible futures, and how personalisation could be used in each future.
- **Analysis and reporting:** based on the results of the diagnostic and prognostic elements of this study, a range of use-cases for personalised services will be generated. Working with interested stakeholders, these use-cases will be developed into detailed proposals for demonstrators which could be developed with JISC support.

## 1.6 Timeline

1.6.1 Key dates in the project are as follows:

- November 2007 – February 2008: landscape study.
- February 2008: focus groups to investigate student perceptions of privacy.
- March – April 2008: scenarios-building exercise to investigate potential futures for the information environment.
- April – June 2008: developing use-cases and specifying demonstrators.
- 26 June 2008: final report.

## 1.7 Terminology

1.7.1 There is not a consistent set of terminology in use to describe the various ways in which services can be personalised. For consistency, we have adopted the vocabulary of the prior JISC-funded work in this area:<sup>2</sup>

*"Types of Personalisation:*

*Customisation (Explicit or Referential Personalisation) - The selection of options is under the direct control of the user who explicitly chooses to include or exclude options.*

*Adaptive Personalisation (Implicit or Inferential Personalisation) - The availability of options is based on knowledge about users gained from tracking user activity and/or other sources of user information. The system identifies items of potential interest to the user and controls what is made available to the user. Note, this form of personalisation may involve varying degrees of user awareness of, and involvement in, the process. Adaptive Personalisation is usually based on one of the following two technologies or approaches:*

- *Adaptive Personalisation based on User Activity (APUA) e.g. Collaborative Filtering – an algorithm which allows a service to identify items of potential interest to a particular user based on the preferences of other users with similar characteristics and/or activity records (e.g. <<http://www.amazon.com>>).*
- *Adaptive Personalisation based on Data held elsewhere (APOD) e.g. Rules Based Filtering – a database driven system based on preset rules about relationships between items and user profiles. The user profiles may be created by the user, based on tracking user activity or based on data held elsewhere (e.g. student records, staff management information system or personnel record.)"*

## 1.8 Overview of this document

1.8.1 The rest of this report is set out as follows:

- Section 2 contains background information on the personalisation of internet resources;
- Section 3 sets out relevant technologies and standards;
- Section 4 discusses current examples of personalisation;
- Section 5 addresses some potential legal issues;
- Section 6 presents some initial ideas regarding potential challenges and opportunities for personalisation;
- Annex A lists the interviews conducted for this phase of the project;
- Annex B provides a brief overview of the data protection legislation in the UK relevant to this project.

---

<sup>2</sup> *Personalisation in presentation services – a follow-up report for the JISC, Neil Smith, Nicky Ferguson and Seb Schmoller, 29 September 2006.*

This page is intentionally blank

## 2 Background

### 2.1 Introduction

2.1.1 This section contains some high-level thematic information relevant to the personalisation of online information services. It is not intended to be a comprehensive review of prior work in the area, but rather to identify the key themes and activities. More detailed information is included in later sections, where appropriate.

### 2.2 The rise of personalised services

2.2.1 The past decade has seen phenomenal growth in the use of the internet for education, research, commerce and for personal communication. Although business information has been collected in the commercial arena for many years, Internet services present the opportunity to directly utilise information about each individual user in order to tailor the interaction that that user has with the service.

2.2.2 The option for a user to customise sites has been available for some time; the option to generate a persistent "user account" is important in establishing a customer relationship for e-commerce sites. User accounts are typically believed to yield improved service to the users as well as facilitating improved customer analyses by the vendor.

2.2.3 A new aspect of internet services is the concept of "attention data" – recording and analysing what users appear to be interested in, and using this information to influence the presentation of the service to its user (known as APUA in the JISC context). Attempts to investigate the phenomenon of attention have previously been undertaken using traditional market research tools, taking a small sample of customers or potential customers and investigating these individuals in detail. Internet resources have the ability to investigate the behaviour of every visitor to the site, and to use that information immediately to alter the delivery of their service.

2.2.4 There has been a gradual adoption of APUA throughout e-commerce, as seen clearly on sites such as Amazon and eBay. Although customisation features are now widely-established on educational resources, the uptake of more complicated APOD or APUA facilities has been slow.

### 2.3 The personalisation spectrum

2.3.1 It is necessary to consider the different requirements for personalising information services, and to this end we use the concept of a "personalisation spectrum":

- **Personal applications.**<sup>3</sup> At one end of the spectrum are applications which are fundamentally personal. Examples might include the user area of a library OPAC, access to a user's own staff or student records, access to their timetable etc. These applications contain data about the user themselves, and allow access to it. Within this region are systems which permit different users different functions based on their role, as access-control is a fundamental property of the application.

---

<sup>3</sup> In this context, "application" refers to the tool which the user interacts with in order to achieve a desired outcome. It need not be a single programme, or even a single system.

- **Portalisation.** These applications collate personally-relevant information into a personal area. These represent a personal service, by presenting in one place as much personally-important information as possible.
- **Customisation.** This is where a user explicitly alters the behaviour of an application, see above (subsection 1.7).
- **Adaptive personalisation.** This is where an application tailors its behaviour to meet the perceived need of the user, see above (subsection 1.7).

2.3.2 The borders between these regions are hazy, for example most portal applications allow some degree of customisation, and some allow adaptive personalisation. It is nonetheless important to distinguish between services which are personal because they are about the user, and those which can be made personal by adapting a service which could be provided impersonally (eg a bibliographic database).

## 2.4 Web 2.0 and personalisation

2.4.1 An important driver for the development of personalised information services is the current trend toward “Web 2.0”. This is a set of related concepts, which together lead to a far more interactive web, where users contribute ideas, opinion, information and content, as well as consuming it. The history, concepts and technologies of Web 2.0 are well-addressed by Paul Anderson’s *JISC Technology Watch* paper (February 2007).<sup>4</sup>

2.4.2 Web 2.0 is frequently seen as a rise in User-Generated Content (UGC), but this is a simplification. Although UGC such as blogs, reviews, Wikipedia, Flickr and YouTube, and social networking sites such as Facebook, Bebo and MySpace are doubtless important in terms of general awareness, much of the power of the Web 2.0 concepts is in the benefits that they bring to service providers and vendors. As more organisations are becoming aware of the potential of these concepts to improve both the user experience and the bottom line, they are becoming more widely deployed.

2.4.3 The focus of this paper is the use of personalisation in an educational context, but as the most successful and widespread use of personalisation so far has been in a commercial context, it is instructive to consider these uses. Perhaps the most commonly-cited example of “Web 2.0” selling is the way by which Amazon tailors its output dependent on the attention data which it collects for each user. There are two main approaches taken, both based on a user’s activity:<sup>5</sup>

- **Intrinsic relationships:** based on properties of the product. For example, after a user buys a CD by The Rolling Stones, new releases by the Stones are recommended to that user. Further criteria can be used, eg type of music (rock & roll), media (CD or DVD), price, etc.
- **Extrinsic relationships:** based on the actions of others. An example of this is “Customers who bought this item also bought” on Amazon. This makes a suggestion as to what a user may find interesting based on that user’s behaviour (looking at a product), and on the behaviours of other users. The inference is that the user is likely to have similar interests to those who have bought that product before, and is thus likely to share an interest in the other products that those previous users have bought.

---

<sup>4</sup> *What is Web 2.0? Ideas, technologies and implications for education*, Paul Anderson, February 2007.

<sup>5</sup> This is in addition to traditional market segmentation (perhaps consumer-intrinsic suggestion), whereby properties of the consumer (eg gender, age, marital status) are used to create profiles of consumers, who can be targeted with appropriate promotional activity.

- 2.4.4 A more subtle demonstration of intrinsic market segmentation is that employed by Pandora,<sup>6</sup> where users can set up music “stations” based on tracks or artists they choose (*eg* making a station based on a Bon Jovi track may result in one which plays primarily melodic rock in a minor key at a moderate tempo with a male vocalist and guitar riffs). This depends on a resource called the Music Genome Database, which contains a large database of music which has been manually categorised according to a large set of attributes. Users can tune these stations with “thumbs-up” and “thumbs-down” buttons, which influence the selection of future music on that station.

## 2.5 Federated access management

### *Introduction*

- 2.5.1 Federated access management “builds a trust relationship between Identity Providers (IdP) and Service Providers (SP). It devolves the responsibility for authentication to a user’s home institution, and establishes authorisation through the secure exchange of information (known as attributes) between the two parties.”<sup>7</sup>
- 2.5.2 There is a gradual worldwide move toward adoption of federated access management across the education and research sectors, and the JISC is currently leading this movement within the UK. The JISC has withdrawn financial support for the previous Athens system from August 2008.
- 2.5.3 The JISC ran two development programmes between April 2004 and March 2006 to plan and prepare for the new infrastructure. These Core Middleware (CM) programmes established the early elements of the infrastructure and generated lessons for institutions in the future. The CM programmes covered internal, third party, inter-institutional and ad hoc collaboration uses of federated access management.
- 2.5.4 Although the Shibboleth technology supported by the UK federation was designed around the concept of anonymous or pseudonymous access to resources, it has the potential to transfer any data held by the IdP in association with the user to the SP. In most cases not all information held will be desired or required by the SP.

### *The UK Access Management Federation*

- 2.5.5 The UK Access Management Federation<sup>8</sup> (“the Federation”) was launched by the JISC in November 2006. The Federation is managed by JANET(UK), with technical elements provided by SDSS, based at Edina.
- 2.5.6 Membership of the Federation is steadily increasing, both by institutions and by SPs. It is unclear what proportion of institutions will eventually join the federation, as opposed to managing access management with individual SPs, or by procuring the Athens system.

### *The Liberty Alliance*

- 2.5.7 The Liberty Alliance<sup>9</sup> is an influential group, with a membership drawn from a broad spectrum of stakeholders across government, industry and academia. Its aims are to “establish open

---

<sup>6</sup> See <<http://www.pandora.com>>. At the time of writing, this service is not accessible in the UK due to issues regarding payments of royalties [Accessed 29 Jan 2008].

<sup>7</sup> <[http://www.jisc.ac.uk/whatwedo/themes/access\\_management/federation/federation\\_faq.aspx#2](http://www.jisc.ac.uk/whatwedo/themes/access_management/federation/federation_faq.aspx#2)> [accessed 12 October 2007].

<sup>8</sup> <<http://www.ukfederation.org.uk>> [accessed 29 Jan 2008].

standards, guidelines and best practices for identity management". The Liberty Alliance has a strong focus on supporting the development of policy and interoperability, rather than focusing on the technological aspects.

2.5.8 The Liberty Alliance does not explicitly consider personalisation, but is focused on identity management, and the associated areas of interoperability, privacy and security. The technological and policy guidance from the Liberty Alliance is discussed later, as appropriate.

### **OpenID**

2.5.9 OpenID<sup>10</sup> is a project to develop independent identity-management systems and standards. Whereas the Liberty Alliance and Shibboleth technologies have been developed around enterprise requirements to protect access to valuable resources, OpenID has principally been intended as a relatively-lightweight and user-friendly infrastructure for "open" internet resources.

## **2.6 User-centric identity**

2.6.1 There has been a gradual shift in the way that identities are managed and asserted, and this shift is especially clear within the UK Further and Higher Education (FE and HE) sectors. The key stages to-date have been:

- 1) **Service-centric identity:** users must enrol with each individual service they wish to use, and each service maintains its own authentication system.
- 2) **Centralised identity:** *eg* the Athens system. A central service maintains identity information which is shared between service providers.
- 3) **Institution-centric identity (federated identity):** *eg* the UK Access Management Federation. Identity and authentication is undertaken by an identity provider (which in the case of UK FE and HE is their home institution).

2.6.2 Now, however there is a move – in thinking, at least – toward user-centric identity. This is the idea that an online identity is owned by the user themselves, and that every identity transaction should involve the user's agent. This is a deeper level of involvement than simply having control over their identity, or consenting to the use of their identity, and is sometimes known as "having the person in the protocol".

2.6.3 Discussions as to the utility of user-centric identity are somewhat hampered by the difficulty in defining identity, especially electronic identity. Further, there are likely to be some situations where it is not feasible or desirable for the user to own their own identity (as a user of the system).

---

<sup>9</sup> <<http://www.projectliberty.org>>

<sup>10</sup> <<http://www.openid.net>>

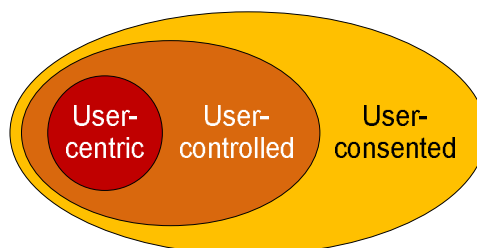


Figure 2-1: user-centric identity<sup>11</sup>

## 2.7 JISC activities

- 2.7.1 This sub-section sets out some of the key strategic efforts which have been undertaken by the JISC, and briefly discusses projects of particular relevance. Operational-level information (technologies and the personalisation which is currently undertaken) is discussed in sections 3 and 4.

### **Personalisation**

- 2.7.2 "Personalisation" has been a theme of JISC activity for several years. A wide range of projects have been funded, across a range of definitions and uses of personalisation. The previous *Personalisation in presentation services* report<sup>12</sup> contains a good summary of the work which had been conducted up to 2004.
- 2.7.3 The current "personalising technologies" stream within the users and innovation programme<sup>13</sup> consists primarily of projects investigating the "personal application" end of the personalisation spectrum (see p.15) within an e-Admin context. The exception is the ticTOCs project,<sup>14</sup> which aims (amongst other things) to allow aggregation and customisation of Table of Contents (TOC) RSS feeds based on user customisation. ticTOCs will harvest usage data, and the current Gold Dust project<sup>15</sup> aims to exploit this "Personal Interest Profile" alongside text-mining and pattern-matching techniques in order to identify articles of particular relevance to individual users.
- 2.7.4 The current Developing Personalisation for the Information Environment (1) project is investigating opportunities for utilising Web 2.0 concepts (and especially social networking) within the IE.<sup>16</sup>

### **The Information Environment**

- 2.7.5 The JISC Information Environment (IE) presents a model by which content-providers are linked to users through "fusion" and presentation layers.<sup>17</sup> The IE has an associated technical

<sup>11</sup> Diagram adapted from <<http://connectid.blogspot.com/2006/06/prototol-for-people.html>> [accessed 4 February 2008].

<sup>12</sup> *Personalisation in presentation services - a report commissioned by the JISC*, Neil Smith, Nicky Ferguson, Seb Schmoller, July 2004.

<sup>13</sup> <[http://www.jisc.ac.uk/whatwedo/programmes/programme\\_users\\_and\\_innovation.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_users_and_innovation.aspx)> [accessed 1 February 2008].

<sup>14</sup> <[http://www.jisc.ac.uk/whatwedo/programmes/programme\\_users\\_and\\_innovation/tictocs.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_users_and_innovation/tictocs.aspx)> [accessed 1 February 2008].

<sup>15</sup> <<http://www.hull.ac.uk/golddust/index.html>> [accessed 1 February 2008].

<sup>16</sup> A website for this project could not be located.

architecture which specifies the protocols which services within the IE should use to deliver and consume information, albeit at a very high level (essentially setting out some common protocols such as HTTP/1.1 and Z39.50). There is no specification as to how to structure or present data.

- 2.7.6 The IE is essentially a conceptual framework within which the JISC has focused its efforts, and provides a vocabulary as to the role of various projects, components and services.

### ***The e-Framework***

- 2.7.7 The IE presents a high-level picture of service interactions and protocols, but does not in itself deliver workflows and models which will lead to the development of real-world services. The focus within JISC for developing service-orientated approaches within the sector is the e-Framework for Education and Research, jointly sponsored by JISC, the Australian Department of Education, Science and Training (DEST), New Zealand's Ministry of Education (NZMoE), and the Netherlands' SURF.
- 2.7.8 The JISC-managed e-Learning Framework (ELF) began work in 2003, to help systematise the emerging service-oriented approach to software engineering, within e-Learning. The ELF approached the problem by defining sets of User Agents, Learning Domain Services and Common Services. These "services"<sup>18</sup> were used as a filing system to categorise JISC research projects, and any appropriate implementations or specifications. The ELF has been extended to encompass a wider scope than originally foreseen, was re-launched as the e-Framework, and now incorporates the international partners. The e-Framework is currently being populated, with a number of projects currently generating Service Usage Models (SUMs), and others providing high-level support.<sup>19</sup>
- 2.7.9 The relevance of the e-Framework for personalisation is discussed at sub-section 3.3.

---

<sup>17</sup> <<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/>> [accessed 10 February 2008].

<sup>18</sup> The term "service" is ambiguous, and the e-Framework has now developed a clear vocabulary. The term "service", for the purposes of the e-Framework, refers to a technical service that promotes interoperability between systems. The "services" within the e-Learning framework encompasses service genres, service expressions, and even entire domains within the new terminology.

<sup>19</sup> Including HiLDA, and the Reference Models: Synthesis of Sources projects.

## 3 Technologies and standards

### 3.1 SAML and Shibboleth

3.1.1 The key protocol within the UK federation is Shibboleth, which is a profile of the Security Mark-up Assertion Language (SAML). Our previous report (for our evaluation of the JISC Core Middleware programmes) contains the following explanations, which have been slightly updated for this report.<sup>20</sup>

#### **SAML**

3.1.2 SAML is a flexible eXtensible Mark-up Language (XML)-based framework,<sup>21</sup> developed by the Organisation for the Advancement of Structured Information Standards (OASIS), to allow the exchange of identity data between various domains on the basis of a message known as an "assertion".<sup>22</sup> Assertions contain authentication, authorisation or attribute information.

3.1.3 The OASIS Security Services Technical Committee is responsible for defining, enhancing, and maintaining the specifications that define SAML. The latest SAML standard is SAML V2.0, approved as an OASIS standard in March 2005.

3.1.4 SAML has gained widespread industry adoption as a basis for federated identity and security environments and is being applied in a number of different ways, for example:

- **Web Single Sign-On (SSO):** SAML enables web SSO through the communication of an authentication assertion from one system to another which, with appropriate trust relationships, can log in the user as if they had authenticated directly.
- **Securing web services:** SAML assertions can be used within Simple Object Access Protocol (SOAP) messages in order to carry security and identity information between actors in web service transactions.

3.1.5 SAML V2.0 is an important step from V1.1, representing convergence between the SAML and the Liberty Alliance ID-FF standard, and incorporating input from the Shibboleth project.

#### **Shibboleth**

3.1.6 The term Shibboleth in this context refers to three things:

- The **Shibboleth project**<sup>23</sup> is an Internet2 project which is developing architectures, policy structures, practical technologies, and an open source implementation to support inter-institutional sharing of web resources subject to access controls. The Shibboleth project controls the development of the Shibboleth profile and Shibboleth technology.
- The **Shibboleth profile**<sup>24</sup> is a SAML profile which defines the protocols for the exchange of identity information. The current Shibboleth profile is based on SAML v1.1.

---

<sup>20</sup> *Federated access management: international aspects*, CC253D018-1.0, 7 June 2007, Dr Claire Davies and Matt Shreeve.

<sup>21</sup> SAML consists of a set of specifications and XML schemas which define how to construct, exchange, consume, interpret, and extend security assertions.

<sup>22</sup> An assertion is a statement by the identity provider about the user, and is used by the service provider to determine whether the user may be given access to a requested service.

<sup>23</sup> <<http://shibboleth.internet2.edu/>>

<sup>24</sup> Also referred to as the Shibboleth architecture.

- The **Shibboleth technology**<sup>25</sup> is one implementation of the Shibboleth profile, which permits the exchange of identity information between IdPs and SPs to allow access to web-enabled resources using a federated administration.<sup>26</sup> This implementation is open source, but is currently being developed in a controlled manner by Internet2.

### **Attribute release policy**

- 3.1.7 An important concept within Federated Access Management is the utilisation by an IdP of an Attribute Release Policy (ARP), which specifies which attributes may be released to which SPs under which conditions.

## **3.2 Schemata**

- 3.2.1 In order to represent information about a user (or indeed any other object), it is necessary to define a schema to structure this information. In order to share information with other organisations, it is necessary to agree on a common schema, or to generate a mapping between the schemata in use.

### **Identity information - eduPerson**

- 3.2.2 Most institutions have some form of directory service, which contains a common store of information regarding users of information systems, and may be used further, for example in student records, finance *etc.* Appropriate schema should contain elements which are of use in typical directory structures, such as real name, nickname, email address, relationship with institution. The eduPerson schema<sup>27</sup> is used commonly for this purpose, and was developed specifically to represent members of HE institutions in the United States.<sup>28</sup>
- 3.2.3 Although designed explicitly for HE, eduPerson is probably well-suited for representing members of FE institutions and schools. eduPerson is the schema in use by the UK federation, and the specific details of that implementation are discussed below.
- 3.2.4 Whereas eduPerson (and other similar identity schemata) provides adequate information to identify an individual and to understand the organisational membership of an individual, it cannot represent more specific information such as is typically stored within a Student Records System (SRS) or payroll system, for example.<sup>29</sup>

### **The UK federation core attributes**

- 3.2.5 The UK federation mandates IdPs to be capable of providing four eduPerson attributes:
- **eduPersonScopedAffiliation.** This attribute describes the relationship between a user and their IdP. Examples include `staff@somewhere.ac.uk`, `student@somewhere-else.ac.uk` *etc.* Users can have several values (*eg* most students and staff are likely to also be members, whereas alumni are unlikely to be).

<sup>25</sup> Also referred to as the Shibboleth software or code package.

<sup>26</sup> Where SPs leave the administration of user identities and attributes to a user's IdP. The IdP is also responsible for providing user information attributes to SPs.

<sup>27</sup> <<http://www.educase.edu/eduperson/>>

<sup>28</sup> Specifically, eduPerson is a schema for the Lightweight Directory Access Protocol (LDAP) – a widely-used directory protocol.

<sup>29</sup> The HR-XML consortium <<http://www.hr-xml.org/>> has defined a schema for the representation of Human Resources (HR) information.

- **eduPersonTargetedID.** This establishes a unique, persistent, pseudonymous identity for each visitor to an SP. The eduPersonTargetedID is not shared between SPs, so the SPs cannot aggregate information about users.
- **eduPersonPrincipalName.** This is a unique identifier for each user, shared across all SPs. In practice, eduPersonPrincipalName is frequently sufficient information to identify a user, as at many institutions it is equivalent to the user's account name and/or email address.
- **eduPersonEntitlement.** This provides a mechanism to transfer additional information regarding a user to a SP. It is intended to identify users who have specific entitlement to a resource, based on whatever conditions are agreed between the IdP and SP.

### **Learner information**

- 3.2.6 Although typical directory schemata such as eduPerson can contain detailed information regarding a user's identity, relationship with the institution, and contact information, they are insufficient for describing *education*, for example courses studied, results, qualifications *etc.*
- 3.2.7 The IMS Global Learning Consortium<sup>30</sup> has developed a range of specifications for use within an educational context. Of relevance here is the Learner Information Profile (LIP), which was modified for UK use (as UKLeaP) and submitted to the British Standards Institute as BS DD8788. However, there has been no significant uptake of IMS LIP or UKLeaP, and the decision has been taken to lapse the draft BS DD8788.
- 3.2.8 One interviewee commented (by email):
- "Currently there is no major focus I'm aware of on profiles of learners. There is exploratory work (e.g. LEAP 2.0) and interest at the level of social profiles [...], and in distributed identity (OpenID) but no concerted effort on learner information. There is basically a split of concerns: learner information overlaps very strongly with:*
- *formal accreditation;*
  - *reporting and statistics (e.g. MIAP CDD);*
  - *enrolment and institutional use of student data (e.g. IMS Enterprise, LDAP);*
  - *portfolios;*
  - *identity management;*
  - *social network profile and internet presence.*
- Each of these areas pulls in a different direction, so a unified generic standard is unlikely."*
- 3.2.9 The IMS ePortfolio specification (IMS eP) provides a schema for the generation of ePortfolios. These may contain much information which would be of use for personalisation; the focus is on collating evidence of achievement, and selection and presentation of that evidence.

---

<sup>30</sup>

<<http://www.imsglobal.org/>>

- 3.2.10 A schema for describing a learner is a prerequisite for developing APOD based on information regarding a user's learning level. Although there seems to be no desire for a comprehensive schema, there are frameworks for describing individual courses (rather than learners):
- **The Scottish Credit and Qualifications Framework**<sup>31</sup> (SCQF) can describe the level of all courses in Scotland.
  - **eXchanging Course Related Information** (XCRI)<sup>32</sup> is an information model (with an associated schema – XCRI-CAP – for describing courses) for exchanging information regarding HE courses within the UK.
- 3.2.11 Either of these frameworks could be used as the basis for describing the current enrolment of a learner, although SCQF does not have an associated XML schema.

### 3.3 The e-Framework

- 3.3.1 The e-Framework defines a series of service expressions, service genres, and SUMs which can provide a starting-point when developing institutional systems using a service-oriented approach.
- 3.3.2 The e-Framework is not normative – definitions are not binding on any party, and developers can mix and match as appropriate. It is unlikely that any institutional-level definitions will entirely meet the requirements of any given institution. At an intra-institutional level, however, the e-Framework is a consistent place to present the requirements for interoperability or compliance.
- 3.3.3 The current content of the e-Framework is limited, and contains nothing directly addressing personalisation. Of potential relevance are those SUMs relevant to authentication and identity (eg the Australian Meta Access Management System (MAMS) federation, and the New Zealand Identity and Access Management) and their associated service genres and expressions, and the service genre *personal competency profile service*. There are two expressions of this service at the time of writing, covering the use of HR-XML and IMS eP - see sub-section 3.2 for a discussion of these schemata.
- 3.3.4 In summary, the e-Framework does not currently contain any definitions specific to personalisation, but will be a sensible place to record the outputs of any work which develops service-oriented personalisation frameworks.

---

<sup>31</sup> <<http://www.scqfcolleges.org.uk/>>

<sup>32</sup> <<http://www.xcri.org/>>

## 4 Current examples of personalisation

### 4.1 Introduction

4.1.1 This section is not intended to be an exhaustive analysis of the personalisation used on a service-by-service or product-by-product basis, but rather as an overview of the key techniques employed across the sector, and the experiences in implementing such systems. Commercial approaches to personalisation are discussed in sub-section 2.4.

### 4.2 JISC content services

4.2.1 The JISC content services have been active in providing opportunities for users to customise their experiences. Typical approaches include:

Approach	Description	Examples <sup>33</sup>
<b>Alerts</b>	A user can save details of a search, and will receive by email any new items added to the dataset which meet their criteria.	Zetoc alerts
<b>Saved searches</b>	A user can save details of a search, and run this again without re-entering the search details. This is of particular use for more complex search criteria, and can be used to establish alerts.	Suncat
<b>Bookmarking</b>	A user can bookmark records, for ease of access later.	MyIntute
<b>Saved outputs</b>	The core functionality of some services involves generating outputs specifically to the requirement of a user. These outputs can be saved for later access.	Mapchest (DIGIMAP)
<b>User interface (UI) customised according to role</b>	The options presented to a user depend on their role, for example teachers could see different options to students.	DIGIMAP, Hairdressing Training Service
<b>Show changes only - "data diffs"</b>	For some datasets, users are expected to download data for use. For frequently-changing data, it is efficient to only download data which have changed since the last download.	DIGIMAP data download
<b>Adaption based on home institution</b>	A service alters its presentation or mechanism of operation based on the institution that a visitor is a member of.	The Depot, GetRef

*Table 4-1: approaches to personalisation within JISC services*

4.2.2 The majority of personalisation capabilities within the JISC content services are customisation, rather than adaptive personalisation. The approaches taken to identifying users can involve use of authentication attributes (from Athens or the UK federation), or inviting users to enrol for the service explicitly.

4.2.3 There is no common infrastructure for this customisation; each service that implements such facilities writes code to achieve it on an ad-hoc basis, and the information thus collected is

<sup>33</sup>

This list is not comprehensive – it is intended to illustrate the approach rather than list every instance. The related DPIE(1) project has conducted a more comprehensive survey of the use of these techniques within some of the JISC data services.

not shared with other services. There are no current moves to establish such an infrastructure.

- 4.2.4 Some services (including DIGIMAP and the Social Science Data Services) require users to authenticate using either Athens or the Federation, and require an additional enrolment step before the user may access the service, in order to comply with license requirements.

### ***Current concerns***

- 4.2.5 Several issues are currently causing concern amongst the service providers, which may impact on their ability to provide personalised services.

### ***Alternative mechanisms of access***

- 4.2.6 Several services are actively developing or considering the development of alternative methods of access to their data. This can include Really Simple Syndication (RSS) feeds, providing Application-Programme Interfaces (APIs) to allow machine-to-machine (M2M) communication, aggregation and "mash-ups", and developing presentations of their sites which are adapted for use on mobile devices (which may be mobile telephones, but could also be task-specific equipment). Two problems arise from this re-packaging of content:

- Providing access to data through RSS or other M2M interfaces is intrinsically not suited to personalisation at the service provider. It is possible that the services which are consuming content are providing a personalised service, but the service provider will not be able to. There are possible mitigations to this (for example, generating RSS feeds on a user-specific basis with opaque locations), but they require careful thought at the planning stage.
- To develop personalised services, it is necessary to allow users to identify themselves and to authenticate on any platform through which they access the service. This can create a problem as the service provider has no control over the site which is presented to a user by their IdP when they wish to authenticate; these sites may not function well on mobile devices, for example.

### ***Migration of user identities***

- 4.2.7 There is currently no ready mechanism by which service providers can associate profiles of users which are associated with their current Athens accounts with those which will be associated with their Federation identities.<sup>34</sup> This is likely to result in users losing any service customisation which they have undertaken. Further, the design of Shibboleth requires any such association to be undertaken at the level of individual service providers. This is an issue of wide concern, and the JISC are working to identify solutions.

### ***IP-based access***

- 4.2.8 Many services allow access based on the IP-address of visitors, and many institutions (especially FE institutions) continue to promote on-site only access to resources (IP-based access precludes – or at best, makes more complex – remote access to resources, and JISC has been keen to migrate users and resources to Athens, and now to the Federation). When authentication is undertaken in this manner, it is not possible to identify individual users

---

<sup>34</sup> Some institutions may be able to associate Athens accounts with individual users, but others do not hold sufficient data to match the identities.

without conducting a separate enrolment step, and therefore it is not possible to undertake personalisation.

### 4.3 Publishers

4.3.1 Although commercial publishers are not the primary focus of this work, they are the conduit through which the majority of academic output is channelled. As such, it is of relevance to investigate their current approaches to personalisation.

#### **Context**

4.3.2 This sub-section is adapted from our previous report.<sup>35</sup>

4.3.3 There are an estimated 20-25k peer-reviewed scholarly journals actively being published worldwide. Many other journals are discontinued but available via archives. The number of journals has grown consistently at a compound annual growth rate of 3-4%.<sup>36</sup>

4.3.4 It is acknowledged that there is only a loose correlation between the number of journals and the number of articles published by any given organisation. The mapping between a journal and its value (*ie* financial sales). Also, some journals have greater prestige and impact factor than others.

4.3.5 The publishing market in the UK is made up of a small number of large publishers, with a large number of smaller publishers (around 90% of publishers with ISI-rated publications publish four or fewer titles).<sup>37</sup>

4.3.6 It is accepted that there is relatively little information available in terms of sizing and segmentation of the publisher market.

#### **Analysis**

4.3.7 The majority of academic publishers are small, publishing a few journals and are probably not suited to developing personalised outputs themselves – they are typically focused on the collation and dissemination of information, rather than the presentation and development of information systems.

4.3.8 The larger publishers are beginning to shift their emphasis from the traditional publishing role toward the provision of online information services, adding value to published material. Although publishers must protect and increase their revenue streams, the approach taken to doing so can vary between traditional “hard sales” approaches, and very community-focused approaches, demonstrating a genuine desire to understand the needs of the customers, and adapt the offerings to better meet them.

4.3.9 Reed Elsevier (“Elsevier”) is the largest publisher in the UK market, and is regarded as being pro-active in engaging with the education and research communities in the UK. As a market leader, it is of interest to investigate the approach that Elsevier is taking to personalisation.

---

<sup>35</sup> *Evaluation of the Core Middleware programmes: final report*, CC253D015-1.0 Matt Shreeve, Geoff Curtis and Alan Grant, 23 March 2007.

<sup>36</sup> *UK scholarly journals: 2006 baseline report*, Electronic Publishing Services Ltd, Final Report, 12 September 2006.

<sup>37</sup> *Reed Elsevier plc and Harcourt General*, A report on the proposed merger, Competition Commission, July 2001.

- 4.3.10 Elsevier produce several aggregator products to facilitate easier access to published work, and these are where efforts to personalise outputs are focused. To take ScienceDirect as an example,<sup>38</sup> many of the techniques used in JISC services are employed (see Table 4-1, p. 25 - saved searches, alerts, bookmarking, and clear visual cues as to whether any given resource is subscribed by the user's institution are provided), but in addition related resources are identified, and listed in a side-bar to the main content. The detailed algorithm used to select these articles is proprietary, but includes full-text matching based on the content of the current content.
- 4.3.11 Elsevier feel strongly that the approach to developing their services should be to understand the workflow of a researcher, and adapt their services to fit this workflow. Whereas this is not exactly personalisation as defined by the JISC (sub-section 1.7), it is intended to make services more attractive to and easier to use by specific groups of users.

## 4.4 Institutional systems

- 4.4.1 All large institutions and many smaller institutions have extensive information systems, and in some contexts these may be the systems with which users have the most contact on a day-to-day basis.
- 4.4.2 This sub-section will briefly discuss three types of institutional system – Virtual Learning Environments (VLEs), portals, and library catalogue systems. There is significant overlap between VLEs and portals, with modern VLE applications essentially presenting a portal to the user, albeit one which is specific to the educational aspects of their interaction with their institution.

### VLE

- 4.4.3 Many institutions use some form of VLE (sometimes referred to as a Course Management System – CMS), many of which claim some personalisation functionality. The VLEs with the largest share of the market<sup>39</sup> are believed to be Blackboard Academic Suite<sup>40</sup> and Moodle.<sup>41</sup>
- 4.4.4 The personalisation offered by these products is outside the definition accepted by the JISC (see sub-section 1.7), and fits more clearly within the application and portalisation regions of the personalisation spectrum (see sub-section 2.2). Examples include the ability of teachers to set up specific schemes of work for individual students or groups of students, and the ability of users to annotate content within the VLE.
- 4.4.5 VLEs can consume resources from elsewhere – for example the learning objects within the JORUM service. We are not aware of any VLE which is designed to feed back information regarding the use of these objects to their original source; this is a desire which has been expressed by several service providers.

---

38 <<http://www.sciencedirect.com/>>

39 The most recent analysis we are aware of is contained within *VLE Surveys - A longitudinal perspective between March 2001, March 2003 and March 2005 for higher education in the United Kingdom*, Martin Jenkins, Tom Browne and Richard Walker, 2005. The position currently is based on anecdotal evidence of discussions with stakeholders.

40 <<http://www.blackboard.com/>>

41 <<http://moodle.org/>>

**Portal**

- 4.4.6 Portals are now widely-used within FE and HE, utilising a range of technologies and architectures. The aim of institutional portals (as distinct from subject portals) is to collect together all the relevant institutional systems and information appropriate to a given user into one location. This information typically includes email, HR/student records, calendars, financial systems, announcements *etc.*
- 4.4.7 As previously discussed, the portalisation of information is not personalisation in the JISC definition *per se*, but nonetheless provides a convenient interface for individual users. The discussions regarding “fat” vs “thin” portals<sup>42</sup> which were summarised in the prior personalisation reports<sup>43</sup> continue, and are made more pressing by the increased uptake of SOAs within institutions and web services across the environment at large.
- 4.4.8 The essential question is to what degree content can be separated from presentation – individual services have developed interfaces to optimise use of that resource, and aggregating these resources into a “one-stop-shop” may not be desirable. Indeed, the service providers interviewed are in the most part strongly negative toward this approach. The degree to which this negativity represents concern over a loss of ownership is unclear, but any moves to aggregate JISC resources should be carefully considered. The views of users of the service are unknown.
- 4.4.9 There is a stream of thought within some of the service providers interviewed 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. Ideas as to how this personalisation should be conducted are not clearly expressed, but are primarily focused toward *eg* generating lists of resources appropriate to particular groups of users.

**Library catalogue and OPAC**

- 4.4.10 Library catalogue systems (and their associated Online Public Access Catalogue – OPAC) are ubiquitous across FE and HE, and most offer a personal application which allows users to keep track of their borrowing, fines, reservations *etc.* Some offer facilities to generate reading-lists for users or groups of users, integrate with institutional finance or VLE systems, and generate detailed information on the usage of resources.
- 4.4.11 Due to the similarity between an OPAC and a bookshop, many of the techniques developed by Amazon can be employed directly within the OPAC. The University of Huddersfield has developed their OPAC to suggest resources based on intrinsic and extrinsic recommendation – both by understanding the activities of previous borrowers of a resource, and by locating resources with similar subjects (Figure 4-1).

---

<sup>42</sup> A thin (aka. shallow) portal presents a user links to the resources which are appropriate to that user, and may authenticate the user to those resources; the resources are separate from the portal. A fat (aka. deep or thick) portal is one which provides access to the appropriate resources within itself; the user need not leave the portal to utilise the resources.

<sup>43</sup> *Personalisation in presentation services – a follow-up report for the JISC*, Neil Smith, Nicky Ferguson and Seb Schmöller, 29 September 2006 pp.17.

University of Huddersfield [Search](#) [My Account/Renewals](#) [Ask a Librarian](#) [Login](#) [My List - 0](#) [Help](#)

[Basic](#) | [Advanced](#) | [Expert](#) | [Short Loan](#) | [History](#)

Search:   [GO](#) [Refine Search](#)

You're searching: **University of Huddersfield Library** [Return to results](#) | [Previous](#) | [Next](#)

**Organometallics / Complexes with transition metal-carbon [sigma]-bonds.**  
by **Bochmann, Manfred.**

Subjects: [Organotransition metal compounds](#) (4 other items)

Publisher: Oxford : Oxford University Press, 1994.

Part no. and title: 1,

Series: [Oxford chemistry primers](#)

ISBN: 0198557507

User rating: [no rating yet](#) with no comments ([add your rating](#) or [your comments](#))

[Add to my list](#) [Request Item](#)

**Print Copy/Holding Information**

Location	Collection	Class (Dewey) No.	Status	Due Date
Library and Computing Centre	Loan Collection Floor 3 - 2 wk loan	547.056 BOC	Available	
Library and Computing Centre	Loan Collection Floor 3 - 2 wk loan	547.056 BOC	On loan	18/02/2008

**People who borrowed this item, also borrowed** ([show more](#) or [everything](#)):

- ◆ [Organometallics / Complexes with transition metal-carbon \[pi\]-bonds 2](#), (1994) by Bochmann, Manfred
- ◆ [Modern inorganic chemistry](#) (1984) by Jolly, William L.
- ◆ [Introduction to colloid and surface chemistry](#) (4th edition, 1992) by Shaw, Duncan J.

**Items with the same or similar subject headings:**

- ◆ [Organotransition metal chemistry](#) (2002) by Hill, Anthony F
- ◆ [Principles and applications of organotransition metal chemistry](#) ([New ed.], 1987) by Collman, James P
- ◆ [Molecular chemistry of the transition elements: an introductory course](#) (1996) by Mathey, Francois



Figure 4-1: Huddersfield OPAC

## 5 Legal landscape

### 5.1 Legal issues

#### *Introduction to the legislation*

- 5.1.1 Personal information is legally protected throughout Europe, and in many other jurisdictions worldwide. The relevant legislation in the UK is The Data Protection Act 1998 (the DPA), which gives effect in the UK law to European Community Directive 95/46/EC. The DPA replaces the Data Protection Act 1984 and was brought into force on 1 March 2000.
- 5.1.2 Annex B presents an overview of the UK legislation and guidance relevant to personalisation (*Nb*: the contents of this report are not, and are not intended to be, legal advice). A more comprehensive primer is available from JISC Legal,<sup>44</sup> and extensive information and advice is available from the website of the Information Commissioner's Office (ICO).<sup>45</sup>
- 5.1.3 The remainder of sub-section 5.1 assumes some familiarity with the terminology and concepts of the DPA.

#### *Relevance to personalisation*

- 5.1.4 Developing personalised services is – by definition – liable to involve personal information. This information may be collected directly from users, either explicitly for customisation or transparently in APUA, or transferred from another source, as in APOD. Any of these operations is liable to fall within the scope of, and must therefore comply with, the DPA.

#### *Transfer of institutional information*

- 5.1.5 One potential mechanism for providing APOD would be to transfer institutional information regarding users to the service provider, and from this to tailor the output of that service accordingly. Whether this disclosure is permissible depends on a number of factors, including:
- **Whether the information transferred is personal data:** clearly, the more information which is transferred, the more likely it is to consist of personal data, given the definitions and interpretations in Annex B.
  - **Whether the user has consented to the transfer:** either at the time of disclosure, or previously (for example during enrolment in their course or module).
  - **The relationship between the institution and the service provider:** if a service provider is contractually bound to the institution, and is processing data for purposes which the subject has consented, this processing may be permitted – but there will be a question as to whether the service provider is a Data Processor or a Data Controller within the meaning of the DPA.
- 5.1.6 Clearly, the most unambiguous approach to legal compliance is to ensure consent from data subjects prior to transferring their information. The point at which this consent is requested, and the efforts taken to ensure that consent is informed must be considered for any such system.

---

<sup>44</sup> <<http://www.jisclegal.ac.uk/dataprotection/dataprotection.htm>> [accessed 6 February 2008].

<sup>45</sup> <<http://www.ico.gov.uk/>>

### *Recording user activity*

- 5.1.7 Records of user activity as required for APUA may well be considered personal data. As such, the DPA will apply. This issue is not specific to personalised services, indeed the content of any web server log files may be personal data, if for example they contain the content of search queries that lead to that site or that were conducted on that site. There is current debate as to whether IP addresses<sup>46</sup> should be classed as personal data;<sup>47</sup> although it may be possible to relate IP address information to an individual, it is difficult to do this reliably – it comes to a matter of interpretation of “reasonable” within the DPA (see B.4.2).
- 5.1.8 In practice, user activity information is recorded by most commercial websites (and many individuals, using tools such as Google Analytics<sup>48</sup>), and server log files are recorded by virtually all websites.
- 5.1.9 In the context of personalisation, it will be necessary to consider how the aggregation of user activity information with information gleaned from customisation or APOD techniques (including information transferred through authentication mechanisms such as Athens or the UK federation) may increase the sensitivity of the overall data.

## **5.2 UK federation policy and guidance**

- 5.2.1 The *Rules of Membership*<sup>49</sup> of the UK federation require that members abide by the eight principles of the DPA. In addition, the Federation’s *Recommendations for use of personal data*<sup>50</sup> present more detailed guidance on the legal situation, and on recommended procedures for the use of personal data.
- 5.2.2 Of particular note are the following:

*"For most applications a combination of the attributes eduPersonScopedAffiliation and eduPersonTargetedID will be sufficient. A requirement to provide other attributes should be regarded as exceptional by both Identity and Service Providers and will involve considerable additional responsibilities for both."<sup>51</sup>*

*"Where a Service Provider wishes to personalise the service they offer to each user this should be done using the eduPersonTargetedID attribute, since this provides the required ability to recognise a returning user and recover their stored preferences or other information.*

*In some cases it may be appropriate for the Service Provider to request additional personal information in order to provide an enhanced*

---

<sup>46</sup> In this context, IP is Internet Protocol. An IP address is a number such as 192.168.1.2 which locates a computer on the internet. IP addresses are often not persistent, and may be used by a large number of computers simultaneously – but there remains the potential to identify individuals, or to add to the information already held about an individual.

<sup>47</sup> <<http://www.washingtonpost.com/wp-dyn/content/article/2008/01/21/AR2008012101340.html>> [accessed 7 February 2008].

<sup>48</sup> <<http://www.google.com/analytics/>>

<sup>49</sup> *Rules of membership*, ST/AAI/UKF/DOC/001, UK Access Management Federation for Education and Research, v1.1 1 November 2007.

<sup>50</sup> *Recommendations for use of personal data*, ST/AAI/UKF/DOC/002, UK Access Management Federation for Education and Research, v1.01 14 Jan 2008. We are working from the final draft of this revised document, replacing the November 2006 v1.0 edition.

<sup>51</sup> *Ibid.*, para. 3.2.2.

*service, for example to send e-mail notices of upgrades to the service or information provided, or to greet the user by name or nickname. As discussed [above], the appropriate way to obtain this information – whether from the Identity Provider or the User – will depend on how critical the information is to the provision of the service. [...]*<sup>52</sup>

---

52

*Ibid.*, para 3.4.1.

This page is intentionally blank

## 6 Challenges and opportunities

### 6.1 Challenges to personalisation

#### *Unclear demand*

- 6.1.1 The demand from users for personalised interfaces is unclear. It is evident that there is great demand for social networking sites, and likewise that those commercial sites which have pioneered the use of personalised sales have been successful.
- 6.1.2 Social networking is outside the scope of this project, but it is important to be clear whether the current "buzz" around personalisation is focused on personalisation in the JISC definition, or rather on the current social networking phenomenon. When engaging with stakeholders during the conduct of this study, initial thoughts almost always turned to user-generated content (including annotation) and social networking (this is being investigated in more detail by the DPIE(1) project).
- 6.1.3 It may be technically and legally possible to develop complex personalised services, but it is essential to carefully consider the use-cases, in order to avoid embarking on a technology-led rather than business-led development programme.

#### *Lack of motivation to significantly develop services*

- 6.1.4 JISC content services have typically evolved from projects which were initially funded through a call or an ITT. This model results in service providers designing services which meet the terms of reference under which they were created and against which they will be assessed. There is currently very limited appetite within the JISC content services to develop personalised services.
- 6.1.5 There is a limited budget available for service enhancement, and the services are more focused on addressing the current, clear demands rather than undertaking extensive research into user behaviour. In a few cases, the expressed viewpoint is that the JISC has decided to fund these services, and as such it is not the responsibility of the service to decide what users want – that decision is taken by JISC. This approach is reinforced by the Service Level Agreements (SLAs) that these services must meet, which address issues such as server uptime and page impressions; these present only a very indirect measure of the value of a service to its users in terms of pedagogical, research or administrative outcome.
- 6.1.6 The nature of personal services demands that the service providers intimately understand user behaviour: the current approach to funding and monitoring services seems unlikely to provide an environment conducive to significant investment in this area.

#### *Understanding the resources*

- 6.1.7 In order to develop services to utilise adaptive personalisation, it is necessary to understand which resources are likely to be of use to which users. This can be undertaken based on extrinsic information (*ie* the behaviour of other users) or intrinsic information (properties of the resource itself).
- 6.1.8 Extrinsic suggestions require sufficient information about the behaviour of other users. Major commercial websites handle tens of millions of users daily, and offer perhaps hundreds of thousands of products, and can generate detailed usage profiles. Services with only a few

thousand users and millions of resources (such as bibliographic datasets) are unlikely to accrue sufficient information density to have real value.

- 6.1.9 Intrinsic suggestions require resources to be categorised and catalogued, either manually or automatically. Many of the JISC content services contain metadata for their resources, but this metadata has been recorded for traditional library requirements; it is likely to be insufficient for adaptive personalisation.
- 6.1.10 For example, a book record may contain keywords such as "poetry, English, 18<sup>th</sup> century", and this may be appropriate for a user who is currently studying a course in 18<sup>th</sup> century English poetry. However, this type of matching is simultaneously too broad, and too precise:
- Too broad: resources which are appropriate for a first-year undergraduate may not be appropriate for a doctoral student. For researchers in particular, any cataloguing short of full-text analysis is unlikely to contain adequate data to locate relevant resources.
  - Too specific: if a student is studying 18<sup>th</sup> century European poetry, they may be interested in resources covering 18<sup>th</sup> century English poetry. Unless a "cross-walking"<sup>53</sup> technique is applied, this will not match.

### **Notions of privacy**

- 6.1.11 The interpretations of "privacy" and related concepts such as identity, personal information and consent can be nebulous, depending on the individual, the situation, the context, the perceived level of individual control and many other factors,<sup>54</sup> and it is outside the scope of this study to discuss in detail the various definitions and models of privacy, except to highlight two points:
- **current discussions of privacy separate the whole into a series of dimensions**, and the definitions of these dimensions are themselves variable, *eg* physical, interactional, psychological and informational,<sup>55</sup> or informational, accessibility and expressive.<sup>56</sup> To talk about "privacy" is not sufficient, issues and concerns must be discussed in their context.
  - **the Internet has specific privacy issues**, many of which are also benefits in other contexts. These have been categorised as: permanence, volume, invisibility, neutrality, accessibility, assembly (*ie* aggregation) and remoteness.<sup>57</sup>
- 6.1.12 Central to the concept of privacy is that the individual must have control over the release of their own information.<sup>58</sup> This concept is extended into UK law through the DPA, which places significant responsibilities on those who process personal information.
- 6.1.13 The move towards user-centric identity (sub-section 2.6) represents an effort to develop technologies in such a way as to comply with the demand, both from users and legislation, for individuals to have control over their own information.

---

<sup>53</sup> Cross-walking is linking metadata, so that related terms can be matched. In the example above, English poetry may not be matched for a search for European poetry, whereas it is actually a subset of European poetry. Cross-walking allows these links to be made.

<sup>54</sup> *Self-disclosure, privacy and the Internet*, Adam N. Joiner and Carina B. Paine, in *The Oxford Handbook of Internet Psychology*, Oxford University Press 2007 pp237-252 presents a good introduction to the issues.

<sup>55</sup> *Maintaining and restoring privacy through communication in different types of relationship*, Judee Burgoon *et al*, Journal of Social and Personal Relationships 1989 **6** 131-158.

<sup>56</sup> *In pursuit of privacy: Law, ethics, and the rise of technology*, J. DeCew, Cornell University Press 1997.

<sup>57</sup> *Privacy: what's different now?*, Karen Spärck-Jones, Interdisciplinary Science Reviews 2003 **28**, 287-292.

<sup>58</sup> Two influential analyses of privacy are those in *Privacy and freedom*, Alan Westin, Atheneum 1967, and *The environment and social behaviour*, Irwin Altman, Brooks/Cole 1975.

- 6.1.14 Given the complexity of perceptions surrounding privacy, it can be challenging to express the nature of individual concerns. Individuals may not have well-articulated views regarding privacy until an event causes them to feel exposed.<sup>59</sup> Any potential personalisation technology which requires the transfer of personal information between organisations must be considered in the context of concern and legislation. Recent well-publicised losses of personal information by both private and public sector organisations have raised concerns about personal information in the public awareness.<sup>60</sup>
- 6.1.15 One aim of the second phase of this study will be to investigate perceptions of privacy online amongst students and staff within UK FE and HE, and how this may impact personalisation. One issue of particular interest is the cost/benefit trade-off involved when users are asked to disclose personal information.

## 6.2 The potential of the UK federation

- 6.2.1 This sub-section sets out an analysis of the potential of the UK federation to support personalised services. In short, the UK federation provides an infrastructure with modest but increasing usage which can provide persistent identifiers for users, and transfer any institutionally-held information to a service provider.

### *Market penetration*

- 6.2.2 The adoption of federated access management is accelerating worldwide, including within the UK. Currently, 198 organisations have joined the UK federation, and this number is steadily increasing,<sup>61</sup> although there are no published data to support analyses of its current level of usage.<sup>62</sup>
- 6.2.3 Few FE institutions have yet adopted the approach, and take-up amongst publishers (especially smaller publishers) is slow. JISC is currently funding a range of projects and services aimed at increasing uptake amongst these communities, and the impending withdrawal of JISC funding for the Athens service may act to promote uptake. All JISC content services will be capable of acting as SPs within the UK federation.

### *Attributes*

- 6.2.4 The four core attributes of the UK federation are sufficient to support customisation, through the unique identification of users by eduPersonTargetedID or eduPersonPrincipalName. The use of eduPersonTargetedID, which is recommended for use due to privacy aspects, prevents the sharing of user profiles directly between service providers. eduPersonScopedAffiliation can facilitate some simple cases of APOD – for example personalising a site based on the user's institution, or whether they are a student or staff.
- 6.2.5 Although the federation specifies the four core attributes, there is no reason not to utilise further attributes from eduPerson or elsewhere, if required.

<sup>59</sup> For example, <[http://www.theregister.co.uk/2006/09/07/facebook\\_update\\_controversy/](http://www.theregister.co.uk/2006/09/07/facebook_update_controversy/)> [accessed 14 October 2007].

<sup>60</sup> See, eg, <[http://news.bbc.co.uk/1/hi/uk\\_politics/7103566.stm](http://news.bbc.co.uk/1/hi/uk_politics/7103566.stm)>, <<http://news.bbc.co.uk/1/hi/uk/7123285.stm>>, <[http://news.bbc.co.uk/1/hi/uk\\_politics/7127951.stm](http://news.bbc.co.uk/1/hi/uk_politics/7127951.stm)>, <<http://news.bbc.co.uk/1/hi/uk/7103911.stm>>, <[http://www.bbc.co.uk/london/content/articles/2008/01/17/queen\\_mary\\_missing.shtml](http://www.bbc.co.uk/london/content/articles/2008/01/17/queen_mary_missing.shtml)>.

<sup>61</sup> <<http://www.ukfederation.org.uk/content/Documents/MemberList>> [accessed 12 February 2008].

<sup>62</sup> There is no information regarding how the members of the federation are using Shibboleth; it is possible that they still use other access-control mechanisms, despite having signed up. In the UK are approximately 470 FE institutions and 170 HE institution, plus a large number of publishers, so it is clear that take-up is currently low.

### ***Institutional data provision***

- 6.2.6 Although Shibboleth could be used to transfer any information about a user, most institutional directory services (which are the source of attributes to the Shibboleth IdP) are understood to hold the minimum information required to authenticate a user, and to provide associated identity information (*ie* eduPerson is a good fit to the data held in many institutional directory services). Richer information which may be of use for adaptive personalisation is not typically stored within the same system; accessing this information and packaging it appropriately for assertion across the Federation is likely to be difficult, and institution-specific, and changes to institutional data structures will become even more complex.

## **6.3 Initial opportunities**

### ***Introduction***

- 6.3.1 One of the aims of this landscape study has been to identify any ideas which show promise for the development of personalised services. Although many stakeholders had good suggestions which were specific to the services they were involved in, these can be broken down into a number of generic themes, which are set out below.

### ***Automatic form-filling***

- 6.3.2 Many services allow users to enrol for email communications, yet find that a large proportion of the email addresses which are entered are incorrect. Allowing a user-consented transfer of email address information (and potentially other information) could be beneficial.
- 6.3.3 In addition, a small number of services require users to enrol for that service specifically, in order to consent to that service's terms of use. The details entered on enrolment are not confirmed with the user's home institution. If the home institution could assert the information required for enrolment, it would help ensure data accuracy and would provide a guarantee of that information's validity.

### ***Improved management information***

- 6.3.4 There is a great demand to better understand the usage of resources, and service providers are enthusiastic about the potential of receiving enhanced information on visitors. Although this would not directly lead to personalisation, it could lead to deeper understanding of the usage of resources, and this information can be fed into the design of services.

### ***Group membership***

- 6.3.5 If an institution could assert membership of groups (for example courses), services could provide simple mechanisms for those groups to interact. One possibility would be to allow a course tutor to recommend resources, and these resources would be recommended to the students of that course. This provides a mechanism for selecting relevant material which is well-tailored to the specific group involved. There would be opportunities for sharing this User-Generated Content (UGC) more widely – perhaps by providing public listings, or as a basis for automatic recommendation.

*Linking to a user's VLE*

- 6.3.6 There is some desire to be able to “push” the content services into users’ VLEs. This could allow personalisation to be undertaken by the institution, and potentially allow the service provider to gather greater information about the use of their resources. The nature of this desired interaction is unclear – even amongst those stakeholders who have suggested it – but it may prove fruitful to investigate further.

**6.4 Summary of landscape study**

- 6.4.1 Current use of personalisation by online services for education and research is limited. Many services offer some ability for customisation, but the type and mechanisms of customisation vary. Within the JISC content services there is no common approach. The situation is similar amongst commercial services, although some large academic publishers are beginning to develop APUA systems.
- 6.4.2 There is demand for APUA amongst JISC content services, but there is currently no real demand for APOD. The views of users are currently unknown; the second phase of this project aims to provide further information in this regard.
- 6.4.3 The UK federation provides an infrastructure which can support customisation and adaptive personalisation, although it would be necessary to develop appropriate schemata for APOD.
- 6.4.4 Several potential opportunities for development have been identified, but there are significant challenges to be addressed before extensively personalised services can be developed. The remainder of this project will consider how to mitigate these challenges, and how to develop the opportunities into proposals for JISC funding.

This page is intentionally blank

## A List of interviewees

A.1 The following individuals were interviewed in the course of this work:

Name	Organisation	Date	Type
Balviar Notay	JISC	7 Nov 2007	Kickoff meeting
Jane Charlton	JISC	7 Nov 2007	Kickoff meeting
Adam Joinson	University of Bath	25 Jan 2008	Telephone interview
Henry Hughes	JANET(UK)	28 Nov 2007	Group interview
Josh Howlett	JANET(UK)	28 Nov 2007	Group interview
Mark Tysom	JANET(UK)	28 Nov 2007	Group interview
Andrew Cormack	JANET(UK)	13 Dec 2007	Telephone interview
Sean Dunn	MIMAS	3 Dec 2007	DPIE(1) meeting
Caroline Williams	MIMAS	3 Dec 2007	DPIE(1) meeting
Mark van Harmelen	DPIE(1)	3 Dec 2007	DPIE(1) meeting
Joy Palmer	MIMAS	10 Jan 2008	Telephone interview
Ale de Vries	Reed Elsevier	8 Jan 2008	Telephone interview
David Chadwick	University of Kent	8 Jan 2008	Telephone interview
Ross MacIntyre	MIMAS	16 Jan 2008	Telephone interview
Peter Burnhill	Edina	17 Jan 2008	Group interview
James Reid	Edina	17 Jan 2008	Group interview
David Medyckyj-Scott	Edina	17 Jan 2008	Group interview
John Casey	Edina	17 Jan 2008	Interview
Helen Chisholm	Edina	18 Jan 2008	Group interview
Andrew Bevan	Edina	18 Jan 2008	Group interview
Fiona Culloch	Edina	18 Jan 2008	Group interview
John Murison	Edina	18 Jan 2008	Group interview
Fred Guy	Edina	18 Jan 2008	Group interview
Theo Andrew	Edina	18 Jan 2008	Group interview
Tim Stickland	Edina	18 Jan 2008	Group interview
Rick Loup	Edina	18 Jan 2008	Group interview
Christine Rees	Edina	18 Jan 2008	Group interview
Alistair Young	UHI	30 Jan 2008	Telephone interview
Paul Hopkins	University of Newcastle	1 Feb 2008	Telephone interview
Mark Elliot	University of Manchester	1 Feb 2008	Telephone interview
Jackie Carter	MIMAS	4 Feb 2008	Telephone interview
Stuart Smith	MIMAS	4 Feb 2008	Telephone interview
Gill Needham	Open University	4 Feb 2008	Telephone interview

Anne Ramsden	University of Leeds	4 Feb 2008	Telephone interview
James Cornford	University of Newcastle	5 Feb 2008	Telephone interview
Marion Tattersall	Sheffield University, JIBS	5 Feb 2008	Telephone interview
Joe Nicholls	University of Cardiff	6 Feb 2008	Telephone interview
John Paschoud	LSE	6 Feb 2008	Telephone interview
Peter Walker	Eduserv	7 Feb 2008	Telephone interview
Martin Weller	Open University	8 Feb 2008	Telephone interview

## **B An introduction to UK data protection legislation**

### **B.1 Introduction**

B.1.1 This annex presents a brief overview of UK data protection legislation and guidance, and draws heavily on the published ICO guidance, highlighting the concepts most relevant to personalisation (*Nb*: the contents of this report are not, and are not intended to be, legal advice).

### **B.2 The data protection principles**

B.2.1 At the core of the Data Protection Act 1998 (DPA) are the eight *data protection principles*:

- 1) "Personal data shall be processed fairly and lawfully and, in particular, shall not be processed unless [one of a number of specific conditions are met – see para. B.3.1]".
- 2) "Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes".
- 3) "Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed".
- 4) "Personal data shall be accurate and, where necessary, kept up to date".
- 5) "Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes".
- 6) "Personal data shall be processed in accordance with the rights of data subjects under this Act".
- 7) "Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data".
- 8) "Personal data shall not be transferred to a country or territory outside the European Economic Area, unless that country or territory ensures an adequate level of protection of the rights and freedoms of data subjects in relation to the processing of personal data".

### **B.3 Conditions for processing**

B.3.1 The conditions for processing under the first principle (for the purposes of the DPA, disclosure of information is considered to be a form of processing, albeit that disclosure has specific conditions in some situations) are:

- *"The data subject has given his consent to the processing.*
- *The processing is necessary –*
  - *for the performance of a contract to which the data subject is a party; or*

- *for the taking of steps at the request of the data subject with a view to entering into a contract.*
- *The processing is necessary to comply with any legal obligation to which the data controller is subject, other than an obligation imposed by contract.*
- *The processing is necessary in order to protect the vital interests of the data subject.*
- *The processing is necessary [for the administration of justice, to comply with law, or for the functions of government]*
- *The processing is necessary for the purposes of legitimate interests pursued by the data controller or by the third party or parties to whom the data are disclosed, except where the processing is unwarranted in any particular case because of prejudice to the rights and freedoms or legitimate interests of the data subject."*

B.3.2 The necessity of processing (as required for several of these conditions) is viewed by the ICO as:<sup>63</sup>

- *"the purposes for which the data are being processed are valid,*
- *such purposes can only be achieved by the processing of personal data and,*
- *the processing is proportionate to the aim pursued."*

## **B.4 What is personal data?**

B.4.1 Personal data is defined within the DPA as:

*"data which relate to a living individual who can be identified:-*

- *from those data; or*
- *from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller*

*and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual"<sup>64</sup>*

B.4.2 The Information Commissioner takes a broad view of how this should be interpreted:

*"Sometimes it is not immediately obvious whether an individual can be identified or not, for example, when someone holds information where the names and other identifiers have been removed. In these cases, Recital 26 of the Directive [95/46/EC] states that, whether or not the individual is nevertheless identifiable will depend on "all the means likely reasonably to be used either by the controller or by any other person to identify the said person".*

---

<sup>63</sup> *Data Protection Act 1998 Legal Guidance*, Information Commissioner's Office, §3.1.6 available online from <[http://www.ico.gov.uk/upload/documents/library/data\\_protection/detailed\\_specialist\\_guides/data\\_protection\\_act\\_legal\\_guidance.pdf](http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/data_protection_act_legal_guidance.pdf)> [accessed 7 February 2008].

<sup>64</sup> Data protection act §1(1)

*Therefore, the fact that there is a very slight hypothetical possibility that someone might be able to reconstruct the data in such a way that the data subject is identified is not sufficient to make the individual identifiable for the purposes of the Directive. The person processing the data must consider all the factors at stake.*

[...]

*When considering identifiability it should be assumed that you are not looking just at the means reasonably likely to be used by the ordinary man in the street, but also the means that are likely to be used by a determined person with a particular reason to want to identify individuals. Examples would include investigative journalists, estranged partners, stalkers, or industrial spies.<sup>65</sup>*

B.4.3 And specifically regarding internet resources:

*"If the information about a particular web user is built up over a period of time, perhaps through the use of tracking technology, with the intention that it may later be linked to a name and address, that information is personal data. Information may be compiled about a particular web user, but there might not be any intention of linking it to a name and address or e-mail address. There might merely be an intention to target that particular user with advertising, or to offer discounts when they re-visit a particular web site, on the basis of the profile built up, without any ability to locate that user in the physical world. The Commissioner takes the view that such information is, nevertheless, personal data. In the context of the on-line world the information that identifies an individual is that which uniquely locates him in that world, by distinguishing him from others."<sup>66</sup>*

B.4.4 Within this definition and interpretation, even information which does not explicitly identify an individual could be personal data.

## B.5 Sensitive personal data

B.5.1 The act defines *sensitive personal data* as information regarding:

- “the racial or ethnic origin of the data subject;
- his political opinions;
- his religious beliefs or other beliefs of a similar nature;
- whether he is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992);
- his physical or mental health or condition;
- his sexual life;
- the commission or alleged commission by him of any offence; or
- any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings.”

---

<sup>65</sup> Data Protection Technical Guidance - Determining what is personal data, Information Commissioner's Office, available online from <[http://www.ico.gov.uk/upload/documents/library/data\\_protection/detailed\\_specialist\\_guides/personal\\_data\\_flowchart\\_v1\\_with\\_preface001.pdf](http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/personal_data_flowchart_v1_with_preface001.pdf)> [accessed 12 February 2008].

<sup>66</sup> *Data Protection Act 1998 Legal Guidance, op. cit.*

B.5.2 For these classes of information, a more rigorous set of conditions must be met before processing can take place.