

Maximising Online Resource Effectiveness

April 2011

JISC Content

A guide to real value from the internet

Producing, promoting and developing content most effectively

Summary

This document has been produced following a very successful series of workshops during 2010, commissioned by the Strategic Content Alliance (SCA) under the title Maximising Online Resource Effectiveness (MORE). The purpose was to promote most effective use of the internet by SCA member organisations, with an emphasis on producing, maintaining, promoting and developing content available as an online resource. The workshops were delivered to around 200 participants from all over the UK, from a variety of further (FE) and higher education (HE) and public sector organisations.

Here we summarise the principles addressed in the workshops and present a condensed version of the two day agenda, designed, as in the workshops, to stimulate thought and action.

It is hoped that readers will recognise the importance of the points highlighted herein and adopt this as a guide to good practice.

Audience

Primarily the SCA member organisations, including all UK FE and HE institutions.

Release date

April 2011

Contents

1. Backdrop	1
2. Fundamentals	3
3. Priorities	5
4. The social web	7
5. Content	9
6. Promotion	11
7. Development	12
8. A checklist for value from the internet	14

1. Backdrop

There's no slow bike coming

If you visit the quayside of Poole on the south coast of England on a particular summer evening you may find almost every inch of road and footpath occupied by a vast array of motorcycles of every kind and beside each a proud owner ready to tell you all about their pride and joy. However you are very unlikely to find a small slow bike among them—they just get bigger and faster every year. So with the internet.



We live in a truly extraordinary time. The human race has never before been so technologically advanced or experienced such an accelerated boom of technical achievements. Nowhere is this more evident than with information technology and in particular the internet.

Over the past ten years the internet has spread to every inhabited place on earth—one continent, Africa, seeing a growth of up to 2000%. In Asia and Africa there is still the lowest penetration, nonetheless Asian users still account for over 40% of total users with this figure rising all the time.

And the most recently connected countries often enjoy better connectivity than those where the internet has been established for much longer. North American and UK users on average have access to much less bandwidth (speed and volume of information transfer) than those in many other countries. Users in South Korea and Japan enjoy average connection speeds of over

50 megabits per second (mbps) compared to around 4 mbps in the UK.

The internet is now ubiquitous, reaching every geographical region and human culture, with current estimates of two billion users.

When we consider what the internet is now used for, we find a myriad of activities and a prodigious amount of information.

A combined 13 hours of video is contributed by individuals all over the world to youtube.com every minute of every day. So, at the time of writing, it's impossible for any one person to watch everything on Youtube—because they would have to sit and do nothing else for well over 400 years!

People are also sharing photographs—3.5 billion photographs on flickr.com now, one for every two people on the planet.

An even more revealing fact is that five billion minutes are spent on Facebook every day.

And it's not just young people participating. The fastest growing age sector is 50-64 year olds.



Furthermore, some estimates put the number of devices now connected on the internet as higher than the two billion human users. Devices that collect and collate information, monitor the world around us, and communicate with other devices or people. The internet has also become an

“internet of things” adding much more to the available information and connection possibilities.

As the internet has grown, other advances in microprocessor and storage technologies provide exponentially growing compute power, processing speed and storage in devices that require less and less space.

So much so that we have now reached a stage where information storage capabilities will allow a person to record on video everything that occurs in their life, and search back through any part of that at any time! In a single iPhone we have around twice the total computing storage that the whole of a typical university required (staff, students, everyone) just 20 years ago! A modern network switch can shift all 13 million articles in Wikipedia in about one microsecond.

The proliferation of computing and network devices has power consumption implications, at a time when energy conservation is a critical consideration for us all. Thus the emergence of virtualisation, with large data centres focussing tiers of physical servers, each hosting multiple and easily regenerated virtual machines. This maximises power efficiency and enables massive computer clouds, capable of running the global reaching services that millions of users can avail of: cloud computing.

This has led to easy access for all to simple social online networking and an explosion of interconnectedness which has “turned marketing on its head forever”. The world wide web is no longer primarily an information resource—it is a means to connect to someone. Many would even suggest that certain

forms of the social web are an important aspect of “quality of life”.

Today we have fast robust connections, with powerful resilient processing at each end, enabling mission critical business over the internet. Media, business, government, almost every aspect of modern society, everywhere, has become “entangled” with it. As a consequence we have a growing, and perhaps dangerous, habitual complex dependence on it.

How do we keep control, address the challenge of transforming burgeoning masses of information into knowledge... ensure that our online activities, and results, are effective?

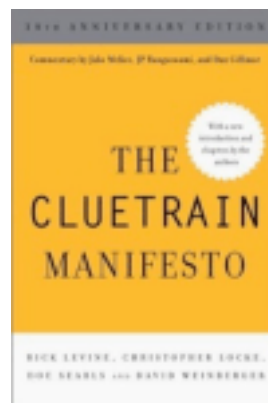
At this point we offer three observations that can be extrapolated to infer useful lessons.

Firstly, more can be much better than less.

We could drown in all of the information now accessible to us, but perhaps with very simple tools, more information can be utilised to produce “better” knowledge. This may require disciplines in the preparation of information so that inherent structures and order can be drawn on to create new visualisations of the whole.

Secondly, a new way of doing successful business has emerged.

With the new regime of internet connectivity, cloud computing, and ubiquitous easy access, a new breed of business has evolved. Lightweight companies, delivering solutions over the world wide web, focussing on doing less, but to a high quality standard, and delivering to many more customers with less overheads.



This totally confirms the insights of the authors of the “Cluetrain Manifesto”, first published in 1999 and again in 2009 (because it was at least as relevant ten years later). Today's new successful companies understand that the internet brings direct connection with an audience, the opportunity for meaningful two way relationships—an approach more like a “table for two” than fleets of trucks delivering loads of widgets to distributorships all over the country.

Thirdly, it is possible to develop online resources to build a more “intelligent” web.

With enough embedded “semantics” to make information “computable” we could begin to extract knowledge from masses of data and gain views and insights never before possible.

Conceivably a “sixth sense” type of device could help us make decisions on a daily basis by observing circumstances and retrieving related information from the information “cloud”, in much the same way that the WolframAlpha resource gleans information from independent curated data.



If all web data was computable, it would be possible to learn from every merging of information resources, user activity, user

and device exchange, and more, occurring online, to provide answers to any reasonable question.

Summarising:

Recognise advantages in having well prepared scalable content that can be utilised in “more is better” scenarios.

Understand the potential of audience engagement using the web.

Consider the longer term benefits of having computable content.

2. Fundamentals

Competing against the millions

The Great North Run in NE England every autumn is the largest road race in the world with more than 50,000 runners every year. The average punter would find it difficult to win the race or even be placed anywhere near the top ten. However this is nothing compared to the challenge of trying to get a web site on the first page of a set of search engine results!



We should not go any further without dealing with Search Engine Optimisation (SEO)—for two reasons: to show how important it is; and to show how unimportant it could also be.

Firstly, to clarify what we are trying to do—maximise a web site’s effectiveness—the following are assumed:

- you have (or will have) an online resource, typically a web site
- to which you want to attract visitors
- the right kind of visitors
- who will find all that’s relevant to them on your site
- and get real value from what they find
- then tell others, who will do the same
- and you will obtain evidence of this

- so that you can justify the cost and other overheads of keeping the resource going
- and improve it
- and perhaps make it more important
- by being the definitive source of something
- in a bigger picture

Starting with attracting visitors. A proportion of web users today currently default to their search engine of choice to direct them to everything they view online. Some of them may not even be aware of URLs (uniform resource locators, the formal name of web site addresses) and how to access a particular web site directly. They simply type in search terms and browse the search results presented to them.

If we have a web site to which we wish to attract visitors then we should not ignore search engines. There are fundamental points to be aware of, to avoid being disadvantaged.

At the outset, It is useful to recognise that there are two approaches to attracting visitors. One is to provide very basic, accurate and honest signposts. The other is to try to captivate using glitzy or sensational billboards that may ultimately be very misleading. The former is much more likely to satisfy the earnest web user in search of specific or even general information. The latter is quite unlikely to do so. The general public has become quite “savvy” regarding conventional advertising, with, for example, some studies suggesting that many television viewers would gladly pay for the removal of adverts.

With search engines, the provider’s aim is to always satisfy the user—their success depends on it—thus their adopted approach will be to promote the accurate signposts and demote the misleading billboards!

So, a common sense principle is that what is good for the reader will most probably also serve to improve rankings with search engines.

Of course there are many search engines available to us. However the market is dominated by a small number with Google having the largest share. In the US Google’s share is around 65% with Yahoo about 16% and Bing 12%. Smaller players include Ask and AOL (about 6% between them). Globally Google is even more prominent, around 86% of the market with Yahoo, Bing and the Asian Baidu, claiming around 3-5% each.



Every search engine employs different (secret) algorithms to determine how search results are ranked. However, on the whole the differences are

insignificant and the Google guidelines, summarised below, are applicable to all.

Perhaps the most important check is to ensure that an appropriate “title” and “description” is available for every page of a web site. These are what appear in search engine results and will ultimately guide a web user to select one site or another from the list on their screen.

Every page should have a different and representative title that reflects the content on that page, contained between <title> and </title> HTML (hypertext markup language) tags in the head of the web page source. Every page should have a different and representative description, contained in the content attribute of an HTML description meta tag in the head of the web page source, i.e. <meta name=“description” content=“...”>.

If a content management system (CMS) is being used that doesn't permit titles and descriptions to be set for each page, then it's time to change it!

The following additional checks have been highlighted from Google's recommendations.

Encourage all "appropriate" links to your site from other web sites.

This reinforces the importance of your site and provided the links are relevant (connecting similar content subjects) will promote your site's ranking in search results.

Use a sitemap file to enable more intelligent indexing of your web site.

A sitemap is an XML (eXtended Markup Language) file that lists URLs for a site along with additional metadata about each URL—when last updated, how often it is likely to change, how important it is compared to all others. Not to be confused with a sitemap page created for visitors to your site, this file is used only by indexing software and enables more intelligent indexing. It is necessary when different sections may not have explicit links between them (e.g. sections meant for different geographical audiences) or when it is undesirable to index certain pages (e.g. empty page month and year permutations in a calendar section). Free web site services will build sitemap files for a starting web address. These can be edited with a simple text editor as required.

Use keywords.

Keywords that summarise a web page can indicate a strong match to search terms provided by search engine users and therefore help increase results rankings. Keywords can be added to the head of a web page source using a `<meta name="keywords" content="...">` HTML tag. These should be different for every page. A "keyword" may consist of several words in a phrase. Keywords are separated by commas. Google provides web based services for deciding on appropriate keywords for particular pages and viewing how often each keyword has been used by web users searching using Google (all data made available for local region and globally).

A recurring theme is to think "text". Use "words" that can be processed, rather than images (or flash movies) which cannot convey the same information.

Test web pages to ensure that all pertinent information is being shown when viewed as text only. One easy way to do this is to select the cached link below an item in a list of Google search results. This will show what content from that page is currently cached at Google. A further link is then presented (top right of screen) to view the text only version of this cached content. If the web page makes sense when viewed in this way, then it has been produced with sufficient text for satisfactory discovery and indexing.

Make sure that all text HTML elements, title, description, keywords, ALT attributes, are descriptive and represent content of web page.

When text is used in key parts of a web page, particularly title, description, keywords and headings, make sure that the words used do also appear in the content elsewhere to confirm that they represent the content properly.

Use a robots.txt file to stop irrelevant files on the web server from being indexed.

A simple text file in the document root of the web server can identify any directories that should not be visited by indexing software—for example test pages that are not meant for public consumption.

Test your site to make sure that it appears correctly in different browsers.



Unfortunately different browsers may render web pages differently to the extent that a web page that looks exactly as intended in one browser, may appear totally distorted in another. Check!



Use Google Analytics, or a similar method, to monitor visits to a web site and measure how successful it is.

It can be challenging, but very worthwhile, to agree a "conversion" metric that would objectively record success for any user visit. This might be time spent on the site, download of important documents, pages viewed during the visit. Some indication is better than none and helpful when assessing the benefits of any changes to a site.



All of the above measures will ensure the fundamentals of drawing visitors to your web site from a search engine such as Google. However these only partially satisfy the desirables listed at the start of this section.

To approach maximising effectiveness, overlapping areas should be addressed more fully. These could be categorised in several ways: we have chosen the following which follow a general thesis of good practice in producing content for an online resource, plus promotion and development:

- web page and site structure
- accessibility
- content integrity
- metadata inclusion
- conveying richer semantics
- involving online social media

These are expanded in the next section 3.

Summarising:

Search engine providers want their users to find exactly what they are looking for, so describe your content accurately using titles, description, keywords.

Use text for all important content.

Monitor and measure how your site is being used and define success.

Search engine optimisation cannot be ignored, but it is not everything.

3. Priorities

Build resources well

The cathedral of Saint James in Santiago De Compestela in NW Spain has been visited by pilgrims for centuries, some of them traveling 500 miles on foot to reach their goal. It is a magnificent edifice, built with incredible attention to detail over more than a millennium. And it stands today drawing the same attention that it always has done.



A report in January 2010 claimed that “shoddy” web sites were costing UK councils around £11 million per month, simply because the web sites were not functioning properly. Users could not find what they knew was supposed to be there, and staff were spending excessive time relaying information on the phone. Even if the web sites appeared at the top of search engine results, much more was needed.

The highest priority is to identify the target audience(s) and ensure the online resource is built for those users, supplying the content they require, in a way that they can discover, access and use it easily.



This involves getting to know audiences—as comprehensively as possible—how they work, what they value, what they expect from your web site (or other resource). These considerations inform everything else.

After audience, structure is so important. It underpins everything in the build. Structure considerations are wide ranging and involve:

- elements of a single web page and their relative importance and readability

- elements of a web site, the sections that together give a user the desired overall perception
- navigation options for a web site
- separation of styles and content using cascading style sheets (CSS) so that content or styling can be modified without effect on the other
- separation of programming (e.g. javascript) and content so that content always remains intact and independent of scripts
- resources that are derived from a web site (e.g. RSS feed)
- options for adding metadata

Much of the latest HTML5 and CSS3 standards work from the World Wide Web Consortium (W3C) is aimed at improving structure in web pages and by so doing enhancing content meaning (e.g. a “footer” should not be considered as central to the information as a main “section”; a “nav” block is not part of the intrinsic content).

With HTML5 old tags that were not designed to label parts of the page (e.g. <center>,) have been dropped, and many new ones to identify specific elements on a page have been introduced (e.g. <section>, <article>, <nav>, <header>, <footer>, <video>, <audio>).



Likewise, CSS3 allows “rich” styling with animation effects, if desired, in order to eliminate the need for graphical text and Flash in building an attractively designed, engaging and more meaningful site.

The result is more portable, easily communicated content because of the inherent structure.

There is some reluctance to adopt the new standards because they have not been finalised and web browser providers have embraced them to different extents. However advocates have developed javascript libraries that will simulate most of the new functionality so that much of the core HTML5 and CSS3 markup can be used with non compliant web browsers by preloading this javascript. With recent developments (speed) to the javascript engines in every browser, the processing overhead is usually completely unnoticed by a user.

Examples of workarounds for browsers that don't support the new standards include scripts supplied by Google for using <video> and <audio> tags, and Google Chrome Frame for Microsoft Internet Explorer versions.



Another aspect of structuring a web site that can be very helpful in extending the reach of a web site, is arranging an RSS (Really Simple Syndication) feed.

This requires an XML file with RSS markup that identifies different “chunks” of information—publication date, author and a link back to a source URL. The feed is normally used for items on a news or updates page in a web site. Many content management systems will generate the required RSS file automatically. Others may require some programming to link the RSS to content.

Once in place the RSS feed can be selected and used on an ongoing basis, by anyone using an RSS feed reader, which usually aggregates many feeds into a single readable list so that news from many sources can be easily reviewed.

When it comes to accessibility, all of the following can have a positive (or negative) impact:

4. The social web

Getting involved

It's happening all around us, people are interacting and engaging with each other in many different ways. We can ignore and be left out, or jump in and connect ourselves.



To many the social web of Twitter, Facebook, Youtube, Flickr, Slideshare, LinkedIn, and several hundred more “services in the cloud”, is something very different and separate from professional online resources. However, today there are many examples of the social web being used to significantly increase the effectiveness of other online resources, or, in some cases, achieve the same or better results.

Nonetheless, social web evangelists should be regarded with a healthy dose of pragmatism which examines whether or how it can work in a particular set of circumstances.

The world wide web is now 20 years old. In the space of that period the world has become a much smaller place because of it—almost any web user can communicate directly with another.

From the very start, Tim Berners-Lee and Robert Cailliau were adamant that the web was not a broadcast medium, rather an interactive, two-way, communication platform.

The web was architected for participation.

As the number of web users and web sites grew exponentially, the urgency was to find and visit the sites of most interest. Very few waited on the platform to discuss matters with others and share recommendations on the best destinations.

All this changed with what is popularly called “Web 2.0”, and the appearance of social networking sites in the web “cloud” freely serving unlimited users. Suddenly there was enough space and interesting “stuff” on the platform to make people want to hang around there.

Through the connections that are made between users, clusters and communities have emerged. Members of such groups recognise they are on “neutral territory” in that all users have equal authority to communicate, discuss, and share ideas, knowledge, observations or just opinions.

And the platform is open 24 hours a day. And people can be there at anytime—often by using a mobile phone app, but in general from any internet connection, anywhere in the world.

Furthermore ordinary people actually want to be part of the social web, because there they find news with qualifying comments, a community that appreciates what they have to offer, pointers to information they can rely on from sources they trust, recommendations based on experience (not keyword matches).

The social web is an open public place where most activities are transparent to all and self regulation is the name of the game.

So what are these popular services and what are the differences between them?



There is a bewildering number of social web sites and services—the attitude of a new user could easily be like that of a “kid in a sweet shop”.

Rather than getting caught up in time consuming diversions, it's important to decide your purpose in using the social web—or perhaps just simply what things are worth sharing in this way.

Others are already actively participating in the following ways:

- blogs, or “web logs”, often led by one or a small group of bloggers with each log entry open for comments from anyone—an opportunity to create and reinforce a reputation, either by leading a blog or offering comments on several blogs (Wordpress is probably the best known service for creating blogs but there are many others)
- microblogging, as inaugurated by Twitter, where each microblog or “tweet” is limited to 140 text characters, encouraging wide participation, often providing pointers to other resources or users, with options for following particular users and seeing only their tweets
- sharing audio (podcasting) by building a library within a common repository, such as iTunesU, or offering recordings for use in other large multiuser projects
- sharing video, as first enabled by Youtube (the largest video repository, but there are many more)—there is so much useful video content on Youtube that many people use it as a search engine and expect answers to be shown to them on video
- sharing career information, as with LinkedIn, and gaining recommendations from employers or associates that everyone can read
- photo sharing, which became so much easier and more social with Flickr—other services existed many years before Flickr but did not provide such a public stage
- sharing presentations (Slideshare is the best known service, others available)



- sharing bookmarks (e.g. Del.icio.us) where the bookmark is a pointer to any other online resource
- sharing event or other diary information (with e.g. Google Calendar) where invitations can be sent that work with all popular desktop calendars
- using Facebook, which aspires to be the ultimate social web site by allowing users to do almost everything they can do at other sites and more



What are the benefits of using these services?

There are some very clear and tangible benefits from employing social web sites.

The cost involved is low, provided, of course, that the time required is not debilitating.

The reach is very wide. Most popular social networking sites have millions of users. If Facebook was a country it would be the third largest in the world.

Services such as Twitter can provide one of the best online personal learning environments. In addition the opportunity to connect to like minded people anywhere is huge.

The effectiveness of passed links via Twitter is greater than those in search engine results because the Twitter based links are usually pre-filtered, coming as a recommendation from someone who is followed.

Resources, such as a photograph collection, can be made easy to find and share by using a social sharing service, at less cost and in a way that engages or even employs other users to add to and improve it. This has been done, for example, with Oxford University's Great War Archive (Flickr).

Many companies claim they have gained significant revenue through social networking and targeted marketing (e.g. Dell on Twitter).

There are also reports of new companies having received funding as a result of promotion on LinkedIn alone.

There are ongoing activities to encourage better engagement with art bodies (e.g. museums) in order to bolster subscriptions and donations—essential in an economic climate that mandates public spending cuts. Such have been reported in national press articles and are common in the US.

What concerns are there?

The social web is primarily about connecting individuals. Participation involves some personal exposure. The user is responsible for the image of themselves they create from what they say and share. For some this is a challenge—they feel they are not trendy enough to go public, or don't have anything interesting to share.

However the biggest concerns about the social web tend to arise from the fuzzy interface between individual and organisation.

An organisation may be very uncomfortable with a member speaking freely or sharing inappropriate material on the social web, when that person is known to be associated with the organisation, or perhaps is using the organisation's resources (computer, phone, network) to participate.

An organisation's view of what is suitable for broadcast does not always match an individual's.

There have been a number of reports of staff losing their jobs because of inadvertent comments made publicly on the social web.

An organisation may also have legitimate concern over the time that individuals may spend on the social web, if use of such was condoned.

In cases where an organisation has embraced the social web, recognising the benefits and coaching staff on how to participate, there is still the challenge of measuring impact, success and justification for investment.

The social web also provides members of an organisation with direct access to services in the cloud which may not be available locally. Most commonly this would be a blogging platform, with, for example, many UK university blogs hosted on public sites. In this way organisation restrictions or policies can be circumvented.

From the individual's perspective, of course, this is also a way of avoiding frustrating exchanges with the local systems administrators and saving everyone time and money.

Organisations may also have already invested in more conventional marketing that appears to be working well and they don't feel that there is enough resources to do something else as well.

Despite concerns and organisational issues that may require resolution (see Section 6), the social web can be used to engage an audience, gain their trust, promote existing online resources, and serve as an extension of those online resources.

We cannot maximise the effectiveness of online resources without the social web.

Summarising:

The social web is not new, it's what the web was always intended to become.

Audiences are already in the social web—it's the best place to engage with them.

Decide on a purpose for adopting the social web and use the best service for that.

Be aware of valid organisational concerns over the use of the social web.

5. Content

There's more to it than you think

Riding a unicycle, especially in front of an audience, requires a lot of concentration and being mindful of everything that is going on around. There is the potential to really impress, or to come a "cropper". So it is with providing content online.



In sections 2 and 3 we have covered fundamentals and priorities when preparing, publishing and maintaining content. In this section we add some additional points and consider issues relating to managing content across an organisation involving multiple players.

What expertise is needed within an organisation to maximise the effectiveness of that organisations's online resources?

Let us first consider the priority of getting to know the audience. To do this well may involve a detailed study, empathy with users, ability to articulate needs to content providers and argue the user case. And with multiple audiences, different individuals may have to be included to empathise with each audience.

That is usually a very different kind of person than a systems administrator who must be responsible for managing the web server or network operation, both of which can have such a drastic effect on the response of a web site.

If all other guidelines have been followed, total failure can still ensue if a web site is so slow to respond that users refuse to wait.

On occasions slow web site response may be due to content being too heavy, simply because the content providers have not understood technical issues. For example typical digital cameras produce 10 Mb files that some may upload directly to

a web site, causing unnecessary long load times and impairing many other users by hogging network bandwidth.

Photographs and other graphics should always be optimised for the web, which will reduce file sizes significantly (possibly one hundredfold, increasing loading speed by the same factor).

Non technical content providers may not realise that multiple images, CSS file references, javascript links and other web page resources (often mandated by a content management system) all create additional download overheads because each resource is retrieved using the hypertext transfer protocol (HTTP) with header information that may sometimes be more than the size of the resource being retrieved (e.g. a small icon image). In cases like this a CSS programmer can build CSS sprites which allow all images to be loaded in a single file then that file positioned using CSS to show appropriate parts at the correct places on the page. (A ploy used by large web sites, with many users, to maximise page response).

When it comes to enhancing the user experience with web pages, it's unlikely that most content providers will have sufficient javascript programming capability to implement javascript AJAX calls back to the server. This would allow, for example, all necessary information to be supplied and checked on a single form page, with context sensitive help messages. Measures such as this can make a web site much more effective.

The official W3C accessibility guidelines provide particular scenarios of users with different disabilities and how to make allowances for each.

Implementing full accessibility is not a well defined art and requires an understanding of a user with difficulties as well as the technical insight on how to adjust particular content.

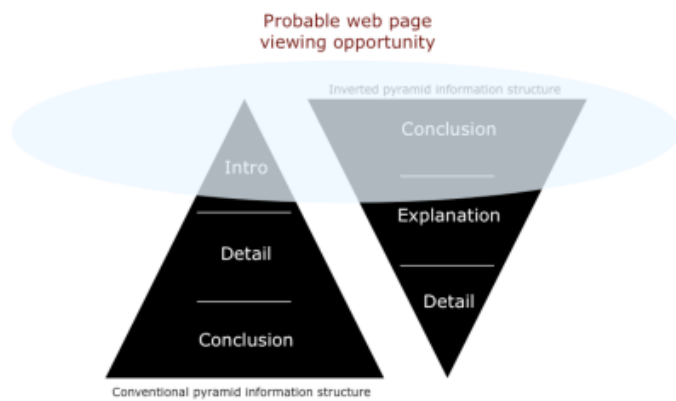
With regard to content, authors must realise:

- often web users "happen" upon the content and will instantly judge the value of a site by fast skim reading for clues
- users are driven by their own goals, what they want to know, not necessarily what the web site owner wants them to know
- users will review headings
- users read about 20% of a web page

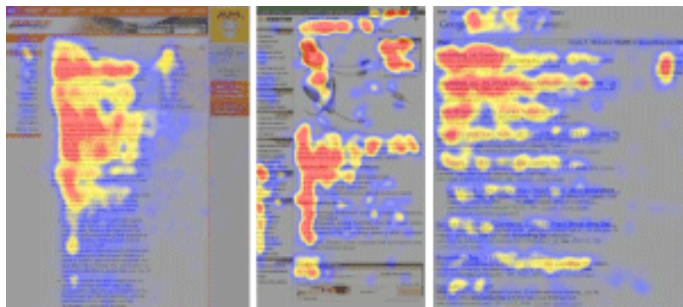
Writing style is usually left to the discretion of authors. However with web content, the author may not fully appreciate that writing for the web is not the same as conventional writing.

Content authors should be encouraged and instructed to do the following:

- be concise—focusing on what the site visitor is interested in
- capture interest, then add detail
- provide useful scannable headlines, headings, emphasised keywords
- use lists instead of long paragraphs
- ensure links are perceived for what they are (not "click here")
- make images meaningful (and include description using "alt" in HTML tag)
- be consistent, and by so doing instill confidence in the web site rather than confusion



- adopt the inverted pyramid approach with which the most important content is presented first, followed by the next most important, contrary to conventional reporting methods
- be aware of the F shape reading pattern—research work by Jakob Nielsen in 2006, which tracked web users eye movements, concluded that users tend to read most web page layouts in an F shape: across, then down, then across again but not so much, then down



All the following considerations can be addressed within an organisational web content strategy:

- what are we trying to achieve
- what do we know about our audience(s)
- are there cultural or political considerations to observe
- how can visitors find all content easily
- what human and other resources are available for content processes
- how and by whom will content be maintained
- have we competitors and what are they doing as compared to us
- are we needlessly duplicating content that is already available elsewhere
- how should editorial control be exercised (if at all) without offending content authors
- what is the overall structure that all content fits into
- what writing style is most appropriate
- are there viewing size imposed or other restrictions for content
- what is our measure of success

The provision and coordination of sufficient expertise to maximise the effectiveness of the organisation's online resources requires roles and responsibilities in the following areas:

- technical, including server, network, programming, CSS, javascript
- editorial, including quality assurance, appropriate style and layout for the web, addressing the right audience(s), and satisfying the organisation's goals

- strategic, defining authorities within the organisation, maintaining overall structure, coordinating roles, measuring success

In any one organisation these areas will overlap in different ways according to available skills.

IPR and Licensing

If you are thinking about preparing, publishing and maintaining content, you should consider the following:

- Have you created all the content or used something someone else has created?
- If it has been created by someone else, you will need to look for a Creative Commons Licence, or similar terms of use before you can use it
- Make sure that you read the licensing terms carefully to ensure that you can use the content in the ways you want.
- If the third party content you wish to use does not stipulate terms of use or does not allow you to use the content that you would like, you will either need to contact the rights holder/s for further permission or find alternative content that you can use

Remember:

- To facilitate legitimate use of the content that you are publishing, and ensure that it has the maximum reach and use, you should choose an end use licence which outlines clearly how you want all your content to be accessed and used by your end users.
- Creative Commons Licences are ideal for stipulating pre-defined terms of use associated with creative works
- Any use that you make of third party content and/or uses that you wish end users to make of content, needs to be compatible with the terms of any end-use licence that you select
- You cannot provide access to third party content that is more than the permissions that you have been granted by any third party rights holders
- Content can comprise of multiple elements created by more than more person. You will need to make sure you have permissions from everyone whose content you wish to use before you can publish it
- You can find out more about IPR and Licensing from:
 - Creative Commons www.creativecommons.org.uk
 - SCA IPR and Licensing Toolkit <http://sca.jiscinvolve.org/wp/ipr-publications/>
 - JISC Legal www.jisclegal.ac.uk
 - Resources on the legal aspects of social networking, including a Web 2.0 Legal Issues Toolkit www.jisclegal.ac.uk/Themes/Web20.aspx
 - Web2Rights Project www.web2rights.org.uk

Summarising:

A mix of expertise is required to maximise effectiveness.

This expertise should be associated with different roles and responsibilities.

The coordination of these roles should be an essential part of the web strategy of an organisation.

6. Promotion

Social policy

If an organisation is to appear “organised” and deliver a strong message to audiences, then all players must play the same tune, even if they are using different instruments. That means a degree of cooperation, agreement and subscription to a common code.



In section 4 we highlighted possible concerns around the use of the social web as an “official” activity within an organisation.

These concerns can be mitigated by requiring conformance by all to a social media policy.

Many such policies have been publicised by different organisations, varying in complexity from extremely simple and lightweight (“don’t do anything stupid”) to heavy bureaucratic documents.

This reflects the dilemma of allowing organic growth or imposing control.

The right answer can only be ascertained by reviewing the purpose of utilising the social web in the context of the business of the organisation, resources available, strategic direction and related factors.

This essential review can occur in the compilation of a social media policy. However lightweight this may end up, it should be the result of a reasoning process.

In the following we highlight suggested elements of that reasoning process.

Match resources and activities to appropriate social web services.

List organisation resources and associated activities (news, reports, presentations, photos, videos, bookmarks) which would be suitable for use with the social web and try to match these to appropriate services (blog, Youtube and so on).

Confirm the rationale for choosing each selected service.

What does the chosen service offer in helping engagement with the target audience.

Specify the goal associated with each service.

Is the primary intention to inform, educate, persuade, influence, encourage uptake, manage expectations, listen, discuss, synthesize, amplify—and does this suggest anything special about how the service should be used.

Agree a desirable communication style.

What kind of communication style suits each service and goal and target audience—formal or informal; serious or trivial; announcement or conversational; factual or opinionated; spontaneous or considered; detailed or brief.

Determine metrics.

How will target audience engagement be measured: views or “hits”; click-throughs; re-tweets; number of followers; subscribers; volume of comments; downloads—and how frequently will measurements be made or formally reviewed. Promotion is a continuous process and as such should be assessed continuously to ensure the efficacy of chosen approaches.

Appropriate integration with other organisational activities.

What should match and be consistent with other official organisation communications and how can this be arranged and coordinated—will posts be syndicated, cross-posted, aggregated, filtered, amplified, reused. Often online activities can complement offline efforts and vice versa and both should be considered in a complete strategy.

Short description of the organisation and its aims in using the social web.

It can be useful to write a public-facing mini bio for the organisation (140 characters, a paragraph or a page) using language appropriate to the service to tell users what they should expect from the organisation through that service and why they should care.

Required social web use disciplines.

It is important to provide guidelines for those members of the organisation who will be taking responsibility for representing the organisation on the social web. These should include:

- what sort of things should be communicated and what should not
- how often should something be communicated or what are the triggers (events, progress, related posts by other people or projects, milestones) that should prompt communication



- is the plan to be proactive or to primarily respond to the audience
- who is ultimately responsible for the content of any communication and what procedures, if any, should be followed for specific services—proofing, link shortening, adding keyword tags or other metadata, cross-posting
- how to respond to positive or negative comments
- any house style stipulations—colours, images, avatar, background, contact information, biographical detail format
- identity management—when can a person act (speak, publish, share) as an individual and when should it be as the organisation

Summarising:

A policy can allay concerns over the use of the social web by an organisation.

There is no magic formula for a organisation wide social web policy—it depends on many organisation specific factors.

The process of compiling a policy should involve a probing review that brings focus to benefits and workable processes.

7. Development

Call a spade a spade

Communication, and related information technology, becomes potent when “things” are described simply with unambiguous terms.



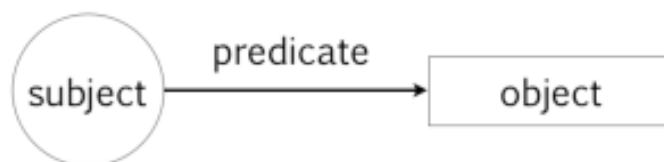
In this final section we will attempt to explain how web content can be made “computable” and why this should be desirable.

We have already made related references in earlier sections. In section 1 we mentioned the possibilities of exploiting ever increasing masses of web based data to extract new knowledge, how suitable order and structure in web data can make the web more intelligent. In section 2 we included “making content part of something bigger” as an aspiration in maximising its effectiveness. In section 3 we indicated that metadata is important in these cases, especially metadata embedded in content which allows software to identify “things” within the

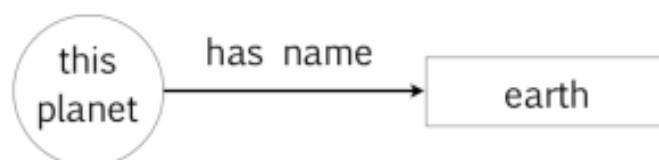
content and do something with them. We referred to RDFa as a mechanism for doing this.

RDFa is simply a way of using the Resource Description Framework (RDF) in HTML. The “a” relates to HTML tag attributes. It uses existing and new extensions to standard HTML tags to include labeling of “things” in the content.

RDF itself is nothing more than a very simple scheme for describing things. With it we associate attributes with some thing. The thing in question is called the “subject”, the attribute is the “predicate” and the value of the attribute is the “object”.



This succinct construct is known as an “RDF triple”. Many triples can be provided for the same subject and triples can be added at any stage. With this triple scheme, anything can be described.



The subject in an RDF triple is identified by a unique identifier called a URI (uniform resource identifier). In a web context it is helpful if this is a web address which makes everything so much more computable, but that is not mandatory.

The object is often a literal value. However it could also be another unique identifier which itself could be the subject of RDF triples.

This option to inter-relate triples about the same or different subjects means that potentially “everything” we know about could be connected, with defined relationships, in a set of RDF triples (a triple store).

A very complex map of how things are related is built from the simplest statements (triples). And the map is completely scalable. With additional triples more relationships appear and the specificity of related subjects increases.

One component of the system requires further note—the disambiguation of predicates.

Unique identifiers and literal values are obvious and understood to everyone, in any language. However the meaning of a particular predicate must be made completely unambiguous if the RDF system is to be used meaningfully by many different people.

This is resolved by the creation and use of standard metadata vocabularies. In these particular predicates are defined clearly (often using other metadata vocabularies). Vocabularies are XML files with a web address that can be referenced in any web page.

Therefore, with RDF triples embedded in web page content using the RDFa standard format, we can have rich metadata and a direct link to the explanation of what that metadata means—all written in a way that software can make sense of and use, i.e. computable.

Hundreds of vocabularies have been created on the web today for use in particular situations. Some of the most common popular ones are listed below along with the type of subject(s) they are used with:

- Google, covering person, organisation, review, event, and recipe
- FOAF (Friend Of A Friend), for person
- GoodRelations, for entities in ecommerce
- Dublin Core, for generic document
- Creative Commons, for licensing statements
- SKOS (Simple Knowledge Organisation System) for higher level concepts

RDFa can be added manually to a web page, but this is likely to be a tedious task for all pages in a web site. Development environments (e.g. Ruby on Rails) and updates to content management systems (e.g. Drupal 7) are now providing native RDFa support. At the time of publication RDFa can only be formally validated when using the XHTML and HTML5 standards.

Google is also very active in this area. Since the introduction of “rich snippets” in Google search results in mid 2009, the Google indexer has recognised RDFa using certain vocabularies and used the embedded metadata to present search results in a more meaningful, useful way to users.

Google also provides online tutorials for adopting and using RDFa in web pages, including web based tools for generating the RDFa HTML code.

The goals of RDFa are similar to the purpose of Microformats, a similar though older method of embedding metadata, mostly for person details and calendar events. RDFa brings more integrity, scalability and robustness by referencing the important vocabularies that make terms definitive. With Microformats there was the assumption that everyone knew what the vocabulary meant.

The common theme has been to make content accessible to software so that interesting integration, combining, “mashing up” of content becomes possible. Content can be reused in different ways, most of which haven’t even occurred to the authors yet!

Is the above an academic discussion or is there evidence of real use and benefits?

During 2009, Mark Birbeck, the initial proposer of the RDFa standard, was engaged by the Central Office of Information (COI) in the UK to assist on a project to consolidate public sector employment opportunities, which at the time were spread across hundreds of web sites, the content being generated from many disparate database systems.

A traditional approach would probably have been to try and rationalise the number of back end systems used with expensive purchase, installation and training costs, and taking a long time. Mark succeeded in avoiding this by encouraging a few tweaks on web publishing systems to include RDFa, then using the web as the integration layer by harvesting the RDFa data and combining on one web site. All this at a fraction of the cost and time of a traditional solution.

In January 2010 The Stationery Office (TSO) announced a commitment to using RDF to improve management of the vast archives of publications under their curation.

Around the same time in 2010 observant users of the Tesco web site realised that Tesco had adopted RDFa.

Later in 2010 Facebook announced adoption of the Open Graph Protocol, (RDF using Facebook sponsored vocabularies) to build the “social graph”, a global mapping of everyone and how they’re related.

In the summer of 2010 the School of Electronics and Computer Science at University of Southampton announced that all publicly available information will be published using RDF standards.



One of the most impressive RDF based projects is

the Linking Open Data project. Here the goal for several years now has been to extend the web by publishing various open data sets as RDF and setting RDF links between the triple stores. For example, links exist when different attributes are provided about the same unique subject in triples from different sources, or where the object of a triple in one source may be the subject of a triple in another. These RDF links then enable navigation from a data item in one data source to related data items in other sources.

Billions of RDF triples exist in this project with hundreds of millions of RDF links between sources.

One of the most linked sources in the project is DBpedia, which is an RDF version of Wikipedia, regularly updated from the live source.

A selection of web based services and browser plugins are freely available for working with RDF, RDFa or exploring linked open data sources, such as DBpedia.

One further standard worth a mention is SPARQL, a protocol and RDF query language for retrieving and manipulating data from RDF triple stores.

More compelling than any specific examples, we are currently observing on the web another phase of its evolution. This has been dubbed “Open Data Movement.”, “Linked Open Data”, “Web 3.0” and “Semantic Web”, often interchangeably but also to point out differences between such names.



The Open Data Movement is generally associated with government (chiefly US and UK but others are involved). This is all about getting data into the public domain— information that can potentially be linked with other data sources and used with software applications to extract new knowledge or build ever more useful and helpful web services.

Linked Open Data refers to when data includes RDF triples and when RDF links can be created between data sources.

Web 3.0 and the Semantic Web are essentially a new and old name for the same thing. This is the web platform that includes sufficient rich computable content to enable a plethora of new powerful web applications that can utilise content from many sources.

In the words of the web’s creator, Sir Tim Berners-Lee, speaking in relation to these developments: “We’ve only just started”.

Summarising:

RDF is a very basic scheme for describing things using unambiguous terms in brief statements known as triples.

RDFa is a way of including RDF in an ordinary web page to embed metadata.

RDF metadata in a web site can be used by software applications to detect semantics.

Web sites containing RDF can be linked when there is overlap in the triples.

There is a growing number of online resources using RDF and real semantics to create a more effective web of resources.

This document is available from the [SCAMORE online repository](#). At this web address additional supporting material is available, including presentations which provide more details about the topics mentioned herein and descriptions of, and references to, many examples.

It has been produced with a Creative Commons Attribution license which lets others distribute, remix, tweak, and build upon this work, even commercially, as long as they credit this source for the original creation.

This document was produced for the Strategic Content Alliance (SCA) by JISC Netskills.

Strategic Content Alliance, Stuart Dempster:
s.dempster@jisc.ac.uk

JISC Netskills, David Hartland: d.w.hartland@ncl.ac.uk.

8. A checklist for value from the internet

Producing, promoting and developing content most effectively

1. Recognise advantages in having well prepared scalable content that can be utilised in “more is better” scenarios.
2. Understand the potential of audience engagement using the web.
3. Consider the longer term benefits of having computable content.
4. Search engine providers want their users to find exactly what they are looking for, so describe your content accurately using titles, description, keywords.
5. Use text for all important content.
6. Monitor and measure how your site is being used and define success.
7. Search engine optimisation cannot be ignored, but it is not everything.
8. Know your audience.
9. New standards enhance the value of content by enabling informative structure.
10. Many benefits of new standards can be realised with older browsers by referencing ready made non intrusive javascript.
11. RSS can extend the reach of suitable web site content.
12. Keep things simple on a web page to prevent creating barriers to accessibility.
13. Employ simple web based services to check the integrity of content and associated keywords.
14. Embedded metadata can open up new possibilities for the use of content.
15. Online social media can play a prominent role in attracting and engaging an audience.
16. The social web is not new, it’s what the web was always intended to become.
17. Audiences are already in the social web—it’s the best place to engage with them.
18. Decide on a purpose for adopting the social web and use the best service for that.
19. Be aware of valid organisational concerns over the use of the social web.
20. A mix of expertise is required to maximise effectiveness.
21. This expertise should be associated with different roles and responsibilities.
22. The coordination of these roles should be an essential part of the web strategy of an organisation.
23. A policy can allay concerns over the use of the social web by an organisation.
24. There is no magic formula for a organisation wide social web policy—it depends on many organisation specific factors.

-
- 25.** The process of compiling a policy should involve a probing review that brings focus to benefits and workable processes.
-
- 26.** RDF is a very basic scheme for describing things using unambiguous terms in brief statements known as triples.
-
- 27.** RDFa is a way of including RDF in an ordinary web page to embed metadata.
-
- 28.** RDF metadata in a web site can be used by software applications to detect semantics.
-
- 29.** Web sites containing RDF can be linked when there is overlap in the triples.
-
- 30.** There is a growing number of online resources using RDF and real semantics to create a more effective web of resources.
-