

**Summer/Autumn
Issue 2002**

Circle of State Librarians

Edited by Mags Griffin

HM Customs and Excise

**Circle of
State
Librarians**

CSL London 2002

Acknowledgements

This issue contains articles with a fairly wide subject coverage. It includes an article on the universality of classification in daily life and KM networking across businesses (public/commercial). The success of departmental web site development is investigated, as is the recognised need for and importance of supply of credible information resources to government.

The Committee is grateful to all authors for their agreement to the publication of their contributions here.

Call for Papers

This magazine aims to cover all subject interests relating to policy and operation information management in government for information specialists. This includes news about changes within your organisation, new developments and new ideas about information/IT/KM policies and operational practises. If you would like to contribute a article please contact the editor at **mags.griffin@hmce.gsi.gov.uk**, or at KM Assignments, IMSD Logistics, 8th floor, HM Customs & Excise, Dorset House, Stamford Street, London SE1 9PY. Mobile: 07776 166681, Fax: 020 8929 6677.

The copyright of the individual articles remains with the authors. The issue and its typographic representation is the copyright of the Circle of State Librarians.

Views expressed are those of the authors.

Contents

CONTRIBUTORS	4
THE AFGHAN WAR AND BEYOND – THROUGH THE EYES OF BBC MONITORING	0
<i>Bet Tickner, BBC Monitoring</i>	
THE ASSOCIATION FOR INFORMATION MANAGEMENT – INFORMATION RESOURCE MANAGEMENT NETWORK	0
<i>Rosemary Miller, HM Customs & Excise</i>	
THE ART OF CLASSIFICATION – MAKING USE OF AN OLD NEW BUSINESS TOOL	0
<i>Mark Field, CLIP</i>	
HOW FRIENDLY ARE GOVERNMENT WEBSITES	0
<i>John Scott Cree</i>	

Contributors

Rosemary Miller

Rosemary is currently working as part of the Knowledge Management Assignments team in HM Customs & Excise. The team aims to integrate knowledge management into normal ways of working within HM Customs & Excise to enable it to achieve its business aims more effectively. She was previously Data Registrar, responsible for implementing data ownership.

Mark Field

Mark has managed library and information services in the construction industry, the media and consultancy. He is an unreformable technophile whose hobbies include crashing systems at work and at home in as many arcane ways as possible without any formal training in programming. His specialisms have developed in establishing or transforming information services. He very briefly achieved guru status when he set up complete Internet services at the Centre for Exploitation of science and Technology using Apple Macintosh shareware. CEST's Internet service provider reduced their annual fee by £1000 when he left.

Bet Tickner

Bet Tickner is a duty editor at BBC Monitoring, currently responsible for press, public relations and marketing. She has worked as a French and Russian linguist and as an editor on the Asia Pacific and East African desks.

John Scott Cree

John is currently working as a librarian for the Department of Health.

The Afghan war and beyond – Through the eyes' of BBC Monitoring

Bet Tickner

BBC Monitoring

The war in Afghanistan provides an interesting case study of the importance of information flows to government in times of conflict. Specifically this article will demonstrate how BBC Monitoring deals with the challenge of providing an effective information channel to a modern government.

BBC Monitoring gathers news, information and comment from the mass media around the world and provides electronic news services to government customers and the BBC itself.

The main consumers within HMG are those concerned with foreign policy, defence and international security. To these BBC Monitoring supplies the full range of its reporting, covering a host of topics - political, economic, environmental, humanitarian and defence-related issues, the media and more. Our reporting is also used by the main BBC newsrooms, and reaches the public through this channel and is available online to members of both Houses of Parliament .

The task in this age of globalisation is to provide relevant and timely reports from all the world's mass media - TV and radio, the press, news agencies and increasingly the internet - to meet the wide-ranging requirements of analysts and policy-makers in government, and to make that service fast and easily accessible.

Clearly flexible online delivery systems are the key to enabling government to be effective in the international sphere. They avoid information overload

and enable news providers like ourselves to get the right information to the right people when they need it and in an accessible form. They are also in line with HMG's policy of providing electronic access to the widest range of information for all government departments.

There is a wealth of information available from open sources in most countries which are potentially of interest to people working in government. However most analysts and policy-makers have little time to surf the Internet, for example, to find it. BBC Monitoring uses up to 3,000 sources across the globe, selecting rigorously what is relevant and translating into English from roughly 100 languages.

Despite this proliferation of electronic and conventional media over most of the world, it cannot always be assumed that the right information can be easily accessed however. Afghanistan is a case in point - a poor country with a derelict communications infrastructure which hampers newsgathering as much as official censorship and the dangers of a conflict situation. There are many other such countries, particularly in sub-Saharan Africa. BBC Monitoring does its best to mitigate these shortcomings by using a wide range of sources and by its constant alertness to new sources.

Afghanistan and the War on Terror

As in all contemporary wars, the international media are vital players in this war, not just reporting events, and reactions from all sides to events, but actively attempting to win over hearts and minds in the theatre of war and more widely among world public opinion.

Before the US bombardment began on 7th October, all mass media were severely restricted in Taleban-controlled areas of Afghanistan. TV was banned and Radio Voice of Shari'ah, the only national, Taleban-controlled radio station broadcasting from Kabul, was on air only for about eight hours daily. When the fighting began, access to the country for international journalists was heavily restricted, and journalists either attempted to enter Afghanistan clandestinely or moved with the advance of the Northern Alliance forces. Their ability to report fully on unfolding events was thus restricted.

In such situations BBC Monitoring's extensive reach puts it in a strong position to fill the information void.

Radio Voice of Shari'ah fell silent on day one of the bombing, to resume only briefly on day two. It then ceased broadcasting altogether when its transmitter was blown up. BBC Monitoring recorded these events and from then onwards turned to other sources to give a full account of what was happening in the country now at war. Balkh radio continued to broadcast from the northern capital of Mazar-e Sharif and the Afghan news agencies across the border in Pakistan as well as the Internet.

On 14th October a new broadcaster was heard on the airwaves - this was the start of the US 'psy ops' broadcasts, telling the Afghan people to stay indoors and keep away from Taleban-run establishments. They also reassured listeners that foreign troops were landing in the country to help the Afghan people. "Information radio", as it calls itself, is believed to be still broadcasting, though no longer from on board an aircraft.

Regional states became important players in this war overnight. The actions and reactions of the Central Asian states of Tajikistan and Uzbekistan, were reported by our unit in Tashkent and attitudes in Iran and Pakistan were carefully monitored.

In October our output increased substantially from an average 800 to 1,000 stories a day, while at the same time interest in BBC Monitoring's reporting rose significantly. To make this surge of information manageable for users, new services were organised, in particular a daily round-up of reports from Afghanistan, its regional neighbours and the wider Islamic world, from Somalia to Indonesia. This was a well-received attempt to keep government customers up-to-date with the main happenings of the day and the main trends in public opinion across the Islamic world.

The future

It is a truism to state that the world has become and will in the future be less predictable, more volatile. Forecasting future trends and outcomes to enable policy decisions to be made has now become a more risky exercise. Equally, a wide range of information, giving an authentic picture of events has become more crucial to help facilitate this decision process.

To be in a position to provide this, BBC Monitoring must maintain its expertise in media research, tracking new sources as they emerge, checking on such issues as ownership, editorial policy, etc. and adding new sources to coverage where appropriate.

Appropriate technical solutions for information flow and for knowledge management are also likely to be key to success for those in the business of newsgathering and information services.

To meet diverse customer requirements and to handle different types of content, in addition to the core text service, BBC Monitoring is now developing a portal which will enable users to link together the different types of material to suit their needs - current reports with archive or reference material, graphics, audio and video. These linked services will add value and a new dimension to traditional textual material,

As HMG applies the advantages of information technology and knowledge management to its business, BBC Monitoring hopes to be in a position to serve more users within government. In particular government-wide networks and internet-based systems will provide a vehicle to make this material more readily available.

In the future it is likely that many government departments with a strong domestic focus will nonetheless require more information on trends in Europe and the industrialised world in particular, as well as on global issues. The Department of Trade is an obvious example of a department with a global perspective. Similarly the Home Office and police need to be up to date on trends in migration with an eye to immigration and asylum issues in this country, and on other issues, such as organised crime and the narcotics trade.

To sum up, the future task for government information providers will be to target effectively in these terms:

- to provide the most authoritative, focused coverage of events, rather than blanket coverage.
- to deliver the right information to the customer who needs it, based on an understanding of their needs, which will be diverse, and using a variety of online delivery systems according to preference.

To find out more about BBC Monitoring, visit <http://www.monitor.bbc.co.uk>

The ASLIB-IRM Network

Rosemary Miller

Background

The Association for Information Management (ASLIB) – Information Resource Management (IRM) network is a special interest group focused specifically on the management of information resources. It was created in 1992 by a small group of like minded individuals who wanted to explore the issues around managing information as a resource.

In order to give the network a focus from the start, one of the group, Nick Willard devised a model (now known as the Willard Model) which identifies five key elements of IRM:

Identification -The discovery of information resources and the recording of their features in an inventory

Ownership - The establishment of responsibility for the upkeep of an information resource

Cost and Value - Assessment of the cost of an information resource and its value to the organisation

Development -The further development of an existing information resource to enhance its value to the organisation

Exploitation - The processes which may allow a resource to generate further value through conversion into an asset or a saleable commodity

In 1993, the Network decided to devote a year to exploring each of these five elements. So, from 1993 to 1998, the Willard Model served as the basis for invited speakers to talk to four or five open meetings each year.

After exploring the fifth stage in 1998, it was again Nick Willard who provided the Network with its agenda into the new millennium, when he proposed an exploration of knowledge management as a natural extension of IRM. That agenda is now well under way and is proving to be highly successful.

Network Objectives

The objectives of the network are:

- To initiate and encourage the development of Information Resources Management (IRM) concepts and practices within organisations.
- To promote the awareness and understanding of IRM among senior managers and information professionals by providing a forum for discussion and exchange of experience.
- To support the adoption of IRM policies and procedures as sound business practice.

Network Meetings

The network has over a hundred members who come from a variety of different backgrounds. There are information professionals, information consultants, IT professionals, from both the public and private sector with a mix of knowledge and experience.

The network meets every couple of months at various venues but mainly within the London area, with the occasional visit to more unusual locations. Such as in 1997 when the network held a meeting at Bletchley Park, the topic of the meeting being Information Exploitation in WWII.

At each meeting, there is a speaker, followed by a discussion on the speaker's topic which often broadens out into other related issues. Sometimes the follow up to the speaker involves taking part in practical exercises.

A list of recent speakers gives a flavour of the wide range of speaker topics

Elspeth Scott	GlaxoSmithKline plc	Information Policy: Winning 100,000 Hearts and Minds
Victoria Ward	Sparknow	Unearthing and transferring knowledge: The role of objects, art and science
Olivia Freeman, Liz Orna, Kevin Miles, Stuart Ward	Aslib IRM Network Members	Business Value from Information and Knowledge: Key processes and Enablers
Alan Gilchrist	Cura Consortium & Associate Consultant TFPL Ltd	Corporate Taxonomies – Their Building and Use
Professor Charles Oppenheim, Joan Stenson Professor Richard Wilson	Loughborough University	The Attributes of Information as an Asset
Myra Johnson	William M Mercer	Knowledge Management to Knowledge Mobilisation
Liz MacLachlan Trician Lynch	DTI and UNITAS	DTI Information Architecture
Prof. Richard Potter	Defence Evaluation and Research Agency	Knowledge Management – The DERA Experience
Anne Deering	A T Kearney	Using Knowledge Management for competitive advantage
David Snowden	IBM	Story telling: an old skill in new context
Dr Victor Newman	Cranfield University	Towards a Methodology for Benchmarking Intranets and Measuring Business Value

The network has a relaxed and friendly atmosphere which encourages open and frank discussion and provides the opportunity to share knowledge and experience. This includes sharing problems – there always seems to

be someone who has come across the problem you've got and can offer a solution. Even if that means advising you what not to do!

The end of the afternoon is then rounded off with a glass of wine and the chance to chat some more.

Meetings are open to all, but a charge is made for non members of the network.

The network has its own website at www.irm.org.uk which has details of lots of IRM and KM resources, plus a discussion forum. The site includes a notice board ('On the Wall'), which has is a useful way to keep up to date with developments in the IRM world.

My Experience

I joined the network in 1997, through an introduction from an existing member. I had just moved into Information Management and saw the network as useful way to learn, particularly from the experience of others. At the time I was responsible for introducing Data Ownership into H M Customs & Excise and I found the Willard model invaluable in helping me structure the work.

Starting at the identification stage of the model, we set about producing an inventory of data. This included recording the IT system where the data was held and what the data meant in business terms. For example if the data held was a code '0001', that this meant that a HM Customs office was a local VAT office.

For the ownership stage, we identified departmental owners for the data and assigned responsibilities to those owners regarding the accuracy and relevance of the data and who could have access to it.

I have to confess that we ducked out of the difficult stage of looking at the cost and value of the data, although we could have looked at things like the cost of collection or the cost to the department if key data had been lost.

I am no longer part of the project, it has now moved on into the development stage, where the inventory is being used to inform data analysis of future risk trends and to assist in shaping the future data architecture of the department.

Over the past 5 years, I have made many useful contacts through the network, learnt a lot and had fun.

Rosemary Miller , 8th floor, HM Customs & Excise, Dorset House, Stamford Street, London SE1 9PY.

Tel: 020 8929 0759.

Email: rosemary.miller@hmce.gsi.gov.uk

If you want to know more about the ASLIB-IRM network, go to the website or contact the Membership Department at ASLIB.

The art of classification

Making use of an old
new business tool

Mark Field

*Professional adviser, Special Libraries and Information Services,
The Library Association*

We are all classificationists even though we may not recognise it. Classification is a critical component of the set of processes that provide our interface with the world. To some extent this is self-evident: we group like things together. People, the means and ways in which we do things, plants and artifacts are all associated in groups because they look the same, they share some common meaning or purpose, or they just ‘feel’ right together.

There can scarcely have been a time when this has not been the case. While I have not considered how stromatolites – currently the first discrete manifestation of life that can easily be detected by humans – may have recognized groups, if nothing else the communities of bacteria they represent had one very simple classification: good stuff and bad stuff; food, temperature and light. Classification is generally seen as the preserve of two communities: librarians and taxonomists, the tidy-uppers of books and the life sciences; people who obscure the obvious nature of things with arcane notation and peculiarly artificial names. This is clearly not my position.

There are other classificationists. Patent agents are energetic classificationists – without a means of associating and comparing what are, in effect, ideas, the whole machinery of patents would become chaotic.

Classification answers one very simple question: 'What is this?' I would propose that this question is part of any critical process and that it is often mauled by unchecked assumptions, fogged by semantics (particularly in English), or just not asked – and yet whole structures of actions and procedures are built on shaky and incomplete understandings of what might be a resource, a difficulty to be overcome, the kernel of an opportunity, or a compound of all of these. Worse still, this thing we wrestle with is almost certainly related to many other episodes in which methods or devices for dealing with the problems at hand, have been developed or discovered.

“ So we come back to the idea that a good indexing language or classification scheme should provide a useful tool for the questioner to define what his question really is. So you have to provide by means of an indexing language, not only a retrieval tool but a conceptual tool to define what a person really wants to know.”

Alan Rees in Faceted Classification Schemes, B.C. Vickery

“ Why are you doing this? Before you sit down and talk to a consulting firm, it would help you to have some idea of what it is you want to achieve. The more clearly the goal is defined, the greater the chance of reaching it. If you don't know what you want to do, don't make the call.”

Dangerous Company, O'Shea and Madigan

CLASSIFICATION AND BUSINESS PROCESSES

It is possible to accept that there are techniques and methods in the sciences of classification – of 'systematics' – that can be applied to large parts of business process. However, there are qualifications to this assertion.

First, classification has been and is applied to business processes. Where it works, like any other investment, it delivers more benefit or value than it requires as cost. Estate agents are classificationists. They organize their properties by price, by location and dwelling type. That's classification. It enables estate agents to present clients with a list of likely properties within minutes of their walking through the door. This is not a profoundly intellectual process. My estate agent files by price first, then area, then

dwelling type. If I were extremely wealthy my estate agent would file by location, dwelling type and then price. Where classification fails, it is ignored and withers.

Secondly, the history of the science of classification is obscure and is not easily related to modern business processes. Accountants may claim that theirs is the second oldest profession – since the Acadian Cuneiform pressed into clay tablets in Bronze Age Mesopotamia recorded inventories and transactions – but they have also been able to demonstrate a central role in successful twentieth-century businesses. The Upanishads of 4th Century BC India classified knowledge, but for most of its history, classification was a secondary activity of philosophy. It does not appear to be worldly. The history of classification is an interesting tale, taking in Greek philosophy and the classification system of the Chinese Imperial Civil service, demonstrating how it has been able to contribute to successful and long living institutions. However, it is in recent developments that the potential of classification methodologies may be most apparent.

APPLICATIONS OF CLASSIFICATION

We can demonstrate that classification does a number of things: it analyses a given thing – articulating all of the meaningful concepts it contains – it discovers relationships – associating clusters of things that share common elements. It can be used to map things on to a pre-existing structure, or it can discover the innate structure in groups of things. It can be used to map things on to a pre-existing structure, or it can discover the innate structure in groups of things.

There is not space here to present a short course in classification, let alone all sciences of systematics, and frankly some of them look pretty weird. Instead, we can look at a few developments that have obvious applications.

Modern faceted classification was developed in 1927 by S.R. Ranganathan, which is more than most will need to know about its genesis. He introduced a number of devices, some of which have been widely adopted.

The basis of the scheme for our purposes is the idea of the 'facet'. Much misapplied since, this is essentially a list of concepts sharing some intelligible relationships. 'Facet analysis' is the process in which the simplest (most discrete) concepts in the thing under consideration are identified and assigned to facets. Some concepts may be assigned to a single facet, some

to several different facets, some to the same facet. At this point, librarians would assign a value to each concept before compiling them into a string of values organized by some rules for ordering the facets: the 'notation', the code-like class numbers you see in any public library. The rules of notation have the practical effect of providing an order for putting books on shelves in one sequence, a good idea in libraries, but not obviously useful for extending the analytical facilities inherent in this process.

To state simply what is happening in the 'act' of classification: we are saying that this document – or equally this issue, this object, this complex thing – comprises these things, and these actions. This is a process of analysis and description that requires that we use the simplest level of naming, not the most specific. In effect, we are attaching to the big thing a series of labels with small, incontrovertible statements about what it is: these are combined to give an accurate and comprehensive statement. If it helps, we might imagine an elaborate piece of machinery to which we are tying brown cardboard labels, each bearing the word describing the part to which it is attached. This is very like the way in which the classificationists of the life sciences attribute plants or animals to their species: they identify a series of significant characters that are combined to pass a (frankly arbitrary) threshold of similarity with other members of the same species.

Let's pause for a moment: what would it be worth to know that the problem facing your company was a 'training' problem, not a 'workforce' problem, and to characterize the nature of training?

E-commerce: have you detected the publishing, tax and logistics expertise in preparation for selling over the Internet?

In general terms, we might agree that there is a methodology or set of methodologies in the science of classification that can be developed as means of assessing problems and suggesting courses of action. I would go further, but we can hold here for the time being. How will this development proceed?

There have been some very timely developments in the study of how we acquire and process and act on information, much of it centered on the Santa Fe Institute. In particular, the work of Marvin Minsky and John Holland suggests a model of the operation of mind in which small detect-and-react agents can be combined in successively more complex

frameworks to achieve higher and higher levels of intelligence in responses to conditions, culminating in anticipation and planning.

THE CONCEPT OF 'AGENT'

Something that is activated in the presence of external or internal conditions may be a good approximation of how the physical mind works, or just a useful model. Either way, it presents us with a tool that can be developed in several ways. It is very simply described by John Holland's frog. Put crudely, the frog must react to external conditions, and one in particular 'movement', something that is moving. But react in what way? That is determined by combining 'movement' with another condition – 'size', big or small. 'Big' and 'moving' is likely to be a threat, 'small' and 'moving' is likely to be food. Holland goes on to develop this simple model as an introduction to complexity theory, but there are parallel applications.

I would propose that there are useful comparisons to be drawn between 'agents' and 'facets'. They both 'match' simple abstractions and then increase their value as they are assembled into frameworks of abstraction.

These 'frameworks' are the structures we derive from, and impose on, our experience of the day-to-day world; these 'abstractions' enable us to associate sets of artefacts and actions – things – into patterns with associated frameworks of response. Planning might be achieved by creating a 'desired state' framework and constructing a framework of artefact and action components and partial frameworks that show some correspondence. It is probable, in this model, that we are looking at abstractions of abstractions: we would not retrieve actual procedures or responses from our stores of experience until the moment of application. This is a model, we do not know now how closely it resembles the actual process of mind. However, it is simple and highly scalable and, when applied to systematics – classification - we can describe new conditions, synthesize an understanding and locate courses of action. In other words, we can know what the question is.

While we do not fully understand our capacities for detection, recognition and ordering, they are more consistent across human societies than we might imagine. The early seventies saw the beginnings of the study of 'folk systematics', the classification systems used by pre-industrial peoples to order the natural world. The correlation between these and

the scientifically derived systems used in the life sciences of the 'developed' world is remarkable; but maybe it shouldn't be. The 'folk systematics' of the high street is equally consistent and equally organic in origin. The filing systems of estate agents are the same in every town in the UK, and as 'realtors' that consistency probably applies in the US as well.

Is any of this practical? Yes. Classification, is a tool to simplify understanding. As long as it is used in that context it can be almost anywhere, and to any degree. Some examples: what began as a long piece of work organizing a surveyor's library became a much shorter project working with the senior partner on the design of a classification system for all the practice's documents, which mapped the business process and could be learned quickly by six-month interns. More recently, parts of the standard practice of indexing were used to describe the 'language of success' in a large retail development, which was then combined with a facet analysis of project episodes that had a good outcome to identify what devices and processes could be transferred to other projects. I have many others, as do other librarians who have taken the widest possible (even heretical) view of their professional tools.

Why is this wonderful tool not an explicit part of business practice everywhere? Maybe the need to deal with ever-growing amounts of information has only recently reached critical mass. That would fit with an emergent Internet-driven business paradigm. Librarians have been a quietish profession in the past. Not anymore.

Reprinted by kind permission of the Confederation of British Industry

Mark Field is professional advisor, Special Libraries and information services, the Library Association. He can be contacted on mark.field@la-hq.org.uk or mark.field@cableinet.co.uk.

How friendly are government websites?

John Scott Cree

The websites of government departments have gone through a number of changes. On the whole, they look more modern and some appear to be more expensively produced than others. But how easy is it to retrieve these sites? In March 2002, I had a quick look at the following sites; it will be interesting to see how much things have changed by the time this article appears in print:

Department of Health (DH)

Department for Education and Skills (DfES)

Ministry of Defence (MOD)

Department for Environment, Food and Rural Affairs (DEFRA)

Department for Transport, Local Government and the Regions

Foreign and Commonwealth Office (FCO)

Department of Trade and Industry (DTI)

Home Office (HO)

Health and Safety Executive (HSE)

Department for Work and Pensions (DWP)

Treasury (T)

Cabinet Office (CO)

Office of the e-Envoy

Govtalk

UKonline

User behaviour

Recent reports in the excellent <http://searchenginewatch.com> indicate that users tend increasingly to by-pass search engines and to enter the url directly when they know it.

The .gov.uk names of most of the above sites are fairly memorable and will act as default even when .gsi is present in a url but is not keyed. MOD, arguably, have the simpler .mod.uk (I thought only BL had managed to get away with that), which is more easily remembered but out of synch with other government departments.

What search engines look for

But what if the user accesses the site via a search engine? It's a given that most users will not trawl through pages of retrieved references to find a site they want. Most expect it to be near the top of the list and, if it's not, they don't bother. Despite some good features of the Home Office site, it does not appear on the first page of Google retrievals for a search " Home Office homepage" . So how do sites get near the top of the list? It depends on how search engine friendly sites are (unless ranking is being paid for).

Browser compatibility

Some document coding is unacceptable to certain browsers. It is important to test sites against known browsers, to see how they look. www.websitegarage.com can produce such diagnostic checks. This is a Netscape product which does not appear always to match against IE5. The check creates a number of warnings that html coding on a site is not compatible with certain browsers. Most warnings are for incompatibility with WebTV. Warnings appear in the form < attribute at the top margin of the tag body, this is not supported by Netscape Navigator 4.0, Netscape Navigator 3.0>. Lowest number of Warnings was for CO (11) and highest were for DTI and T (32)

Websitegarage ratings for browser compatibility of government sites visited:

Excellent	-	CO
Good	-	DfES, e-Envoy
Fair	-	all others except:
Poor	-	DTI

Load time

We've all experienced slow loading results. Sometimes we switch off in impatience (how long do you let a telephone ring?). We've blamed our PCs, our Internet link, our network, but the fact is that website design is often unhelpful. Designs should be tested on lower spec PCs for example. Some homepages are simply too big. Others have very large graphics and, without height, width and ALT information in (even in the "New" or "Updated" images), they take even longer to load. The coding not only lets the reader know what's coming when it takes a while to download, but search engines also pick up on it.

Some homepages have become cluttered with links because of the (usually house style) approach that everything must be accessible within three clicks. This does not seem to be a good approach. Users are likely to become frustrated by inability to see the wood for the trees on a cluttered page. They are prepared to drill down to find what they want, provided load time is not too long. In website design, there is an increasing appreciation of "The rule of 9" - there should be no more than 9 links on any one page.

Websitegarage's diagnostic checks how fast the page loads up under 6 common modem speeds and points out that slow load times may lose visitors. Surveys indicate that the majority of visitors would not wait longer than 20 seconds for a page to download. Yahoo! loads up in only 6 seconds on a 28.8K modem. Load time is dependent on the number of images on the page, the size of each image, and the speed of the modem. Below is an example from a site (which I shan't name), which merited the ranking "Fair".

Connect Rate	Connect Time
14.4K	52.37 seconds
28.8K	29.33 seconds
33.6K	25.31 seconds
56K	20.06 seconds
ISDN 128K	7.72 seconds
T1 1.44Mbps	2.58 seconds

As you can see, these are rather slow readings for a not uncommon 56K modem.

Websitegarage advises keeping the size of the homepage under 40Kb. There was considerable variation in the size of the government homepages visited. Below are the respective sizes of homepages and their attachments

Department	Homepage size	Total size with Attachments
DH	30960	49252
MOD	30677	61711
DTLR	14937	29171
FCO	18232	61612
DTI	29398	74435
HO	27547	83817
HSE	18758	79643
DEFRA	11257	17537
DFES	26586	96005
DWP	14483	39045
T	42863	109872
CO	36802	92822
Office of the e-Envoy	31473	39536
govtalk	20349	45415
ukonline	57234	85335

Based on the above, the Websitegarage ratings are no surprise:

- Excellent** - DEFRA
- Good** - DTLR, DWP, Office of the e-Envoy, govtalk
- Fair** - DH, DTI, MOD and FCO
- Poor** - HO, HSE, DfES, T, CO, ukonline

Dead links

All sites scored top marks with websitegarage for having no dead links on their homepages.

Website submission

Search engines no longer rely on continuous trawling to identify new and appropriate sites. There are web-based companies and many individual search engines, which invite paid submissions of sites, to guarantee high placings in the retrieved listings. There are other methods of achieving this. **www.freepint.com** includes a click on link to submit a site free of charge to 8 or so major search engines. This includes

www.scrubtheweb.com which first alerted me to free diagnostic tools. **www.jimtools.com** has a mass of helpful information for web site constructors and includes a link to submit to almost 200 engines. However, it carries the warning that submitting in this way will generate large amounts of email which should diminish after a few days. Jimtools dismisses this phenomenon as a fact of life today.

By sticking with the modest 8-engine submission facility of Free Pint, the spam email problem will be less. Submitters receive a message showing which submissions are successful and unsuccessful. Some are unsuccessful only because the submission tool has exceeded its quota with a given search engine. If you look at the homepage of engines which failed to accept the submission, you will find a "submit" or "suggest new site" link, but some require payment to guarantee early processing.

If you don't pay, wait for up to a month before searching each engine to check that it retrieves your site. Many engines will ignore multiple submissions, so it's best not to submit again until a month has passed. Once the engine has accepted your site, you need to resubmit every 3 months or so, but there should be editorial changes on your site to justify this. You can optimise your ranking or multiply the links retrieved from your site, by submitting the url of each page on your site. However, some engines require a 24-hour gap between such submissions, otherwise they treat them as spam and ignore them.

http://searchenginewatch.com/webmasters/ includes a five-part instruction guide on search engine submission which supplements the existing submission tips area of Search Engine Watch.

Link popularity

http://searchenginewatch.com suggests that you cannot rely on submitting sites, because this facility has been abused by spammers. Instead, you must build relevant links from good web sites, to increase the likelihood of being found by the engines and included in their listings. Search engines can measure the number of other websites linking to your site i.e. those who believe that your page is important enough to link to. Search engines use this as an indication of the reliability of the information on your page. One way of generating such links is to include a Useful Links page on your site, then invite the hosts of the sites you name, to reciprocate by trading links. Websitegarage includes a link popularity check, but was unable to apply it to the government websites. However, you can

check this yourself by entering <link:> before your url in the search box of each engine.

Spelling

Hurrah! Websitegarage awarded all the government websites “ Good” for their spelling. Terms challenged are only those like UK, FCO, EU, which are not in their dictionary.

HTML Design

Again, there was considerable variation in the performance of the various government sites. Errors may confuse some browsers. Below is websitegarage diagnostic for one site whose HTML design it rated as “ Poor”

Count	Error Messages
8	invalid attribute 'accesskey' in <a>
3	invalid attribute '/' in
1	invalid attribute 'marginheight' in <body>
1	invalid attribute 'xml:lang' in <html>
1	invalid attribute 'valign' in <p>
1	invalid attribute 'marginwidth' in <body>
1	invalid attribute 'margin' in <body>
1	invalid attribute ' ' in <option>
1	invalid attribute 'xm:ns' in <html>
1	invalid attribute ' ' in <tr>
	Count Warning Messages
4	 loads faster with a 'width' attribute.
3	 loads faster with a 'height' attribute.
1	Good HTML style uses an 'alt' attribute in .
1	<table> loads faster with a 'width' attribute.

Errors highlighted most commonly elsewhere were for missing “ close” tags.

Websitegarage rating:

- Excellent** - DEFRA, DWP and DFES
- Good** - DH, DTLR, HSE, govtalk
- Fair** - HO
- Poor** - DTI, MOD, FCO, CO, T, e-Envoy, ukonline

Metadata

Let's not avoid the issue. Search engines are looking for metadata. You can check this by looking at sites which rank highest on a search and viewing their source code.

Advice to most government websites, is that metadata is not present and should be. This is in accordance with the mandated e-Government Interoperability Framework, as well as being in the name of common sense and customer focus. Assuming you have a passing acquaintance with Dublin Core and metadata, there are several tools available which can analyse sites for search engine friendliness. Meta Tag Analyzer on www.scrubtheweb.com and Meta Scanner on <http://www.submitcorner.com/Tools/Metascan/> offer analysis of metadata in headers. Interestingly, Meta Tag Analyzer says that metadata is wrong and will not be captured by search engines, if not presented in the form "keyword keyword" i.e. space but no punctuation. Similarly, the excellent interactive html tutorial at www.davesite.com says "Commas are not needed to separate keywords". This may cause confusion with compound search terms, although presumably search engines will do the equivalent of a free text search in the Keywords tag. The Meta Scanner tool on Submit Corner prompts for the Keyword delimiter used to separate keywords. The default is a comma delimiter but there is an option for space with no punctuation.

When I first read up on keywords about 2 years ago (HTML for dummies?), the instruction was "keyword,keyword" i.e. comma with no space. At the time of writing, the e-Government Metadata Standard offers the model delimiter semi-colon space. It appears that some search engines may not capture some metadata. It was also held to be necessary for keywords used to appear in the first 200 words of the text. This would not appear to be the case universally now, but is still very common.

Looking at pages on <http://www.bubl.ac.uk> which seem to have comprehensive metadata, they separate what they label "AltaVista metadata" from Dublin Core (DC) tags.

I believe that we should not get bogged down with using thesauri for keywords. We need to use meta keywords which the users use. www.scrubtheweb.com provides a Wordtracker keyword report service (you can remove porn terms) to ensure you're using the right keywords in your tags. They run this free every week on the top 500 terms searched on

all major search engines, plus a retro look at the 200 most popular terms for the last six months. Even if terms like "Lord of the Rings" are not relevant to your site, what is useful to see, highlighted, is the need for alternative spellings, misspellings, plurals, alternative capitalisation etc. You can also use the service to find the most used terms beyond the first 500, which are applicable to your site.

The Description tag should include expanded terms, alternative spellings, the title of each document on the site, plus any useful headings. You're a librarian, this is what you're supposed to be good at. It is not necessary with the Description tag to ensure that terms used also appear in the first 200 words of text.

Metadata on government websites

And who are the good guys among government websites who use metadata? Only five of the fifteen viewed - HO, FCO, DWP, T and e-Envoy. How successfully do they use it?

The websitegarage diagnostic advises HO and T that they have missed an opportunity and left blank the META Description tag. The diagnostic also warns that there should be only one <TITLE> tag per page.

DWP are advised that it is permissible to use phrases to boost rankings in their META Keywords tag. The DWP META Description tag uses up less than 50% of the permitted length; up to 200 characters (approximately 25 words) of information are allowed.

Both the HO and DWP META Keywords tag use up less than 20% of the permitted length; up to 1000 characters (approximately 150 words) of information are allowed. HO and DWP are advised that some search engines ignore frame-based pages and it is necessary to ensure that the <NOFRAMES> tag contains information about your site.

HO redirects users to another page. Search tools will usually index the address of the page that users are redirected to - is this the desired effect?

FCO and DWP have repeated terms in the META Keywords tag, which can cause search engines to stop indexing keywords.

Overall rankings and Conclusion

Allowing for the fact that the websitegarage may be one programmer's subjective view of what constitutes a good website, government websites were all subject to the same analysis. Results show that government

websites visited are a mixed bag and generally not very impressive. There is no model of excellence among them. They would do well to learn from each other's good and bad points.

Excellent - None.

Good were: DEFRA and DWP.

Fair were: DFES, DH, DTLR, HSE, HO, MOD, FCO, CO, T, e-Envoy, govtalk

Poor were: DTI and ukonline.

What is to be done? The involvement of Librarians in website design and construction is not universal. Librarians have much-vaunted professional skills in identifying organising and exploiting information resources. More particularly, they have at their disposal the excellent Free Pint, with its tips, procedures, discussion boards etc. Librarians are more likely to be aware than non-librarians of resources like Search Engine Watch, Jimtools, Scrubtheweb, Websitegarage etc Why aren't they using them to advise colleagues or to inform their own work in this area? The excellent <http://searchenginewatch.com> includes very helpful web page design tips among much other useful information.

But most especially, where is the missing metadata from these sites which ensures indexing by search engines and retrieval by users? We are Librarians. This stuff should be bread and butter to us. If it isn't, we should make it our business. It's time to make sure we have these skills and to reclaim the territory.

The author writes in a personal capacity.

Email: johnscottcree@britishlibrary.net

Address: 102 Balcombe Road, Horley, Surrey RH6 9BP

