





# Contents

Conventions used in this document.....	4
Chapter 1: Introduction.....	5
Objective .....	5
Who is it for? .....	6
Chapter 2: Why sustain web resources?.....	7
How and when to sustain? .....	7
What are “Sustainable web resources” all about? .....	8
Some practical issues: experiences from ReStore.....	9
Chapter 3: Types of web resources.....	10
What is a web resource all about? .....	10
Simple project sites .....	10
Reference sites .....	10
Resource sites .....	10
Content management systems and virtual learning environments .....	11
Personal web sites, blogs, twitter, etc. ....	11
Chapter 4: Alternative initiatives on archiving/preserving/sustaining online resources .....	12
Maintenance by universities.....	12
Maintenance by a commercial company.....	12
Maintenance by a community of users.....	12
The role of repositories like ReStore.....	13
Proposer vs. funding provider .....	13
Preservation of web resources by search engines .....	13
Chapter 5: Assessment of web resource content quality.....	14
Appropriate content referencing.....	14
Quality of academic content.....	14
Consistency of quality .....	15
Understanding your users .....	15
Appropriate description of artefacts .....	15
Superfluous material .....	15
Content typography.....	16
Frequency of content updates.....	16
Web resource usage statistics .....	16
Promoting the website.....	17
Getting specialist help.....	17
Chapter 6: Intellectual property rights (IPR) .....	18
Disclaimer .....	18
Scope of this guidance.....	18
What are Intellectual Property Rights?.....	18
How does IPR apply to my online resources? .....	19
Experiences of the ReStore team .....	21
Using Links safely .....	22
Managing IPR in a resource.....	23
What sort of things should I do to manage the IPR in the resource?.....	24
What are ReStore’s IPR requirements?.....	24
Some other things to consider when setting up your resource .....	25
Frequently Asked Questions .....	25
Chapter 7: Technical guiding principles for sustainable web resources .....	27
Web site Accessibility .....	27
Usability Standards .....	27
Web Browsers.....	29
Character encoding.....	29
Web Server .....	29
Use of Cascading Styling Sheet (CSS).....	30
Modularisation.....	30
File naming.....	30
Descriptive hyperlink text .....	31
What does a web page contain?.....	31
Search Engine Optimisation .....	32

## Conventions used in this document

These guidelines material are intended to be of use to any creator of online resources, particularly those holding funding from the Economic and Social Research Council (ESRC). The document has been developed as part of the ESRC-funded ReStore project. Two types of rectangular box are used here to:

- (a) Highlight the approach we have adopted for sustaining online resources in the ReStore web repository (Vertical blue header at the left of the box)



- (b) Provide important information largely based on the specifics of the ReStore project (Horizontal blue header at the top of the box)



Any discussion of the type presented in this document inevitably leads to rather specific use of some commonly-used terms. In particular, we would draw attention to the following, which arise out of the ReStore project:

- ReStore specifically refers to the online resource restoration approach which has been followed in the ReStore project. We use the term as a verb meaning to preserve, update and maintain a web resource for a specified period of time, a step beyond merely "preserving" an exact snapshot of a web resource.
- ReStoration is defined as the process of transferring an updated web resource into the ReStore web repository while maintaining active links with the resource author, at every step of the process.
- ReStorable is an adjective which refers to a web resource which could be restored (preserved, updated and maintained) after it has met the necessary criteria set out for ReStoration.

## Online version of guidelines

*A web based version of these guidelines can be accessed at <http://www.restore.ac.uk/guidance>. The document is available both for download and as an interactive version which provides various ways to navigate through the contents of this document.*

# Chapter 1: Introduction

Authors increasingly cite web pages and other digital objects on the Internet, which can “disappear” overnight. In one study published in the journal *Science*, 13% of Internet references in scholarly articles were inactive after only 27 months ([www.webcitation.org](http://www.webcitation.org)).

The online information revolution has provided numerous opportunities to communicate more quickly and easily. The web has evolved into an enormously rich but largely unstructured source of data and information. The growth of the web reflects widening access and developing technologies, but also the increased availability of really useful content. Creating, publishing and accessing content have become commonplace, especially due to social networking, online communities and a series of technologies which have become known as Web 2.0<sup>1</sup>. In such an environment, where most web content is uncontrolled, it has become very difficult to assess and track the highest quality content.

During the last decade, numerous technical standards have emerged for the mark-up and scripting of information on the web. These include HTML, XHTML, CSS, ASP, JSP, Pearl and PHP. The software platforms developed around such languages and technologies are continually evolving to create a complex hybrid web landscape. In such a situation it is increasingly likely that today’s content will disappear or simply be incompatible with tomorrow’s web technologies. Every user is familiar with the frustration of attempting to follow a broken link or opening web content which can no longer be read, played or viewed. Clearly it is possible to continuously maintain every website, but not all content is worth such effort. An important question which therefore arises is how best to discern the quality of web content? In particular, how should we determine what needs to be preserved for future use? Associated with these questions would be issues relating to content ownership and legislation which aim to protect the rights of the creator or owner.

The relative scarcity of funding for research and teaching demands that funders should invest extremely carefully in the creation of online resources. Inevitably, we should consider how best to maximise the impact of funding dedicated to web resource creation and dissemination and to avoid duplicating effort by rigorously promoting high quality online resources. The full realisation of this objective would only be possible if we could be assured that investment in online resources will not be rapidly undermined by changing web standards and technologies.

The remainder of this document is specifically concerned with how to maximise the value of online resources created with ESRC funding, particularly by making explicit the consideration of sustainability from the very outset of each project.

Many ESRC-funded projects include the creation of resources designed to assist other researchers and promote specific research methods. Typically, the development of an online resource is time-consuming and expensive and the full value of the resource only comes into play close to the point at which funding ends. This document is concerned with both:

- Those web resources for which the initial funding is complete and which are at risk of deterioration at exactly the point where they are beginning to be used and valued.
- Resources which are in the early stages of creation and for which there is therefore the greatest opportunity to increase future impact by thinking about sustainability issues now.

## Objective

A wide range of web content, creators, owners and aspects of quality will be considered here. These guidelines, primarily aimed at the creation of sustainable online resources, address the issues which arise at the very start of web resource creation and also those which are faced by authors and content owners at the end of the funding period. It is to be hoped that research council funding will result in many high-quality and useful web resources (although it is neither a necessary nor sufficient condition!). It is therefore strongly desirable to have a strategy in place for the preservation of the web resource in a way which remains accessible and valuable to users before the team moves on to another project or even alternative employment. These guidelines are intended to address exactly this challenge.

Specifically, this document offers to (a) explain, (b) standardise and (c) streamline the process of creating high quality online resources from ESRC-funded projects, with a particular focus on research methods. This is not a technical document and in many areas where there are already excellent published guidelines and standards we will direct the reader to these.

More generally, our aim is to raise awareness amongst web resource project proposers, researchers, authors, editors, contributors and students about the importance of preservation of their valuable web resources before they venture on unplanned publication in the enormous unstructured ocean of the World Wide Web.

---

<sup>1</sup> Web 2.0 refers to a second generation of web development and web design. It is characterised as facilitating communication, information sharing, interoperability, and collaboration on the World Wide Web (Wikipedia).

## Who is it for?

These guidelines for sustainable web resources follow standard web resource creation principles and also draw on practical experience of restoring ESRC-funded online resources in the ReStore web repository project. They include specific recommendations for social science researchers who are about to create a web resource having secured ESRC funding or who are considering options for the preservation of their online resources following the end of the initial project funding.

We would draw the reader's attention to the ESRC Research Data Policy<sup>2</sup> which sets out to support award holders who collect, produce and re-use data, for example by advocating the adoption of a data management plan. See also the ESRC<sup>3</sup> impact toolkit and other resources about the management and sharing of research data<sup>4,5</sup>. Many of the issues applying to data are broadly have parallels in web resource creation and it is our hope that this document will aid award holders in a similar way, although not yet formally adopted as a research council policy. Users of this guidance may therefore include postgraduate students, research assistants/associates, teachers, project investigators, authors and co-authors.

There are issues here which are (or should be) important to anyone who contributes to the creation of online content – if for no other reason than because they are contributing intellectual property, the ownership of which will be an issue for anyone who attempts to work with the resources after they have moved on. Having gone through the relevant sections of the guidance, the reader should be able to:

- Find the information required to plan and create a variety of types of web resource
- Understand the major issues impacting on the sustainability of web resource development
- Understand the basic principles governing intellectual property in web resource creation
- Plan in advance the actions necessary for subsequent restoration of their own web resources
- Add value to their work from the outset by planning for its long-term availability to other researchers

---

2 [http://www.esrc.ac.uk/\\_images/Research\\_Data\\_Policy\\_2010\\_tcm8-4595.pdf](http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf)

3 <http://www.esrc.ac.uk/funding-and-guidance/tools-and-resources/impact-toolkit/tools/interactive-media/website/>

4 <https://dmponline.dcc.ac.uk/>

5 <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

## Chapter 2: Why sustain Web resources?

The web is extremely ephemeral in that 80% of the pages are updated or disappeared after 1 year (Gomes et al., 2011)<sup>6</sup>

In this chapter, we will consider the nature of online resources, the risk of deterioration, and the challenge of actively preserving a resource before it is too late. By sustaining a resource we aim to make sure that it remains available in the same fashion to users beyond the original project funding period. It is in this context that we have chosen to adopt the term “Restoration”, which refers to the preservation, updating and maintenance of a web resource for a specified period of time. This goes beyond merely “preserving” an exact snapshot of a resource and embraces dynamic curation of the site, adapting its content and presentation as necessary to ensure continued utility to users.

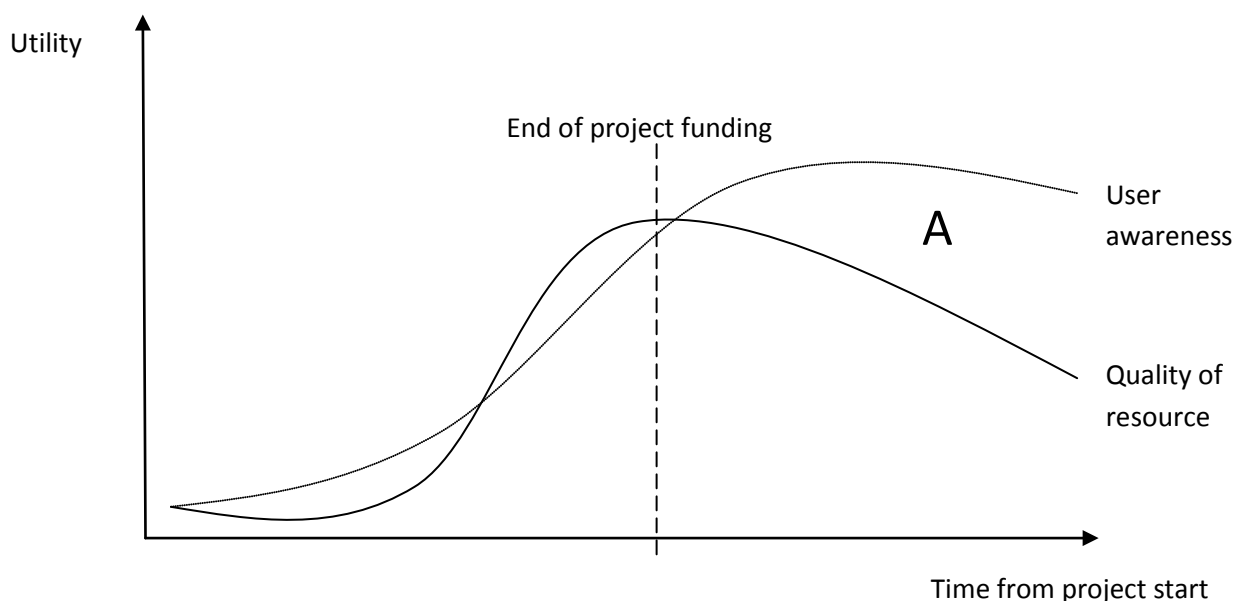
### How and when to sustain?

Before we proceed any further and discuss the “how” part of sustainability, it would be helpful to take as an example a typical research council-funded research project which produces an online resource as part of its activity. ESRC invests heavily in research methods projects which create, as part of their activities, online training and resource materials, often with considerable interactive or reference-value content.

Typically, the development of an on-line resource is time-consuming and expensive and the full value of the resource only comes into play close to the point at which funding ends. The example of an ESRC-funded project which advances a particular research method and produces online tutorial materials for the guidance of other researchers is presented graphically in Figure 1. At the start of the project there is no online presence and user awareness of the research is low.

As the project team present their work at conferences and create an initial website, user awareness increases but the utility of the actual online resources is not realised until the end of the project when the content of the website is complete and the resource is widely publicized. The resources are likely to be extremely useful to a wide range of users within and beyond the academic sector. Indeed, the research councils’ focus on Pathways to Impact<sup>7</sup> should make it even more a priority of UK award holders to ensure the maximum reach and value from their online resources. Resources reach their peak utility at around the time that funding ends but user awareness continues to increase as the materials are cited in other publications and presentations and also spread by word of mouth. Unfortunately, it is during this period of peak usage that, without any follow-on strategy, the online resource begins to decay.

Figure 1



<sup>6</sup> Gomes, D., Miranda, J. & Costa, M. 2011. A Survey on Web Archiving Initiatives. Research and Advanced Technology for Digital Libraries. In: Gradmann, S., Borri, F., Meghini, C. & Schuldt, H. (eds.). Springer Berlin / Heidelberg.

<sup>7</sup> <http://www.rcuk.ac.uk/kei/impacts>

Perhaps a key textbook is published to which it makes no reference, a government department is renamed and a series of important URLs become broken, or a new release of widely-used browser software does not correctly render illustrations within the resources. These deficiencies could all be readily addressed but the researcher has left and the investigators must prioritise other funding applications.

The size of the gap between the user awareness and the quality of the resource, shown as area A in Figure 1, represents the “missed opportunity” for a return on the original investment – in this case missed impact for both the project team and ESRC. Without further action, the site will become unusable and cease to be recommended or used. These issues are not addressed by static web archiving projects which effectively freeze online resources at the peak of their utility, but take no action to moderate the decay of quality over time and do little to increase user awareness.

## What are “Sustainable web resources” all about?

Sustainability generally refers to something having the quality of enduring due to its resilience and robustness. In this sense a web resource is termed sustainable if it stands the test of time and maintains its original utility with the least possible human intervention. The preservation and upgrading of a typical web resource ideally needs to be addressed before its content are outdated, its links become broken or it ceases to be presented appropriately by users’ web browsers. The need to sustain a web resource could be driven, for example, by the fact that it supports an active user base, contains valuable or definitive information, analysis software, research publications etc. The popularity of most resources declines when the resource creator stops updating and maintaining them and their deficiencies multiply and become more apparent to users. The challenges relate to technical, academic and organizational/legal aspects and a resource cannot usually be made sustainable without addressing all of these areas in parallel.

Some online resources continue to be maintained by their original creators or their institutions, especially where they are the work of well-established institutes or research centres that support a broad portfolio of online material within a particular field of research. This is an entirely acceptable solution providing there is some guarantee of ongoing active management. Most of the basic considerations raised in this document will continue to apply to such resources.

The problem of how to sustain online resources beyond their initial period of funding is one that has been long-recognized, for example by the Digital Curation Centre<sup>8</sup>.

8 <http://www.dcc.ac.uk>

An initiative called the UK Web Archive launched by the UK Web Archiving Consortium, has been in the business of preservation and curation of web resources mainly representing subjects of cultural, societal, religious, political and scientific significance to the UK. Common to these initiatives is the fact that they archive web resources in their current form without explicitly addressing the challenges of ongoing maintenance. In some cases, this solution may suffice but it can be problematic in situations where web pages are generated and updated dynamically. Search functions, navigation and external links are particularly prone to failure under this approach. This approach is well-suited to resources which provide reference information that does not change over time, such as metadata or instructions relating to a specific version of a dataset – in this sense it is more akin to the archiving of research datasets.

More frequently however, online resources become unserviceable because content updating has become necessary to ensure continuing utility, or where resources have been created using specific software tools or web standards that are superseded. It may be expected that they will cease to be serviceable over time, even though their core content remains valuable to researchers. Once funding has ended and project teams have dispersed there is little opportunity to update or maintain resources with the consequence that valuable materials are lost to the research community.

It is in response to these challenges that ESRC funded the “Sustaining online resources in research methods” project now entitled ReStore: a sustainable web resources repository<sup>9</sup>. As the name suggests, this type of active restoration comes in to play when active development of the web resource stops. The average lifespan of a web page is 44 -75 days (Internet Archive<sup>10</sup>). The very short lifespan of online resources, if unmaintained, suggests that active restoration and preservation need to begin very soon after the end of funding of a web resource project if maximum utility is to be retained. This requires a clear framework for identifying resources that are appropriate for restoration and the tasks that are required. Many of these are greatly facilitated if they have been anticipated earlier in the project; hence it is most certainly possible to make a web resource more “sustainable” by planning for this phase right from the outset. Questions for consideration include:

- Identification of “what” and “what not” needs sustaining (e.g. methods and principles do, project news items do not)
- Identification of any inappropriate web content (e.g. content for which copyright permission has not been obtained)
- Identification and remedy of technical problems (e.g. broken links, incorrectly rendered images)
- Gathering contact details of team members, contributors
- Readying 3<sup>rd</sup> party consent documents, software licenses and ad hoc contract documents

9 <http://www.restore.ac.uk>

10 <http://www.archive.org/web/web.php>



- Gathering details of technology transfer and relevant institutional involvement (generally the creator's employing institution will be the owner of the material and a technology transfer office or legal advisor will need to be involved)

## ReStore – Wise

ReStore considers that a web resource becomes “restorable” following completion of the initial funding period but while the core content still remains valuable to the wider research community. At this stage there is an opportunity to keep the resource updated and maintained by transferring it into a suitable repository such as ReStore, thus preserving the effectiveness of the resource. ReStore adheres to a policy of restoring the content, appearance and behaviour of a typical web resource. We have developed a set of questions, available on the ReStore website, to help the original author of a resource to assess the readiness of their material for ongoing preservation in this way.

- i. It is recommended that wherever possible, standalone resources should not be restricted to certain users, for example by using federated access management (the system allowing academic access to web resources such as online journals and databases using institutional user accounts) or local username and password systems. Not only does this reduce the user base of the resource but it can create extremely complex maintenance implications, requiring continual security updates, adaptation to reflect broader access control technologies and active management of, and communication with, the registered user base. It may be essential to protect specific content but in this case it is far better to consider placing the resource with an established archive or data centre that is able to support user access controls on an ongoing basis.
- ii. It is generally only practical to guarantee the active maintenance of online resources for limited periods of time – it is recommended that ongoing maintenance be reviewed after no more than three years to ensure that the resource is still being used and is not in need of more major updating or reworking. Factors that need to be considered include: the likely cost of further maintenance; the level of use; whether there is likely to be a change in technology that may either make the resource redundant or would require such a fundamental re-write that it would not be cost-effective. If a well-maintained resource ceases to merit active continuation, it may be an ideal time to consider its transfer into a static web archive.

## Some practical issues: experience from ReStore

Experience with the ESRC ReStore project has highlighted that some online resources require particular investment in sustainability:

- i. Resources that will not ‘work’ unless updated and maintained. Failing to update these resources means that some or all of the investment would be lost. Examples of such resources include web resources that relate to a particular analysis package that routinely changes, e.g. SPSS, STATA, NVIVO. In some cases exemplars will no longer run if users are not working with the same software version as that illustrated in the web resource. In other cases, resources become unhelpful if they are not changed to reflect software updates.
- ii. Resources that will deteriorate in quality and value if not regularly updated. Many resources require links to other pages that, if no longer available, will quickly reduce their reference value. In some cases these resources will continue to be updated by the authors because they have a continuing work programme in that area, but this is unlikely to provide a coherent maintenance strategy.

Other considerations include the control of access and that of for how long a resource should be maintained:

## Chapter 3: Types of web resources

### What is a web resource all about?

There are a very wide variety of sites on the web. We now need a more specific definition of the types of resources to which this guidance relates.

A “web site” is a collection of related web pages, images, videos or other digital assets that are hosted on one web server, usually accessible via the Internet. The web site further defines the web page as a document, typically written in (X)HTML that is almost always accessible via HTTP, a protocol that transfers information from the web server to display in a user’s web browser.

Images (jpg/gif/png), audio (mp3, mp2, wav) and video files (mpeg, dat) etc. are examples of further digital objects which could be added into a web page. HTML also serves as a foundation for other scripting languages (such as ASP, JSP, ASPX, JavaScript etc.) which are used in a web site if it has to respond to user actions automatically for example by running database searches and displaying the results – we refer to these as dynamic web pages to distinguish them from static pages, which do not offer any such response.

The browser software running on the user’s own device (e.g., Internet Explorer, Mozilla Firefox, Google Chrome, Apple Safari, etc.) translates the coded information on the web site into displayable typographic and image elements for viewing. An important aspect is the ability to click on defined elements that then automatically display other web sites (“hyperlinks”). This is an important area when it comes to web preservation and ongoing maintenance. There is no formal distinction between different types of web site but we here characterise some of the most common combinations.

### Simple project sites

These contain a few static web pages and typically list administrative details about a project, in much the same way as would be used in a grant application form or final report. They set out the membership of the team, investigators, researchers etc., contact details, aims and objectives, perhaps a few presentations or working papers, useful links. Much of this detail is also required to be reported to the funding body e.g. via the Research Councils UK web outcomes system<sup>11</sup>. Although useful during the lifetime of the project, simple project sites contain limited academic content and the key information is likely to be retained elsewhere on the web. The direct benefits to end users of archiving such sites are limited. They would normally fail to meet the criteria set out for the ReStore repository.

11 <http://www.rcuk.ac.uk/research/researchoutcomes/pages/ROS.aspx>

### Reference sites

These may contain all the information from a simple project site but also contain a library of important reference material - perhaps datasets or publications resulting from the project activities. These materials can be of considerable academic value, although unless the project has had a very high-profile, they are unlikely to provide the most visible location. Funding bodies are generally likely to require the deposition of important datasets in some form of archive (e.g. the Economic and Social Data Service<sup>12</sup> (ESDS) or UK Data Archive Store<sup>13</sup>). These major repositories are effectively very large reference sites. Similarly, there is increasing interest from universities in the collation of all their staff’s publications into institutional repositories and Research Councils UK require the upload of all publications and outputs arising from the projects they fund. In addition to publishers’ own websites and library catalogues, publication details are becoming increasingly searchable and retrievable by multiple routes and there is a growing trend towards pay-to-publish models for Open Access content. A final consideration is that the original author will rarely have the rights to put copyrighted material on their own website for general distribution, hence it may not currently be possible to share the most prestigious project publications in this way. Although useful for work-in-progress, project-specific sites rarely provide the best long-term containers for important research materials, which are better placed into well- recognised, maintained archives and repositories.

### Resource sites

These sites generally result from projects with a methodological development/training and capacity building focus and at least a part of their aim is to provide training in research methods or techniques. Such sites may be assembled either as a series of static pages or with some level of interactivity facilitated by use of a scripting language and/or database. By their nature, these types of resource may contain a variety of content including relevant publications, sample data, quizzes, presentations, etc. and use a variety of media types. These sites are not readily placed within publications repositories or data archives, nor are they easily assimilated into research project catalogues. Their utility is greatest when they are readily accessible and well-maintained as users seeking to learn new methods will be easily put off by broken functionality. This latter category is the particular concern of the ReStore project - and of these guidance notes. ReStore aims to provide a well-recognised, maintained archive for precisely this type of online material. It is also worth noting that for the purpose of opening up teaching materials for re-use authors may wish to consider placing their materials within an Open Access educational repository<sup>14</sup>.

12 <http://ukdataservice.ac.uk/>

13 <http://www.data-archive.ac.uk>

14 <http://www.jiscinfonet.ac.uk/infokits/repositories/drivers/oer>

## Content management systems and virtual learning environments

This category of web resource may encompass content of any of the types listed above, but is distinctive because of the way in which the content is managed and delivered on the web. Staff in universities may be familiar with two major examples without realising it.

Firstly, it is increasingly common for corporate websites to be managed through a Content Management System (CMS). This software allows individual users to update specific pages (e.g. individual staff research interests) while other material is authored centrally (e.g. academic regulations) and yet other information is extracted automatically from databases (e.g. academic staff publication details, telephone numbers, news). All this is set within a series of overarching page style sheets that allow a central web team to manage the overall 'look and feel' of the site, ensuring that common elements and styles apply regardless of who has generated the detailed content of each page. Essentially, the CMS is a database of web pages, which are served up to the web browser according to a set of standard conventions. A single page on the university's website will be dynamically generated when a visitor clicks on a link and may comprise an amalgam of individually and centrally-authored content together with information extracted from one or more databases. Advantages are that individual users can update pages quickly and easily by simply filling in a web form without needing any web programming skills, while the corporate style is maintained strongly across a wide range of contributors. Examples may be proprietary (e.g. Teamsite<sup>15</sup>), Open Source (e.g. Drupal<sup>16</sup>) or local custom solutions.

Secondly, it has become standard practice for course materials to be delivered to students through Virtual Learning Environments (VLEs). A VLE can be thought of as a specialised CMS which is geared towards the delivery of learning content to students. An overall institutional style is maintained, which may include centrally-maintained links to learning resources, library catalogues etc., while an individual tutor is able to upload and manage materials within their own course such as lecture slides and handouts. A variety of online learning tools are often provided for the tutor such as resources for building quizzes, maintaining discussion lists, running course blogs, handling coursework submissions and making announcements often combined with some form of usage statistics recording. Access to these systems usually requires students to login and hence there is a connection to an institutional database of registered students. Again, examples may be proprietary (e.g. Blackboard<sup>17</sup>), Open Source (e.g. Moodle<sup>18</sup>) or locally specific.

The author of academic web resources intended for use by the broader research community may be recommended to use the CMS or VLE already available within their institutions. While this approach has attractions, there are also specific risks to be considered which include the very limited control over design, institutionally-specific branding, potential difficulty of extracting content from the system, dependence on future corporate IT strategy and the particular challenges of access being restricted to users registered with the host institution. Materials hosted within such systems are generally less readily indexed by search engines and thus the material will be harder to find and less used than might be the case if it were hosted in a more open environment. In many cases, these may make the university CMS or VLE unsuitable as a sustainable platform for research resources. From a sustainability perspective (for example, moving the resource to another server or placing into an archive), it may prove almost impossible to extract pages from such systems while retaining their "look and feel", resulting in a substantial programming burden rebuilding menus and style sheets.

## Personal web sites, blogs, twitter, etc.

Personal web sites, as the name suggests, are created by individuals and generally contain personal information, interests, opinions, beliefs, available on the web. Traditional personal pages are technically no different to the site types listed above, but numerous additional web formats will be found in use, particularly blogs (literally "web logs") in which the author records their thoughts or views sequentially by adding small snippets of information (or in the case of twitter, extremely small snippets) for sequential publication on a web page. Where an academic researcher is well-known and publishes high quality content by any of these means, they may be of significant academic interest but generally share the same weaknesses as the sites noted above, in particular their liability to fall into disuse if that individual should move institution or otherwise cease to maintain the entire project to a high standard.

Blogs are a ubiquitous component of online life, having emerged in recent years as a pervasive, interactive medium for communication and information dissemination<sup>19</sup>.

Such web sites, besides having attributes and components of other web sites, maintain an active connection with a remote or local database for retrieving and displaying content. Parts of such pages are generated dynamically after specific requests have been submitted to it by users such as searching for keywords. An example would be the LSE Impact of Social Sciences blog<sup>20</sup> in which the user can search the database of blog posts in addition to retrieving them by category, date and other criteria.

15 <http://www.interwoven.com/promote/products/teamsite.page>

16 <http://www.drupal.org.uk>

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

18 <http://moodle.org/>

19 [http://www.digitalpreservationeurope.eu/publications/briefs/preservation\\_blogs.pdf](http://www.digitalpreservationeurope.eu/publications/briefs/preservation_blogs.pdf)

20 <http://blogs.lse.ac.uk/impactofsocialsciences>

The ReStore search engine is a typical example of a dynamic web site. We store all the web links, keywords and titles harvested from the web resources on the ReStore site to an MySQL database which improves precision in search results.

## Chapter 4: Alternative initiatives focusing on archiving/preserving/sustaining online resources

In this chapter we will briefly review the merits of alternative initiatives for preservation, encouraging the author to consider whether any of these would be appropriate for their own resources.

### Maintenance by universities

A very viable option for the maintenance of online resources is for this role to be taken on by the original host university. However, it is strongly recommended that there is a clear maintenance strategy- either that the resources will continue to be actively maintained by a relevant department or institute as part of its own web presence or that they become part of an institutional repository that continues to be accessible to external users. These options can be a sound choice when a project has only recently been completed and the original authors still wish to actively edit or manage content. However, there are many instances where leaving resources on their original institutional server leads to rapid loss of functionality or availability. These can result from institutional reorganizations, staff departures, changes to IT services, or content management systems. In some cases, relatively new materials have been entirely lost to the community because of this 'leave things as they are' approach.

There is an increasing move towards the creation of 'institutional repositories<sup>21</sup>' in which institutions seek to maintain the research outputs of the institution, while sharing descriptions of what is available with a range of external directory services. To date, the scope of these repositories has varied with initial efforts usually focusing on research publications and to a lesser extent research data. It is therefore appropriate to consider whether a suitable repository environment exists for the maintenance of web resources and to investigate any particular requirements early in the life of a project if this is considered an attractive option beyond the end of project funding.

Simply leaving web resources within an institutional CMS is generally not a reliable strategy as changes in the CMS software, especially if a proprietary product, may require updating or conversion of content after the project team has disbanded. A prospective resource author would do well to discuss future maintenance options with the IT department in their own institution and discover what options are available for the extraction or transfer of materials hosted by any of the institutional systems on offer.

### Maintenance by a commercial company

In general it is unlikely that a methods related resource would be adopted by a commercial company and made freely available to users on-line. However, there may be some exceptions with respect to software companies where, for example, an on-line resource to provide software training could, conceivably, be adopted and maintained by the software house, and made freely available on the web. Although resources that promote specific commercial software packages tend not to be funded by research councils, there may be some situations where adoption by a commercial company is an appropriate option. However, these situations are likely to arise rarely and will depend heavily on individual contacts and agreements.

### Maintenance by a community of users

There are various instances of academic research projects forming the core for an ongoing online community, which effectively self-maintains a set of online resources. Users who benefit from a specific resource (e.g. they direct their students to it) and also have a direct interest in the content (i.e. it maps onto their own research interests) may be willing to adopt responsibility for its long term maintenance. Building a community of interest could be facilitated by setting up an advisory committee, at the point of project funding, which represents that community. Long-term maintenance could be recognised by providing public recognition of the quality of work being undertaken.

21 [http://muse.jhu.edu/journals/library\\_trends/toc/lib.57.2.html](http://muse.jhu.edu/journals/library_trends/toc/lib.57.2.html)

(This could take a number of forms and is not discussed in detail here.) The “community of users” may be a viable option for some resources. However, it requires a continuing level of commitment from a small number of core people.

The CD-LOR (Community Development of Learning Object Repositories) project investigated these issues in more detail, particularly examining the characteristics of learning communities<sup>22</sup>. It is perhaps significant that the community-building and user engagement activities which appear to be necessary prerequisites for success in this area are not readily developed as part of the type of ESRC-funded research and development projects which create the online resources of interest here. The alternative option of using a wiki as a tool that would enable anyone to contribute or edit material was not seen as providing sufficient quality assurance by those consulted. Successful communities are either very large (e.g. the Linux open source software project) or underpinned by commercial software vendors who incorporate community-generated content (e.g. STATA).

## The role of repositories like ReStore

This section identifies the position of ReStore within the broad range of options discussed here, and stresses in particular the funding model and the importance of ongoing collaboration with the original resource authors. Initiatives like the UK Web archive<sup>23</sup> focus mainly upon preservation and curation of web content by preserving a snapshot of the web resource, and do not actively engage with the creators and/or authors of the web resource. ReStore is different in this regard as it seeks active collaboration with the authors before and after the ReStoration in order to keep the web resource up to date and fully functional. This approach comes at a cost but it fulfils the very basic criteria of web resource preservation in its true spirit. It also helps in maintaining the resource by publishing up to date content with the collaboration of the original creator.

It is also important to note the role of Open Educational Repositories<sup>24</sup> such as Jorum<sup>25</sup>. The principal objective of these initiatives is to permit teachers to freely share educational content with other teachers. Some research methods resources may be suitable for deposition in such a repository, although the emphasis is on making the materials available for others to re-use, as opposed to maintaining the entire operational resource on the web directly for end users. Authors who are interested in the concept of placing their materials in an open repository should give consideration from the very start of the project to the adoption of an open licence such as Creative Commons<sup>26</sup> and need to be sure that this is both suitable for the type of material they are creating and will be acceptable to their institution.

---

22 <http://www.gcu.ac.uk/media/gcalwebv2/academy/content/cdlor/learningcommunitiesreport.pdf>

23 <http://www.webarchive.org.uk/ukwa/>

24 <http://www.jiscinfonet.ac.uk/infokits/repositories/drivers/oer>

25 <http://www.jorum.ac.uk/>

26 <http://creativecommons.org/>

## Proposer vs. funding provider

Current financial pressures are enormously increasing the pressure on funding bodies to consider the utility and impact of awards to investigators. In most cases, a proposer has to explicitly articulate potential impact and to articulate an exit strategy for their award. Research Councils UK defines both academic and non-academic impacts<sup>27</sup>. It is generally recognised that simply being able to disseminate research results does not in itself constitute impact. The creation of web resources has become far easier in today’s world of social networking and web communities, but the question of how to maximise impact and preserve valuable content is much harder.

Apart from the very basic idea of collecting or harvesting web resources, ReStore has gone further towards a strategy which aims to help proposers include the importance of post funding scenarios in order to stress “value” and “impact” from the very beginning of the project. This approach should not only engender a sense of responsibility among investigators, developers and managers but also maximise the chance of creating sustainable resources. If an online resource were created with preservation in mind from the outset, it would save a greater amount time and money that is currently required to restore a web resource after the cessation of initial funds. Sustainability considerations are enormously helpful in maximising impact as they force explicit consideration of target audiences, communication channels and stakeholder engagement and also prompt careful measurement and evaluation of usage. Such an approach is a low cost means to augmenting the impact of the finished web product, thus maximising the return on the investment made by the original funder. This highlights the fundamental importance of bearing preservation in mind when creating web resources.

## Preservation of web resources by search engines

Most of today’s leading search engines e.g. Google, Yahoo, and MSN etc. will index almost all web sites that impose no specific restriction on indexing. This does not necessarily mean that all indexed web pages would be accessible to users after a few months. This type of preservation is normally termed shallow web preservation and does not take into account long term sustained access. In other words, once a web resource decays and a large number of files in it become inaccessible for various reasons (discussed in previous chapters), these search engines stop indexing those particular web pages. These factors reduce the apparent quality of the resource in the eyes of users, thereby reducing usage.

---

27 <http://www.rcuk.ac.uk/kei/impacts/Pages/home.aspx>

## Chapter 5: Assessment of web resource content quality

In this chapter we review the criteria that can help to determine whether or not an academic web resource needs to be sustained and what work will be involved. Clearly, investing in maintenance of a web resource which is of no practical value or significance to researchers would be a waste of time and money. Factors which have to be considered before deciding to actively sustain a resource include the following:

- Does the resource have an active user base?
- Are the contents of the web resource being used and referenced by researchers and students as part of their academic activities?
- Are the contents of the resource of high quality and up to date?
- Have the developers and investigators taken sufficient care to avoid copyright infringement while uploading content, research papers, software tools and datasets?
- Is there an up to date IPR register which records the ownership of content?

Just because there may be problems with some specific aspect of a site (for example, a particular area which has become out of date), this does not mean that it should not be sustained. Rather, the answers to these types of question can help determine whether the benefits of restoring and sustaining the resource outweigh the likely costs.

### ReStore – Wise

To properly assess a web resource for potential inclusion in ReStore, we have developed a review process involving author, academic and technical reviews. Authors of new web resources may find it helpful to consider these review forms during resource development, in much the same way as a journal paper author will want to consider the journal's review criteria and instructions for authors.

The ReStore review process is implemented online, much as in an electronic submissions system for an academic journal. The process begins with a review request to the original author who is encouraged to reflect on the strengths and weaknesses of the resource and details of its construction. A typical author review form can be accessed at <http://www.restore.ac.uk/resources/resourcereviewerAuthor.php?id=JdXAtZ>. Further reviews are sent to external academic reviewers and a technical assessment undertaken directly by the ReStore team. Academic reviews focus on content quality, scope, beneficiaries, usability and potential ongoing value of the site. A typical academic reviewer form is available at <http://www.restore.ac.uk/resources/resourcereviewer.php?id=twpCcf>

### Appropriate content referencing

For academic resources, much the same principles regarding referencing apply as to conventional academic journal and book publication.

Firstly all sources should be appropriately acknowledged and key, up-to-date academic sources should be cited. In addition, it is particularly important to be sure that no third party content has been included in the resource without the permission of its original copyright holders (something we consider in more detail in the section of these guidance notes on Intellectual Property Rights) and also that references to other online materials are properly described and correctly working. One of the greatest sources of frustration to users of online resources is to find that interesting links do not work. It follows that there is a significant maintenance burden associated with ensuring that any web resources are regularly checked for broken links and that these are promptly corrected or replaced with appropriate alternatives.

### Quality of academic content

Again, the academic content in a web resource should strongly reflect expected standards for conventional academic publication in the relevant field. The quality of content on a web page is essentially what determines its potential value to other researchers, although there is great scope to harness the power of the web to present academic arguments in ways which are simply not possible through conventional publication, for example through use of interactivity and combination of a variety of media. A high standard of writing, as would be expected in a refereed journal paper, is equally applicable online. Particular areas requiring care are spelling and grammar, paragraph length and composition, lists (static and drop down), sentence length and contextual relevance etc. It is very easy for authors whose writing medium is not primarily online to become distracted by the technical requirements of web authoring and to inadvertently produce content which is of a lower standard than they would apply to conventional academic outputs.

Adoption of a clear message and perspective across the entire web resource usually helps to establish a good relationship with the user of the materials. Because web resources are not subject to a regular framework of academic review and the associated indicators of quality that come from publication in a recognised book series or highly-cited journal, the web user must rapidly form their own judgement about the quality and reliability of the resource. Although web pages comprise many different elements such as text, images, navigational buttons, menus, etc. it is still the textual composition which is particularly influential. For example, search engines cannot read graphics or parts of a page containing client side scripting code (see the section on standards for further details) and are therefore wholly reliant on the textual content.

Well thought-out keywords and keyword phrases make a big difference in conveying a positive message to users about the authenticity of content and capacity of resource author's knowledge on the subject.

Part of the strategy for producing quality content on a web resource site should be adherence to "quality first and quantity second" rule. Application of such a rule should result in providing users with a satisfying experience and encouraging them to return in the future. A good question to consider is whether the material is of sufficient academic and presentational quality that an interested researcher is likely to want to bookmark the site. Another factor which clearly contributes to the quality of content in a resource site, is to ensure that contents are updated regularly and links to external web resources (URLs) reflect updates in the linked resources (e.g. data releases, software versions, etc).

## Consistency of quality

Content quality must be exhibited consistently throughout a web resource site. It frequently happens that the home page draws much more attention from resource creators and less care is given to the more detailed pages. The primary reason may perhaps be the idea that users always start from a home page but this is not necessarily the case: it is entirely likely that search engines will deliver users directly to a page deep within the site which appears to contain topical content. Thus every page potentially has to function as a first entry point to the entire resource. To maintain consistency of quality:

- Ensure that all content on your site are properly reviewed and tested before they are uploaded.
- Ensure that the global web site design and style is followed on every page.
- Adopt a single standard practice for hyperlinking to internal and external web pages. Maintain the same colour and text style for such links throughout the site.

## Understanding your users

A particular challenge of academic writing for the web is the wide exposure which the author's content may receive. Whereas the tutor of a conventional academic workshop or taught course will usually have considerable understanding (and perhaps even control, via registration prerequisites) of participants' background and prior knowledge, this is not the case with freely available online resources. Users will most often arrive at relevant-looking web pages as the result of entering key phrases into a search engine.

Effective communication in this environment requires the author to think very carefully about their intended audience<sup>28</sup> and to explicitly signpost this in their materials.

---

28 <http://www.esrc.ac.uk/funding-and-guidance/tools-and-resources/impact-toolkit/tools/interactive-media/website/understanding-users.aspx>

Writing for the web does not mean that content is limited to an introductory level, but it does require that the intended audience be clearly indicated.

Sections such as "Who is it for?" can be particularly useful, as can clear links to essential background concepts or terminology and labelling of the material from introductory to advanced. These devices all help the inexperienced and the more expert user each to progress through the materials at an appropriate pace.

Authors should write online content with a clearly defined audience in mind (web designers refer to this as "user-centric design") and this should guide the choice of language and content. A helpful approach can be to present different pages for different audiences. Sections entitled "For students...", "For teachers..." or "For journalists..." can provide faceted introductions to the same academic content for different user groups, and may then provide alternative entry routes to the same material. Adaptive sites may even be set up to allow the user to classify their own levels of expertise and then be presented with appropriately selected content.

Once materials have been drafted, feedback can be invaluable: authors should try to identify members of their target audiences and invite them to test the material before general release. Informal feedback about language, presentation and assumed knowledge can greatly enhance the value and usability of the finished site.

## Appropriate description of artefacts

In web terminology, anything usable by users of a web page (e.g. links, clickable buttons, menus, etc.) are referred to as artefacts. It is essential that such elements are labelled and presented in a way which clearly describes their purpose and encourages the user to make use of them. Thus menu items or buttons whose purpose is not entirely clear are unlikely to be effective in drawing users in to additional material and functions.

## Superfluous material

Authors used to writing academic text may find it tempting to create materials which are over-long for online use. Key considerations should be to identify and remove all superfluous material and to ensure that everything is clearly structured, particularly avoiding long passages of unbroken text. A simple device for reducing superfluous material in the web resource may be to provide a link to a lengthier conventional document which can be downloaded and studied separately by the user who prefers to work in this way. Such a strategy has been followed for this document online. It should be assumed that web users will rapidly move on to another site if they feel that the page they are reading has wandered off the topic in which they are interested. Also, the inclusion of significant superfluous material may reduce the effectiveness of indexing by search engines and thus reduce the overall number of visitors to the site.

## Content typography

Good typography improves the organization and aesthetic appeal of text and other artefacts on a web page, making it more legible. Research users will generally be most interested in the core intellectual content of the site, as conveyed especially through its text. The resource author should therefore give consideration to at least the following elements:

- Language of the content
- Typesetting e.g. creating headings, captions, bold/normal texts, simple and drop down lists etc.
- Hierarchy in content e.g. prioritisation of content to make certain parts appear more prominently and arranging content hierarchical categories
- Font e.g. which font size should be used for a heading, sub heading and normal text in a web page
- Layout e.g. intertwining typesetting with other graphical elements such as images, logos, video files and other illustrations
- Colour e.g. colour of text, main and sub headings, hyperlinks, page and menu backgrounds and foreground colours etc.
- Rhythm e.g. the structural arrangement of the various artefacts in a page and overall web site

## Frequency of content updates

Frequently updating content on a web resource site (e.g. every few weeks or months) will encourage users to return. Updating should take into account activities such as:

- Regular random checks of user interface components such as buttons, home page link, footer links, and links embedded in obvious menus etc.
- Replacing outdated links with updated (internal and external) links. The use of Google Webmaster<sup>29</sup>, a free tool for monitoring all links in your site, will be of tremendous help to carry out these kinds of update.
- Updating links to any internal and external software download pages
- Updating links to publications i.e. replacing old versions with newer ones.
- Reviewing suggestions from users (normally received through an online contact or feedback form) and taking action by updating content or links on the site.
- Redirecting web pages to new pages in cases where the older one has been deleted or is no longer valid (usually for external sites).
- Carrying out multiple browser compatibility tests on a regular basis to ensure that web pages are rendered correctly in the major web browsers. Such testing is necessary because browser software is regularly updated, which in some cases may result in changing the look and feel of web pages.

29 <http://www.google.com/webmasters>

- Updating text, images and other content if identified as necessary by regular review of the site.

## Web resource usage statistics

The collection of web resource usage statistics is key to understanding users' activity on a web site. Such statistics form an important aspect of consideration of whether a resource merits further investment in maintenance, for example through the ReStore project. The following are some commonly used freely available usage statistics tools of the type which we would encourage resource authors to use from the outset of their web resource creation projects:

- Google Analytics<sup>30</sup>
- Google Webmaster & Google Sitemaps<sup>31</sup>
- AWStats<sup>32</sup>

### Google Analytics

Google Analytics will provide insight into user activity and web site traffic on a daily, weekly, monthly and yearly basis which is sufficient for the needs of most academic research resource sites. It is straightforward to install and configure and easy to use, providing a range of activity reports.

### Google Webmaster and Google Sitemaps

Google Webmaster<sup>33</sup> is in many ways an extension to Google Analytics as it can aid in improving web resource site traffic which will be reflected in the analytics report. The tool also helps in recording users' actions on the web site such as keywords searched, listing functional and non-functional links, showing errors in various web pages when they were requested by users, etc. In order to feed this tool with all the URLs in a web site so that it can track activity, Google Sitemap<sup>34</sup> is widely used. Google Sitemap is based on an XML file containing all URLs within a web resource site (including those normally hidden from search engines), using Google Sitemap Protocol. All of the URLs contained in the file are available to search engines when they crawl the web seeking pages to index, which ultimately raises the usage ranking of a web resource site thereby increases its visibility to potential users. Once a Google Sitemap has been manually created, it is submitted to the Google Webmaster tool.

### AWStats

Unlike the Google-based site statistics tools, AWStats requires resource developers to have some technical expertise in web server configuration. It may also require a basic understanding of web programming in some widely used scripting languages such as PHP, Perl etc.

30 <http://www.google.com/analytics>

31 <http://www.google.com/webmasters/>

32 <http://awstats.sourceforge.net/>

33 <http://www.google.com/support/webmasters>

34 <http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=156184>



The ReStore team's top tips to academic authors for web resource quality enhancement are:

- Firstly, ensure that all parts of your resource are of high academic quality, adopting the same standards that you would consider in preparing materials for peer review.
- Use a consistent style, ensuring that your website is clearly presented, easy to read and easy to navigate.
- Frequently update the content, checking internal and external links and actively seeking and incorporating user suggestions.
- Use one of the readily available tools to monitor site usage.
- Keep records, which may be especially helpful to you when writing the end of award report or making the case for further funding!

## Promoting the website

There are many important considerations in good website design, and technical suggestions will be found in Chapter 7 including consideration of search engine optimisation techniques. However, when launching a new website, use should be made of a wide variety of different strategies to advertise the site to potential audiences. This might include direct promotion at relevant conferences and workshops, inclusion of the site URL in print media and promotional materials and particularly placement of links on cognate websites which are likely to be visited by the desired audience. Users can be reminded about the site by telling them immediately if there are changes or news items, especially by advertising these on social media such as Twitter, as well as on subject-relevant blogs and mailing lists. It is also good to encourage website users to share the website and resources to their own social networks. This can be done, for example, by embedding LinkedIn, Facebook, Twitter or Diigo sharing buttons within each page of a website (as on the lower right corner of ReStore pages such as <http://www.restore.ac.uk/> and <http://www.restore.ac.uk/guidance/>). It is not necessary to have a profile on social media platforms in order to be able to include such sharing tools on a website. At the very minimum, research award holders should ensure that their sites are included in funders' and institutional websites and appear under the appropriate categories. Investigators on ESRC-funded projects should make sure that their online materials appear as outputs in the Research Outcomes System and that URLs appear in project abstracts which will then be found by searches on the funders sites', as well via departmental and institutional web pages and repositories.

## Getting specialist help

Academic authors are rarely also the smartest web designers! Authors needing to put their academic content onto the web should therefore give careful consideration to the places from which they may obtain specialist help – and also be aware of the implications. For most authors in UK academic institutions, options are likely to include using IT staff within the institution, employing project staff with web skills or seeking external expertise. Each may provide the best solution in different circumstances.

It will be essential to have the help of specialists within the institution when setting up websites on institutional servers and with regard to matters such as the registration of URLs or project email accounts. Institutional IT staff will have knowledge of locally available software tools, server setup and other relevant opportunities and constraints. IT staff within a local department or faculty may additionally have some subject expertise and greater flexibility to meet authors' requirements, whereas central IT staff may be more constrained in the time and customisation of the service they can offer, and their services may be harder to access. In general terms, when working within an institutional environment authors are more likely to be advised to use institutionally-supported solutions, including content management systems and virtual learning environments which in some circumstances may prove unduly restrictive.

A second option, and one which is widely used, is to employ project staff with web skills who are thus able to build and maintain the project website. This can be very effective, especially when a member of staff has deep knowledge both of the subject content and web design, but investigators should be cautious not to foster the production of unduly non-standard solutions which are only understood by the one member of the team that has worked on them. A common difficulty is that when the key individual leaves for a new post, it becomes clear that no one else knows how to maintain the website or to transfer it to another server or institution. Adherence to widely-used standards and simply writing down clearly the details of how a site has been set up are the best insurance policies in these circumstances. We would advise that passwords and account details for a project accounts should be known by at least two members of a project team.

A third option, frequently employed by award holders, is to outsource web design and content generation to external consultants. Again, this approach has advantages and can be very effective in bringing in skills and experience that are not locally available and which may not justify the expense of a dedicated member of staff. In these circumstances, it will be essential to understand how the external expert is able to interact with local IT staff, or whether the entire resource is to be hosted externally. If great reliance is placed on external knowledge, then investigators should plan ahead as to how maintenance and updates will be made once the project funding that paid for that expert input is no longer available.

## Chapter 6: Intellectual property rights (IPR)

Attention should also be given to the IPR arrangements to ensure that any subcontracting does not hamper the subsequent use and sustainability of the materials being created. Apart from IT aspects of a project, authors are strongly recommended to seek local guidance on IPR expectations and to meet with legal or contracts staff within their organization who can provide expert assistance in ensuring that online materials are properly licensed and carry disclaimers and acknowledgements appropriate to the institutional environment. Again, authors should be aware that institutional advice to restrict access and copyright materials may in fact run counter to other sustainability principles such as promoting open access and sharing of content. Attention should be given to the wide range of excellent online resources and also to the requirements of research funders, for example regarding use of logos, adoption of open standards etc.

### Disclaimer

These guidelines do not constitute legal advice and are offered based on the team's knowledge and experience gained during our work with web resource creators as part of the ReStore project. They are intended to assist online resource authors, particularly those who may want to consider using ReStore in the future, to think through the Intellectual Property Rights (IPR) issues relevant to their project at the outset. We strongly recommend that authors also consult their own legal advisors. Although some of the general principles of good practice could be useful to those setting up web resources for research by other types of organisations and in other countries (as there is has been much harmonisation of intellectual property law internationally) this guidance has been developed predominantly from a UK University perspective.

### Scope of this guidance

It is beyond the scope of this guidance to give an in-depth explanation of all relevant IPR considerations and there is a wealth of existing guidance available, including through the UK Intellectual Property Office website<sup>35</sup>. Rather, this guidance will focus on aspects directly relevant to the creation and maintenance of online resources, particularly those arising from funded academic projects. Our aim is to give a practical understanding of how to work through the process.

### What are Intellectual Property Rights?

The term intellectual property (IP) describes a variety of concepts including copyright, database rights, patents, trademarks and designs. The rights of the owner and/or creator of such intellectual property are known as IPR and generally enable them to prevent third parties from using the IP without their consent.

#### Copyright

The most relevant IPR in the context of web resources is copyright, covered by the Copyright, Design and Patents Act of 1988<sup>36</sup>. Protection is given to literary works (including computer programs), written work including databases, dramatic, musical and artistic works, sound recordings and films, broadcasts, cable programmes and typographical arrangements for published documents. A work will generally qualify for copyright protection if it is one of the types listed above, and it is captured, published on a web site and recorded in some form (e.g. in writing, by a sound/video recording, on a computer or in printed form). Websites, for example, would typically qualify for protection as a literary work.

It is a requirement that the work meets the requisite degree of originality such that it must be new and have been generated by individual effort. Rights exist regardless of whether the author has registered them and the duration of copyright protection will depend on the type of work but with literary works this is 70 years after the death of the author (or last surviving author).

Copyright is usually owned by the author (or joint authors), except where the materials are created during the course of employment, ownership is otherwise specified in a contract or is explicitly assigned to someone else. For most UK academic staff, copyright in the materials they create for the purpose of a research project will be owned by their employing institution although it is worth checking with the relevant institution's policies or regulations in relation to IPR. Authors would normally retain moral rights, for example the right to be identified as the author, to object to derogatory treatment of their work and not to be falsely attributed.

Establishing who is the copyright owner can be important, as it is the copyright owner has, for example, the right to copy, publish or otherwise disseminate the materials and grant licences to others to do any of those things. Copyright is infringed if consent is not obtained. If there are two copyright owners, consent from both would be required. It would also be worth noting that students, consultants or subcontractors are likely to retain ownership of work they create unless otherwise dealt with in the contract engaging them on the project.

It is worth bearing in mind that copyright protects the expression of the idea, not the underlying concept itself.

<sup>35</sup> <http://www.ipso.gov.uk/types.htm>

<sup>36</sup> [http://www.opsi.gov.uk/acts/acts1988/UKpga\\_19880048\\_en\\_1.htm](http://www.opsi.gov.uk/acts/acts1988/UKpga_19880048_en_1.htm)

So to take ReStore as an example: the concept of a website hosting and maintaining research methods resources would not be protected but someone could not just copy the material or content. Using the same selection of project websites, organising them in a similar way and having a similar look and feel would in all likelihood bring it into the realms of copyright infringement. With regard to software it is the actual coding that is protected (and any visual look associated with it) and if it is possible to write different code to deliver the similar functionality or tool then this is unlikely to be copyright infringement (although there may be a breach of the terms of the software licence if software is reverse engineered to find a work around). When authors choose to explicitly publish their works under a Creative Commons licence, this greatly facilitates sharing and re-use but does not absolve the user from requirements to in turn observe the Creative Commons license and to correctly attribute the author and licence.

### **Trademark**

Most web resources will include trademarks. These are brands and/or logos used to distinguish a product, service or organisation, and are covered by the Trademark Act 1994<sup>37</sup>. These can be registered or unregistered rights but it is generally easier to enforce a registered trademark. At registration the context of use of the distinctive branding and/or logo will be defined through the choice of classification under which it is registered. It would be infringement of the trademark to use the brand/logo or even one very similar for the same products or services that protection has been sought for. It would, for example, be an infringement to use Coca Cola (especially with the logo) to describe a sugary fizzy pop/drink. It is also important to get permission to use another's trademark on materials or publications (including websites). This would include permission from the funder and any collaborating partners who may have registered the relevant logo and branding surrounding the use of their name as a trademark.

### **How does IPR apply to my online resources?**

When you create an online resource, you are creating IP for which there will be associated copyrights. Just in the building of the website you are likely to be using third party software, code or tools under licence which you may have paid for or obtained for free (such as open source software or freeware). Understanding the licence terms that apply is important as some are more permissive than others. You may even have developed code yourself, or had a consultant do it for you, if your project needed some specific functionality not readily catered for in existing software scripts. So copyright will exist in all of this.

Then there is the whole design of the site to establish its look, feel and navigation. Copying the design of another site could take you close to copyright infringement.

It may depend on the extent to which what is copied was original, the degree to which you copy it or even the site's web terms and conditions. So using a distinctive colour scheme, your own branding, making the layout, menus and content different should be considered when building a new resource.

All this comes before you start populating the website with words, explanations, tools, papers, presentations, links and any other content. It is highly likely that at this stage you will make use of materials and content created by others and brands of partners in which IPR will exist. It is sensible to make sure that you have the rights to do so and far easier if you get these permissions up front rather than retrospectively.

Below is guidance on how to deal with some of the common types of content online. It is worth bearing in mind that sometimes the pitfalls are not purely related to IPR.

### **Data**

Unless you have been the creator of the data, you should not post third party data online without the owner's consent. Even if you have obtained the data under an open access policy from another site, it is likely it will have been distributed under a specific licence which restricts its onward dissemination or which specifies the form of acknowledgement to be used. You should be particularly careful with personal data and make sure you comply with the Data Protection Act 1988. Also bear in mind that data from projects funded by a third party may be considered confidential information so you need to fully understand whether you have the right to make it available before doing so.

### **Film clips**

Using even a small part of a film can be copyright infringement, particularly if it is an iconic scene. The defence of use for critique or review does require evidence of actually critiquing and reviewing rather than just using the clip as emphasis or illustrating a point. Even then, any use must be fair dealing and it is quite probable that having a clip freely accessible and/or downloadable from a website would take it outside of the realms of fair dealing.

### **Graphs, diagrams and/or charts**

Often these are used to illustrate a point. Commonly people use a graph or chart they have seen in a paper or report that they think is particularly useful and think that as it is a small part of the whole paper it is acceptable to use it. However you need to be cautious of doing so as copyright is likely to exist in the graph, diagram and/or chart in its own right and may be a fundamental piece of the whole work in which it is found. Even where you have generated it yourself you need to be careful where you have included it in a paper or journal article for which you have assigned the rights to a publisher. It is not unusual for publishers to require assignment of all copyright in materials included in the paper (including graphs, diagrams etc).

<sup>37</sup> [http://www.opsi.gov.uk/acts/acts1994/Ukpga\\_19940026\\_en\\_1.htm](http://www.opsi.gov.uk/acts/acts1994/Ukpga_19940026_en_1.htm)

When this is the case, even where the publisher allows you and your institution to use this in other works and make available open access, you won't necessarily have the right to pass on that permission to another organisation such as ReStore. It does really depend on the agreement with the publisher and their respective policies on making available online. Some guidance may be found at Sherpa/RoMEO<sup>38</sup>, otherwise you need to consult your agreement with the publisher and/or get permission.

### **Icons** (used to identify file formats)

The inclusion of the correct icons solely for identifying file formats (e.g. pdf or Word) should not be an infringement of copyright, although care should be taken to use the correct icon and not to make changes to it and guidance should be sought from the software provider's website. To the extent that software may be needed to read the relevant file format then only a link should be provided to enable the user to download this for themselves.

### **Links**

This is a difficult item to address, not least as linking is often given as the sensible way to minimise risk of infringing copyright. However, it is possible that by providing a comprehensive set of links you are in fact assisting others to infringe. The safest course of action is only to link to legitimate sites and not to deep link but to effectively link to the front page and allow the user to navigate to the specific resource themselves (unless the linked site agrees otherwise).

### **Logos or trademarks**

It is really important to get consent before using another's logo or trademark on your site. If you are doing a collaborative project with other institutions or having partners on sub-awards or as subcontractors then it is easy enough to include provision in the collaboration agreement that you will probably need anyway. This provision should last as long as the website is still live rather than just for the duration of the project and allow for the possibility that you will transfer the website to another organisation to keep it going. A partner should not get concerned about the fact that there won't be a time limit to the permission as it can be granted for the sole purpose of identifying their involvement in the project and only for use on the website.

### **Photographs**

Copyright will usually be owned by the photographer but if it was taken of another copyright work (e.g. a painting) then it may infringe the rights of the owner of that copyright work. When taking photographs of people and posting on line you also need to consider whether or not you had the requisite consent from the subjects. Clearance for photographs is needed even if they are included in a bigger paper/article/presentation that is to be included on the site.

### **Presentations**

Even if as a whole the presentation and its words have been created by one person it is common practice to include diagrams and images within them. Where these belong to a third party you should also get permission prior to releasing them online. Remember that whilst it may be fair use to make these presentations that include small parts of third party material in the classroom for educational purposes this is not the same for publishing on a website.

### **Published papers or journal articles**

It is common practice for publishers to require assignment of copyright in academic journal articles, hence these should not usually be reproduced in websites without prior permission from the publisher. Some do allow the authoring academic to make available pre and/or post prints but normally this will be after an embargo period and restricted to the author/their institution. You may find some guidance on publishers' policies at Sherpa/RoMEO<sup>39</sup>, alternatively check with the publisher concerned. It is likely that you won't be able to give another organisation permission to have the paper on their website.

The alternative to seeking permission from the publisher is to link to the place on your own institutional publications repository (many institutions now have these e.g. using software such as EPrints). The increase of Open Access publishing does not remove the need observe copyright restrictions.

### **Reports**

Even when you know you own the copyright you need to be careful making reports from the project, or other projects, available. It is possible the report will contain confidential information of another party and sometimes commissioning organisations (e.g. government departments) insist on having first publication rights. You should make sure you know the terms of any contract and/or grant to ensure that you are compliant with it.

### **Scanned images**

The scanning of an article or newspaper cutting (or any other copyrighted work) would in all likelihood constitute copyright infringement unless the permission of the copyright holder is obtained. It is therefore not advised that these are incorporated. A link to the relevant news article would be preferable.

### **Screen shots**

These have similar considerations as photographs and more than one copyright may be involved. Even when using them simply to demonstrate a third party's software or website it is generally advisable to seek permission before making them available online. Sometimes you may find guidance about this on the software providers' site, for example Use of Microsoft Copyrighted Content<sup>40</sup>.

---

38 <http://www.sherpa.ac.uk/romeo>

39 <http://www.sherpa.ac.uk/romeo>

40 <http://www.microsoft.com/About/Legal/EN/US/IntellectualProperty/Permissions/Default.aspx>

## **Software**

The copyright will be owned by the person(s) that wrote it. It is not recommended to post other people's software on your resource but to link to a place where others can access it legitimately for themselves. Remember it does not necessarily follow that freely distributed software is licensed for unconstrained use or for onward distribution to others.

## **User-posted comments or material**

To the extent that it is original, content from blogs or other comment forums and other materials posted by users would strictly speaking belong to the user who made the post. Their consent would therefore be needed to keep it on the web resource but also to transfer the resource to another site such as the ReStore repository. Ways to handle this are dealt with further below.

## **Videos and/or audio podcasts**

Naturally, copyright will exist in the video/podcast as a whole and is likely to be owned by the person filming/recording, perhaps jointly with the editor. There will also be a mixture of other copyrights such as performers' rights for anyone presenting and/or featured in the material and there could also be copyright in any script that was written to aid its creation. It is quite possible that these rights would be owned by different people and use of the whole needs all their agreement. So it is important to understand who was involved and ensure all relevant consents and permissions are appropriately captured prior to posting on your resource.

## **Written text of the web pages**

This will generally be protected by copyright to the extent that it is original. It could throw up joint ownership issues if more than one person has contributed to the pages.

## **Experiences of the ReStore team**

In some of the examples below we provide examples based on the experiences of the ReStore team as we have worked collaboratively with resource owners to resolve issues prior to depositing resources in ReStore. It is hoped that sharing these will enable you to think about what actions you could usefully be taking now as you start to build your resource on the premise that you want to observe best practice and may need to transfer it to a repository such as ReStore at some point in the future.

## **Issues with user generated content and/or posted materials**

It is increasingly common for websites to include the ability for users to post comments, respond to the information posted on the website or even to post their own papers and resources. When this is enabled then it is important to consider who would own the IPR in anything posted (see User posted comments or material above) and how you will manage the risk of them posting material that infringes third party rights. Unless dealt with appropriately when you allow this, there will be parts of your resource that you don't otherwise have permission to keep available or to transfer to another institution should you so wish.

Getting consent retrospectively can be tedious if not impossible where users have changed email or contact details since posting content.

It is recommended that if you are inclined to allow posting of comments or papers by users then it is best to require them to accept terms and conditions that cover ownership and grant necessary licences including the ability to transfer the whole site to another organisation. You could do this by getting users to register and accept the relevant terms and conditions. Thereafter they can only post when logged in. An alternative is an acceptance required each time they try to post. The advantage of getting them to accept terms and conditions of posting in this way is that you can also require them not to post anything that would infringe another's work and also require them to adhere to principles of posting (e.g. relevant and constructive, no rude or offensive comments and your absolute right to take down any comment for any reason, etc). Moreover, people are very used to having to do this. The BBC web pages are an excellent example and even Facebook and other networking sites on which comments can be posted have such policies. The research support or legal office at your own institution may have some standard terms they would advise you to use. If you have already allowed material to be posted then your best course is to go back and get the necessary permission unless you believe your resource can be transferred without these postings.

There are other challenges with user generated content, particularly whether it does in fact add to the academic value of the resource as you may attract time wasters, spam attacks or even highly offensive comments. This really needs to be considered on a case by case basis and mostly beyond the scope of our current guidance. In any event you should ensure that you plan your resources appropriately to permit moderation of comments posted.

## **Using networking or other collaboration sites as the structure on which to build your resource**

There are many sites that are free to use and are aimed at allowing people to post material, network and share information. Ning Groups, Google groups, Facebook, Flickr are but a few of these in addition to various tools for hosting wikis where multiple contributors can collaborate in creation of shared content. Of course these can appear to be ready made solutions and seemingly an answer to the prayers of a researcher who may not otherwise have anyone in their own institution to help them build such a resource. Caution should be exercised before using these, however. It is worth really understanding the relevant network site's terms and conditions and trying to objectively evaluate whether or not they really are suitable as a tool for project delivery. One aspect to consider is that even though users will sign up to terms and conditions, these will be granting the relevant network site provider the rights they need but will not pass any IPR licences to you or your institution as project convenor. It may be possible to create a semi-open site by having the security settings such that whilst all can view content only a limited few can post or contribute.

Then, as approval for posting and contributing needs to be given, you could grant it conditional on obtaining permission to use contributions in the future, or even to remove these as you please. In any event, as most of these network sites' registration is personal to you as an individual rather than your employing institution, then these permissions will also be personal to you.

So whilst these networking or collaboration sites can be useful it does make it trickier to transfer materials assembled in this way into a repository such as ReStore. Setting aside the substantial IPR issues, remember that even if you can get agreement from the networking site provider that they will download and pass relevant content to you then the work involved in repurposing and building a coherent structure could involve considerable time and effort (i.e. a project in itself). There may also be additional charges by the provider for such non-standard services.

If the provider ceases operation or otherwise chooses to stop providing the site for free, or at all, then this will put at risk the continued availability of the material in your collection. This is something over which you will have no control, not least as their terms and conditions will have given them the absolute freedom to make such decisions.

### **Workshops materials and presentations from events you may have run**

Workshops or conferences often form part of a project and if you organize these you will probably invite some external speakers. It is likely you will want to put any presentations or videos of the proceedings up on your website after the event. Remember if you are doing this to get the relevant consents and licences in place first from external presenters as this is much easier than doing it retrospectively. This can be captured by asking them to confirm their permission for this to be done when inviting them, always ensuring that the permission is transferable to another organisation that may host the site in the future, as in the case of ReStore. Of course if you are paying for the presentation then it could form part of the contract for engaging them.

In a resource we have been recently working on for ReStore, this had unfortunately not been done at the time, so we have developed a simple template<sup>41</sup> that the depositing institution would be able to use to retrospectively obtain the permissions required from contributors.

Also be cautious if you decide to video your event. Obviously, you will need consents from the presenters, but don't forget that anyone asking questions and/or easily visible in the video will also need to give their consent. When running events you can legitimately capture this consent when they register, whether or not you are charging a fee for attendance. Otherwise you may be wise to ensure that any filming simply focuses on the actual presenters. Also ensure that if using an external organisation to do the filming, you get the rights passed over in the contract engaging them.

<sup>41</sup> [http://www.restore.ac.uk/guidance/downloads/Contributor\\_permission\\_form.doc](http://www.restore.ac.uk/guidance/downloads/Contributor_permission_form.doc)

## **Using Links safely**

As mentioned above, the safest course of action is only to link to legitimate sites and not to deep link, but to effectively link to the front page and allow the user to navigate to the specific resource themselves. However this can be frustrating and the user may not get to the specific item to which you intended to refer. You should consider the balance of risk/benefit and needs of the project. Really this is a judgement call. It is always worth checking the relevant site's terms and conditions to see what their policy on linking is. If you are deep linking, then you may want to get specific consent for doing so. Usually these are forthcoming as it is a form of marketing to be linked from sites which potential customers visit. At the very least, it is worth including provision in your own terms and conditions to say in effect that links are provided for ease of reference and users should abide by the linked-to sites terms and conditions and ensure they are not infringing copyright. Of course the other downside of links is that they break and/or become out of date! When this happens your own resource can become less useful. This is one of the specific areas in which ReStore can provide assistance.

### **Including instructions and/or code to enable better use of other people's software or tools**

Research methods resources will often include guidance on how to use other organisations' software or tools. When doing this, it is quite likely the instructions or guidance could include screenshots to help describe what researchers may see at various stages. The ReStore team have encountered this issue in a situation in which the software supplier had not published any guidance. It was considered advisable to seek permission directly and the company were happy to grant this.

These considerations may not stop at screenshots and guidance but the development of sample syntax and lines of computer code that could be used to undertake statistical analysis within particular software packages or if you write and publish instructions on how to do clever stuff with Fred's PhotoShare v2. If this syntax or code can stand alone from the software there should not be a problem, but it is important that you don't grant your users any licence to the other party's software and only link to a legitimate site to enable them to access it for themselves. You should also be certain that in providing such example code you are not in fact infringing the terms of your own licence for the software, which may be designed to protect the rights of the software company to run its own software training materials and courses.

Where anything you write or develop also includes code taken from the someone else's software by way of assisting the user to interface with it, then there may be a risk of infringing copyright by including it, small amount though it may be. If you are intending to do this, it is really recommended you understand the licence under which you are using the other's software and probably obtain express permission to include it in your own code. Bearing in mind that this could improve the functionality of the supplier's software it is highly likely permission will be forthcoming.

A collaboration of researchers from different institutions Projects increasingly involve collaboration across institutions and even where one party will take on the responsibility to host the web resource all are likely to contribute to its generation. One resource we dealt with had been authored in this way and prior to transfer to ReStore it was necessary to establish the nature of the contractual relationship between the two institutions and for the one which held the main award to obtain written permission from the other, who had acted as a sub-contractor, for their content to be transferred as part of the site. The university legal teams were able to resolve this relatively simply, but it had not been foreseen or specifically addressed as a likely consequence when the project had been set up and it was time consuming to resolve. This is one reason why it is recommended to get a collaboration agreement in place tackling these sorts of issues when two or more institutions work together on a project. If you are setting up a new project that will be creating a web resource that may need to be transferred (to ReStore or another service) at a later date, then it could be worth including specific agreement on this to prevent delay or difficulty later on.

### **Using third party software or freeware to build the structure of your web resource**

On the whole, if you use commonly-held or easily obtainable software then this should not pose a problem as even though the licence under which you were using it probably prohibits you from sub-licensing the software it should be relatively easy for a repository such as ReStore to obtain the requisite licence to operate it. Using hard to obtain or expensive software in order to maintain or operate your web resource may be a significant obstacle to its ongoing sustainability, whether hosted by ReStore or your own institution. If it seems that this issue might be relevant to a website that you are considering transferring to ReStore, then please discuss this with us as early as possible. Where appropriate, consideration should be given to the use of Open Source software tools such as Moodle and Drupal, which are inherently more transferrable than software requiring licences and maintenance fees.

### **Managing IPR in a resource**

By now it should be apparent that the creation of an online resource requires some consideration of the IP ownership of all the materials to be included and, if necessary, the seeking of permission from the owners. Think about an appropriate IPR management strategy right from the inception of a web resource project as this will help minimise the risks relating to copyright and other IPR infringement. On a simple level it is all about knowing what you have included in your resource, understanding who has what rights over it and/or any items contained in it and what are the constraints on you making them openly available or transferring to another organisation.

IPR Management is really all about being organised. It needn't be a time consuming task but it is sensible to get it right at the start because trying to resolve these issues can be extremely difficult once a project has finished.

Tracing consultants, past-colleagues and students and getting cooperation from other partners will be much more time consuming retrospectively than when you are all working towards the same project delivery.

You may ask yourself if it is really necessary to do this. While it is relatively easy for an author to create a web resource without regard to IPR issues, irrespective of whether it is in fact a breach of their own institution's regulations, transfer of the resource into an archive or repository will almost certainly bring these issues to light including IPR infringements which future hosting organisations will not want to inherit. Remember liability for infringement will shift to them in the first instance if they do take on your resource.

Even if you are not contemplating transferring your resource to another organisation at the moment do try to think beyond the lifetime of the project. Potentially you, or one of your collaborating researchers, might move institutions; or perhaps you are delivering a service that the funder will want to maintain at the end of the term of your project. Good management will enable you to be flexible and readily move the resource should you want to transfer it or merely to keep it going.

Having a good handle on the IPR in your resource could also make it easier for you to defend spurious demands to remove allegedly infringing materials. It would enable you to quickly identify if you have the material legitimately and potentially to establish that the material was in fact developed independently by yourself and at what date. Even when you know you are not at fault, receiving threatening communications is distressing and being able to revert to a sensible and well maintained record of IPR will make this less frightening.

The consequences of failure to properly manage IPR can include that you end up needing to take down or remove materials. Depending on the amount of material that is considered infringing then this could really be detrimental to the value, substance and potential sustainability of your work. In the worst case, if you are sued and there is validity in their claim of IPR infringement, then if you can demonstrate that you did your best to keep an appropriate record but the offending material must have slipped through the system, this will certainly be viewed favourably by most courts. This can also mean the level of award is not inflated for your wilful disregard and/or deliberate infringement of other people's rights.

Ultimately we believe taking a responsible approach with IPR management will contribute to the quality and legality of your resource and enable you to keep it, or for another organisation to maintain it, long after your project is completed. All this can add to the impact of the work you are planning on doing now and add value to your project outputs as they are used in the future.

## What sort of things should I do to manage the IPR in the resource?

### Put an agreement in place with all contributors

It is now fairly normal for research to be carried out in collaboration between institutions and when this happens it is advisable for a collaboration agreement to be put in place. Indeed, most research councils now make it a condition of award and some even insist on seeing the signed agreement before releasing funding. Talk this through with your research support office (or equivalent) and explain that you will be creating a web resource as it could be very sensible to have specific agreement on what can happen to the resource after the end of the project.

You should also make sure you have a contract in place with any consultant or subcontractor you engage on the project specifying what will happen to the IPR they generate as under law you will not automatically get rights to it. You can decide whether you think you need assignment of the copyright or whether a licence is sufficient but if you only get a licence, make sure you can sub-licence otherwise you won't be able to transfer the site to someone else to host.

### Keep an IPR register / record

Managing IPR does not need to be complicated. Authors building an online resource should maintain a list of elements which make up their resource and identify the owners, particularly considering whether they are employees of an institution, students or consultants. Owners' permission should be sought for any elements not directly created by the authors and records kept of any correspondence, including copies of standard licences, for example relating to use of software employed in creating the site. Particular care should be taken to understand the terms of any 'open access' or similar licences to ensure that your intended use is covered. Where possible, it may be preferable to link to an external resource rather than to incorporate it within your own material.

In our opinion there are two main elements to keeping a sensible IPR record:

1. A summary document, database or spreadsheet that can be used as an easy reference tool; and
2. A file/document store where supporting documentation can be deposited in case more detailed investigation is required.

The spreadsheet/data base should capture a full description of the material concerned and identify what other material is available. At the University of Southampton we have developed a simple spreadsheet<sup>42</sup> which can be used as a template. What system you use to keep this record is entirely up to you and/or your institution. You may have the capabilities to integrate this into the structure/underlying architecture of your web resource and it is quite possible that your institution has developed their own tool, database or other system you can use.

42 <http://www.restore.ac.uk/guidance/downloads/IP%20Register%20Form%20Web%20Resource%20record.xls>

The document/file store should ideally have some form of version control or document management system as a way to prove the document has not been tampered with subsequently. If this is not readily available to you then at the very minimum try ensuring you save as pdf documents, dated and locked to read-only. The alternative of course is keeping original hard copy documents but if you do have these it is always worth ensuring you have electronic copies for back up purposes. Items to ensure are contained in the file/document store are any collaboration agreement, institutional IPR policies/regulations, contracts with external providers, licences to software, consent and/or permissions given from contributors, etc. Where the relevant terms and conditions you rely on are located on a third party's website then it is worth taking a pdf or printout of the relevant page and time stamping it as evidence of the terms and conditions that applied at the time you accessed the material. These too should be held in the document store.

If running a collaborative project then it is worth ensuring the summary document is regularly discussed at project meetings, agreed and signed off so this could usefully be a standing agenda item. All partners should be required to ensure relevant documentation is contained in the document /file store. We recognise that tools and practices for IPR Management will continue to develop so if you think you have found the answer please do share it with us. In the meantime just think how much easier it will be to answer the due diligence questions (see below) that a repository such as ReStore requires if you have all the information you need at your fingertips!

## What are ReStore's IPR requirements?

It is entirely up to you and your institution as to the risk you and they are prepared to take with copyright or other IPR infringement. By transferring your web resource into a repository such as ReStore the hosting institution would be vulnerable for law suits arising from any copyright infringement being levied at them as they would now be hosting and making available the infringing material. ReStore is a free service to depositors wishing to keep their work sustained. All we are asking for is a due diligence check prior to transfer to help us understand the risks concerned so we can effectively manage them. We want to work collaboratively with you to try to identify whether or not there is something that could put us at risk of copyright infringement. This is recorded in the form of a completed standard questionnaire addressed to the principal author<sup>43</sup>. We also ask the depositing institution(s), on the premise they are also the copyright owner(s), to confirm that everything we have been told is true and you and they don't have any other reason to believe that what is transferring would infringe another's IPR. By going through the due diligence at the review stage it will make it easier for your institution to sign up to such a provision in the deposit agreement.

43 [http://www.restore.ac.uk/guidance/Author\\_Reviewer\\_Form.doc](http://www.restore.ac.uk/guidance/Author_Reviewer_Form.doc)



The alternative to undertaking due diligence would be to ask for a full warranty that the transferring website does not infringe another's IPR and an indemnity in the event there is a claim against us. However from our experience negotiating with other institutions we have found provisions such as this even more problematic. Most institutions simply will not give that type of warranty and indemnity.

All this is enshrined in a formal licence agreement with the hosting organisation for ReStore – currently the University of Southampton. This is a formal legal document and it is extremely rare that an individual researcher would have the authority to sign such an agreement on behalf of their institution. Normally this needs to be done by or at the very least approved by the depositing University's research office, intellectual property office or legal department. The licence seeks to record the IPR status of the resource, including the members of the authorship team and any IPR policies under which it was created; the efforts made to seek permission to include any third party materials, for example software used or photographs reproduced, etc.

Essentially, the licence records the efforts undertaken to observe the principles of the guidance provided here and aims to reassure the new hosting institution that appropriate care has been taken.

## **Some other things to consider when setting up your resource**

### **Have a takedown policy**

A take-down policy is an important means of mitigating the risk of IPR infringement claims by providing a mechanism for immediately removing any material which transpires to infringe someone else's rights. It should be made clear to a user of the site what they should do if they believe that their rights have been infringed by some material on the site – it should specify who to contact, how to identify the material in question and what action will be taken. During the period of any investigation, it can sometimes be advisable to remove material (even temporarily) from visibility on the website, although this may depend on the seriousness of the complaint and/or how quickly any investigation can take place and be decided upon. ReStore's own take-down policy<sup>44</sup> is available online. In the course of our work we have encountered a situation in which an external organization identified that their material had been inadvertently included in a web resource and required this material to be taken down. In such circumstances a clear take-down policy and the means to carry it out can be invaluable. This type of situation could be particularly problematic to an investigator whose project had ended and who had no readily available technical support when such a claim arose.

44 <http://www.restore.ac.uk/takedownpolicy.php>

### **Website Terms and conditions**

We similarly recommend that your web resource have clearly stated terms and conditions, reflecting the matters discussed in these guidance notes. You should discuss with your own institution whether it has model terms and conditions which should be used, especially as for university-based projects it will usually be the institution that is ultimately responsible for material which you publish to the web as part of your research project. ReStore's own terms and conditions<sup>45</sup> are available online.

## **Frequently Asked Questions**

### **I am doing research. Doesn't the research exemption apply?**

The so-called research exemption (section 29 of the Copyright, Design and Patents Act 1988) only applies to non-commercial research and private study and even then it must be within the fair use provisions. This allows an individual to copy for their own research purposes but not to copy to enable others to use for research and private study unless it is within the provisions protecting a library.

Indeed it is worth noting section 29 (3) (b) which provides that it cannot be fair dealing where the person doing the copying 'knows or has reason to believe that it will result in copies of substantially the same material being provided to more than one person at substantially the same time and for substantially the same purpose'. Therefore posting on a website would fall outside this and is more publishing, distributing and disseminating than use for your own research purposes. Further, it is hardly fair use where there is open access to the website that hosts it.

This is before we have even considered the difficult question of what constitutes 'not for commercial purposes'. Just because an educational institution may be non profit making, this does not mean its actions will never be considered commercial, particularly where it could be seen as a marketing vehicle for future fee-paying students or to raise profile to increase the chance of winning more research grants. In short, it is unlikely that the research exemption will provide much protection to those compiling web resources.

### **My website is for educational purposes. Surely that is fair use?**

There is an exemption under the act for copying done for the purpose of instruction and examination (section 32 CPDA 1988) but it must be done by the person delivering the instruction, be for non-commercial purposes, not done using a reprographic process (e.g. photocopying and scanning to electronic format etc) and due acknowledgement be given. However this exemption applies to 'copying', whereas posting on a website is publishing, distributing and broadly disseminating. Therefore it does not in our opinion assist in the uses anticipated by this guide.

45 <http://www.restore.ac.uk/terms.php>

**Surely by posting something openly on their website they have given me the absolute right to use it?**

This is a commonly-held view but is not the correct position at law. By making something available, the copyright owner is simply allowing individual visitors to have access to it without charge. In the absence of any other licence (Creative Commons or otherwise) then all other rights of copyright would be preserved. This means that it would be an infringement to copy, publish, show, make publicly available and or make an adaptation of the work. Therefore even if you clearly identify the copyright owner you would be infringing by a) copying electronically and b) making available on your website without consent. It is worth checking on the site's terms and conditions as it is possible there are greater rights afforded and it is possible they have released on a Creative Commons licence. Of course it is always possible that the copyright holder does want their work to be broadly disseminated but simply didn't understand copyright themselves. Where you think this is probably the case then it is easy enough to ask for the permission you want but it is unwise to presume it to be the case without asking.

**My project website has been up for years and we have never been sued. What is all the fuss about?**

The fact that no-one has sued doesn't mean there is no copyright infringement. Copyright will exist for more than 70 years in most cases (likely to be more than 100 and considerably longer if the author lives to a ripe old age!) so litigation can be instigated at any time. The fact that no-one has sued could be because a) it has not been spotted yet, b) the copyright owner doesn't see this as a threat at the moment, c) they may be concentrating on other more serious infringers and will come to you later, or d) they may not have an issue with you doing it but could have an issue with your materials if placed in a repository site (where it may receive more visitors). This is not an exhaustive list.

**Do the funders of my work have any requirements?**

It is clearly important to ensure that the requirements of any funders of a web resource creation project are observed. Currently, ESRC does not attempt to exert rights in resources which are created as part of its funded projects, but it will expect appropriate acknowledgement in accordance with the guidance issued to award holders. In most cases, the copyright in such materials will be owned by the institution which employed the staff undertaking the work. Other funders may have specific requirements, particularly if the work is undertaken as part of a consultancy contract or other commissioned research whether with a company or government department – in these cases the contract should make it clear who will own the IPR in the resulting site and who is therefore able to maintain and preserve it into the future. Sometimes funders put blanket IPR terms in place and it is worth talking with the research support or IPR department in your institution to determine whether the contract is suitable for your planned project website: it may be possible to negotiate the inclusion of more suitable terms on a project-specific basis.

**It is not possible to identify who the author is or who owns the copyright. Can I use it how I please?**

These are sometimes described as 'orphan works' and just because the author or copyright owner is not readily identifiable does not mean they do not exist. However, in practice it will make it very difficult to obtain the consent necessary to use the work if you don't know who they are. It is for this reason that you should take care to ensure that where you do copy something or change the format of a work electronically (either your own or another person's with their consent) then you should take care to make sure that copyright owner is clearly identified. It is not recommended you use a copyrighted work that does not appear to have an owner, largely because there will be one out there so alternative material should be used where possible. It is worth doing some due diligence to see if they can be identified either by looking at the site where the work originally came from or searching for similar material elsewhere. In the event there really is no alternative then you are taking a risk that they will emerge and sue for infringement.

## Chapter 7: Technical guiding principles for sustainable web resources

The sustainability of web resources is greatly enhanced if they are well structured and technically stable. Not only does it make these sites easier to maintain in the long term, but these are of course also characteristics that will endear them to users.

The purpose of this guidance is not to replicate the many relevant standards that have been published and are available elsewhere, but to highlight principles that should be considered by the author of any research methods resource. Many of the themes covered here have been identified through our experiences working with creators and users of ESRC-funded research methods resources and we particularly emphasize issues relevant to the authors of such resources. Although many of these principles may seem obvious to readers familiar with using the web, it is surprisingly easy for creators of academic sites to overlook some of the basic standards of web design, especially when this has to be combined with authoring the academic content which is naturally the main focus of attention. Getting this right may appear to be a distraction from an academic project, but getting it wrong may well cause users to leave the site before they reach the academic content! It will be seen that paying due regard to web standards can in many cases improve the suitability of resources for long-term preservation and improve their indexing and ranking by search engines.

Our starting assumption is that the reader already has a good idea of what the academic content of their site will be and understands that web pages are assembled from instructions written in hypertext markup language (HTML)<sup>46</sup> which can be created using a wide variety of software editors.

While this chapter guides the reader through various technical principles relevant to building their own web pages, there are particular issues associated with the use of content management systems or virtual learning environments (explained in our section on types of web resources). The resource author using these tools will generally find that they have very much less control over the technical issues considered here and must live with various important decisions that have already been set by their institution. This can pose enormous difficulties if it is desired to extract content from the system at a later stage because although it may be possible to extract the page content in some way, it may not be possible to preserve the styles, menus, layout and other elements which are actually part of the system rather than the individual pages. Any author considering the use of these systems should find out exactly what options are available for exporting and archiving. If there are good reasons to use such a system, then much greater flexibility and control will be afforded by using open source software such as Drupal or Moodle, where the entire environment can be moved to another service provider without undue licensing and technical restrictions.

We would generally recommend resource authors to build stand-alone pages unless there are compelling reasons for using a CMS or VLE and to use open source solutions in preference to proprietary ones. If an institutional CMS or VLE must be used, an up to date external copy of all the content should be retained with minimal formatting. Some specific comments on these issues are given in the individual guidance sections.

### Web site Accessibility

*"The power of the Web is in its universality".*

Tim Berners-Lee, W3C Director and inventor of the World Wide Web

The term accessibility has more than one meaning in relation to web resource design and development. In general, the term refers to reliable fault-free access to web content regardless of the software, hardware, operating system, etc. employed by the end user. In addition, a fully accessible web resource will be able to deliver its content to users with disabilities in order that they are able to benefit equally from use of the web.

### Usability Standards

In simple terms, the usability of a web site reflects how easily and efficiently a user can navigate through the constituent pages. Achieving good usability standards requires creators to come up with content that is engaging, appropriate and relevant to the primary users of the site. Thus it should be clear that usability standards are not merely about navigation, visual design, functionality or interactivity. Nevertheless, a good user interface does play a central role in enhancing the usability of content, thereby increasing value and creating impact.

### Consistency in design

The human mind constantly searches for patterns and this is especially the case when exploring a web site. If no consistent pattern is apparent, the user is soon likely to look elsewhere. Consistency in design involves much more than presentation style, extending to organization of content and the entire experience of user interaction. Design of a single web page or entire site should always keep in view the intended primary users. Once finalised, the design should be replicated across the entire site, usually using a single Cascading Style Sheet (CSS) file – described more fully below – which controls the appearance of every page and avoids users becoming distracted by variations in style.

A well-crafted style should still be flexible enough to accommodate necessary changes such as the introduction of a new link or button on every web page on the site.

46 <http://www.w3schools.com/html/>

If the design is such that a developer or author has to go to every single page to insert the code for a new link or button, this will be highly inefficient and will often lead to errors and inconsistencies. On the contrary, if the style and layout files are kept separate from the page contents, such changes are relatively quick and simple.

### User-driven navigation

In any web site the structure and organization of content plays a pivotal role in guiding users around the site. The structure encompasses the formation and placement of link vocabulary (link titles, names, phrases, etc); availability of core links on every page to facilitate easy navigation; visibility of each navigational entity in individual pages and across the entire site, and flexibility to accommodate future changes. This is by no means an exhaustive list but indicates factors which deserve careful attention at the design stage. Several measures may be adopted to enhance navigation such as construction of a well-structured main menu (typically including home, contact details, an accessibility statement, terms and conditions, etc) and a detailed sitemap, which are made available from all the core pages. Such devices help users to understand what content is available and where to find it. An omnipresent main menu can give users an immediate impression of the coverage and depth of a site, helping them to see what is on offer and encouraging them to explore further. Some of basic navigational elements are:

- Linking back to the home page from every page on a site
- Displaying a “breadcrumb” trail which always clearly shows the user where they are within the site structure (e.g. “Home > Contact us > Complaints department”).
- Placing page jumps which link the various sections of a single page, aiding exploration of more lengthy pages
- Having an omnipresent facility to search across the entire site

### Memorable layout

Alongside making a web site usable, making the way it works memorable and consistent are other important characteristics. Ideally, a user should be able to find their way around a site and memorise its navigational style in one, or very few, visits. If they have to keep relearning the function or location of links, buttons, menu items, or widely varying page layouts, they will most likely not return. This will decrease the impact of the original development and its academic content. Working with an inconsistent site design is analogous to cooking in a kitchen in which there is no governing logic for where utensils are kept, and continually having to learn, for example, that forks are not in the same drawer as knives, or essential ingredients are scattered across a variety of cupboards!

### Memorable URLs

A Universal Resource Locator (URL) in the form `http://sitename.ac.uk/filename.html` comprises three parts, namely the protocol (`http`), domain name (`sitename.com`) and file name (`filename.html`).

URLs which are simple, short and meaningful can be easily memorised and shared among a community of users, and may make the difference between one user telling another about the site or not. The URL of a web resource should not be longer than 78 characters to avoid being broken into multiple lines inside an html editor, email message or browser.

Shorter URLs are easier to spell and people often type them directly into their browser rather than accessing them from bookmarks or searches. Authors should also make use of relative URLs when linking within their resources (i.e. filenames expressed relative to the structure of the folders containing the site, not absolute URLs starting with the site name) which will therefore continue to function if the entire resource is moved to a different server.

A web resource creator should take care to choose a short, meaningful domain name and file naming convention. If registering a new domain name, consideration should be given to finding a name that will ideally tell users as much as possible about the site and its content. It should be descriptive, meaningful and free from jargon or special characters (e.g. \* \_£#@?>< etc.) Using spaces between the characters in a domain name is also strongly discouraged. Some basic guidance on naming files is provided later. UK academic users will find that sites hosted on their institutional servers will typically take names that begin with their university domain and possibly a departmental element e.g. `http://www.myuniversity.ac.uk/mydepartment/myproject.html`. These can become hard to find or remember if the number of levels is too great. Major initiatives which are expected to span many years may be registered to their own non-institutional domain names such as `http://www.restore.ac.uk/` but this involves an application process and case to be submitted. Resource creators considering this option should consult the JANET website<sup>47</sup> and their own university IT service for guidance.

### CMS and VLE considerations regarding accessibility

These will generally provide very limited options for the author to control the style and appearance of web pages, as they will generally have been set up to ensure that everything conforms to an institutional style. The advantage to the author is that many of the elements considered here such as consistency in design and navigation are already assured by. However, there will be little choice over page layout, style or names. Pages (being generated dynamically) may not have consistent identifiable URLs and are therefore less likely to be exposed to search engines or remembered and shared by users. Links between pages within these systems are unlikely to function correctly if the content is exported. Elements such as images or video clips are often stored in a database and will not necessarily be extracted as part of the pages in which they were intended to appear. All these obstacles can make it very difficult to extract a resource from such systems as a functional whole, requiring substantial additional web programming.

---

<sup>47</sup> <http://www.ja.net/>

User registration may seem attractive in terms of gathering usage statistics and controlling use within a research community or institution, but generally runs counter to the principle of open access, restricts findability by search engines and will in fact limit usage.

## Web Browsers

A web browser is simply a software application used for viewing web pages on the Internet. The most commonly used web browsers are currently Internet Explorer, Mozilla Firefox, Opera, Google chrome, Flock, Safari, etc. In order to be confident that a web page will be displayed correctly on a user's web browser, browser compatibility tests must be carried out before uploading the page to the web site. We strongly recommend that web resource creators test every single web page in multiple browsers as and when pages are created and uploaded into their sites. All of the above browsers are freely available for download and can be installed with minimum features to conserve disk space. Resource authors should bear in mind that the leading academic in their field, or indeed those charged with reviewing the outcomes of their project, may be dedicated enthusiasts of a different web browser to themselves!

A typical web page consists of many elements e.g. text, images, audio and video files, style sheet, JavaScript code snippets, etc. which the user's browser interprets and displays as an integrated whole. However, in some cases, a browser may not be capable of interpreting one or more of these elements. For example, it may fail to interpret JavaScript code because the user has not enabled JavaScript reading in their browser setup. If this is the case, the browser will simply skip that part of the page and would render the rest of the page with a message indicating potential errors. It is therefore strongly recommended that developers using JavaScript come up with parallel server side scripting, which runs on the server rather than the browser and therefore substantially reduces the probability of serving an incomplete page.

## Character encoding

Information within computers is stored and transmitted as bits (binary digits), generally grouped as bytes, which must be converted to characters before being readable by the user. This applies equally to the content of web pages: it is a series of bits that are sent to the user's browser and not the characters which we expect to see on the screen. These are assembled into recognisable characters following a process called Character Encoding (CE), which is the set of rules telling the browser how to convert the bits and bytes to characters. There is more than one such standard, such as ISO-8859-1 which is usable for most West European languages. UTF-8 is another CE standard using different number of bytes for different characters. Detailed explanations are not necessary here, and the user will find more details on the [W3C web site](http://www.w3.org)<sup>48</sup>.

48 <http://www.w3.org/TR/html4/charset.html>

## Which encoding to use and where?

CE is either automatically set in a web server configuration file or the web resource creator can insert a CE directive in the header section of their HTTP web pages (header sections are explained further below).

Regardless of any CE instructions sent by the server, the browser will render the page on the user's computer using the specified CE and it is always a good practice to add the directive regardless of whether or not the web server has sent a CE directive. The line required for manually adding CE into the <head> section of a web page is:

```
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
```

OR

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8">
```

Based on our experience of ReStoring various web resources, we would recommend the use of ISO-8859-1 for all of web pages unless there are specific reasons for doing otherwise.

## Web Server

A web server is the hardware and software system responsible for processing user's requests submitted from a web browser: it is this system that acts as the container for the web site. Many academic site developers will have little awareness of the setup and configuration of the web server they are using, because it will be maintained and operated by dedicated IT staff within their institution and they may have to do little more than place their web pages in a specified folder. However, there are aspects of the server which are worthy of consideration here and there may even be instances where the requirements of a specific project merit the setup of a dedicated server.

A server delivers web pages (usually developed in HTML) and associated content (e.g. images, style sheets, JavaScript) to clients. If a web page contains elements such as programming script e.g. Java Server Pages, ASP (Active Server Pages), VB Script, Perl, etc, then these are processed by the server before being displayed to users in their respective web browsers. Such a page is generally termed a dynamic web page. An online HTML form where a user enters their contact details is a typical example of a dynamic web page. This may result in both the display of a welcome message in their browser and the storage of their records on the server. These types of pages cannot be created without some understanding of the server. No detailed consideration is given to these systems here, but resource authors should be aware that institutions will most likely have specific policies about the types of web server (e.g. Apache Web Server, IIS (Internet Information Services), Apache Tomcat) they are able or willing to support.

There are numerous licensing, security, maintenance and support implications of server decisions and resource authors are therefore recommended to discuss their requirements with their local IT service provider at an early stage in order to understand the options available to them and whether there will be additional costs involved.

## Use of Cascading Styling Sheet (CSS)

CSS is a simple mechanism for adding style (e.g. fonts, colours, spacing, captions, titles and special effects) to web documents. Different CSS style elements may be applied to an HTML document in order to make a web page appear in a user's browser in the way the developer intends. However, some CSS style properties are interpreted differently by the different browsers. This is one of the reasons why we recommend testing all web pages during development in all the major web browsers after any CSS has been applied. It is also strongly recommended to use a single CSS file to control the look, feel and style of the entire web site. By using a single CSS file, any style-related anomalies in a page or across the whole site can be fixed after making modifications in just one place. Much more extensive details are provided about CSS on the W3C website<sup>49</sup> which offers up to date tutorials and other learning materials for developers from beginner to expert.

### CSS vs. Frames

Frames are another way of adding a style to a web page or embedding one page within another. Using CSS is slightly more difficult than implementing frames but is the correct and most effective way of controlling the style of a web resource consistently between different browsers. Frames are also ignored by leading search engines which means that site content inside frames will not be indexed, thereby reducing the chances of potential users finding the site.

## Modularisation

Modularisation is the splitting of web pages into different sections (e.g. header, body, footer, etc.) and saving each of them with a unique file name for the sake of good maintenance and updating throughout the site. Modularisation

is a good practice which we would strongly recommend. One way of achieving this is to develop web pages using SHTML, which is similar to HTML except that such pages are assembled on the server and not on the client side browser. (In SHTML, the S stands for Server Side Include<sup>50</sup>). An SHTML web page easily lets you "INCLUDE" other pieces into HTML using instruction lines such as "`<!--#include virtual="filename"-->`".

## File naming

It should by now be apparent that a website comprises a collection of files on a server. Rather than naming these files without any overall guiding conventions, it is always worth considering the following:

- Is the file name meaningful and does it reflect the content of the file?
- Is it memorable and ideally no more than 10-15 characters long?
- Is there any reason whether it needs to use upper case, lower case or both? The implications may vary according to the type of server being used
- Does it really need to contain any special characters such as underscore ( `_` ) or hyphen ( `-` )?
- What should the file extension be? `html` would reflect a static web page, while `asp`, `jsp`, `php` would reflect the scripting language used

This list does not cover all relevant considerations but should help enormously in ensuring the maintenance and sustainability of the site.

### ReStore – Wise

While preparing a web resource for the ReStore repository (which uses Apache web server), we faced many problems with a resource in which the naming of the same web page was inconsistent within the site. At one place, the author of the resource hyperlinked using lower case e.g. `www.example.co.uk/directory/filename.htm` and at another place the same file was hyperlinked in upper case e.g. `www.example.co.uk/directory/FILENAME.htm`. Apache web server considers `filename.htm` and `FILENAME.htm` to be separate files. This type of difficulty would easily be avoided by naming files consistently and hyperlinking everywhere within the site in exactly the same case.

### ReStore – Wise

During ReStoration of another web resource, we found that the same file was hyperlinked by the author using both `htm` or `html`. In this case, `www.example.co.uk/directory/filename.htm` will be interpreted differently by an Apache web server from `www.example.co.uk/directory/filename.html`. Some web servers (such as IIS) ignore such differences and interpret the URL, but not all. When considering a web resource for long term accessibility and sustainability, it is strongly recommended to decide at the outset of the project which file extension should be used and then to follow this consistently.

49 <http://www.w3.org/standards/webdesign/htmlcss>

50 [http://en.wikipedia.org/wiki/Server\\_side\\_include](http://en.wikipedia.org/wiki/Server_side_include)

The ReStore team have worked with several websites which had originally been assembled inside institutional content management systems. Even for the authors' IT departments, it proved extremely difficult to extract page contents without completely disrupting the overall design and navigation of the resource. The extracted materials, not having been designed for standalone use, failed to conform to many of the standards of good design outlined here and substantial and disproportionate effort was required to reconstruct them so that they worked consistently. Most of these difficulties can be directly traced to the decision to use a CMS in the first place.

This can be easily done by placing a simple attribute "target" in the hyperlink, for example:

```
<a href="moreinfo.html" target="_blank">Click here for more information</a>
```

would open the contents of the file moreinfo.html in a new unnamed window.

## What does a web page contain?

In order for a site or a web page to be easily found by new users, the creator needs to try to ensure that the site is ranked amongst the top results displayed by search engines such as Google, Yahoo, MSN, etc. Thus, in terms of increasing impact, the way in which the content is presented to search engines may be far more important than the visual appearance of the site.

Search engines are the principal means by which users find new material on the web. Developing and enriching web pages with good metadata and well formatted content can help to ensure sustainability by increasing the chances of being found by these search engines. Quality of content, organization and structure of a web page all help to establish trustworthiness of the entire site. We list below some sections of a typical web page, each of which can be designed to increase impact and sustainability.

- Header
- Title
- Metadata
- Content (body)
- Web programming scripts (clever stuff)
- Footer

### Header

The header is the most important graphical element of a web site as it provides instant recognition. The header and any images (e.g. logos, organizational slogans, etc.) contained inside it will be seen more than any other element of a typical site. It is therefore critically important that it looks good on every page. It is common practice for each page to have its own header script contained in the top section. This needs to be standardised by placing the header section into a separate file and including that file in each web page using the "Include" directive. We have already discussed how to use this directive with reference to the use of SHTML.

By doing this, a developer gets maximum control over the whole site and any modification only need be made once, as it is immediately and automatically applied to every constituent page. This type of consideration is especially pertinent to long-term preservation where a small change (for example an acknowledgement reflecting the conclusion of a project) may need to be made to every page in a site. Another method would be to control the header through the CSS file. As discussed earlier, a CSS enables a web resource author to control the overall look and feel, managing the available space in each web page so as to give a distinct identity to the site.

Under no circumstances should a special character such as an apostrophe be contained within a file name (e.g. filename's.htm). Such file names are not recognised by some servers and if the file name has to be included in any web scripting code, this will always cause errors.

Another area already mentioned is whether static or dynamic files are to be created. Most dynamic web page files end with php, jsp, asp, aspx, pl and semi dynamic with shtml. The decision to develop a dynamic page which will be processed by the web server and then passed on to the user's browser should be made with reference to the contents of the file. If the resource author wants to render a dynamically generated web page in the user's browser such as the web form mentioned above, then the correct filetype should be used for the scripting language and server in question.

We strongly recommend that resource authors consult their institutional IT service providers before deciding to create dynamic web pages. This discussion will immediately bring to light the type of web server on which their resources will be running and whether the intended operations can be implemented and supported.

## Descriptive hyperlink text

A hyperlink is descriptive when the linking text is sufficiently meaningful that the user can correctly predict the nature of the page it is linking to. Thus, authors should avoid using text such as "Click here", "Go to", "Read more" etc. as these will be meaningless when the page is indexed by search engines. Meaningful hyperlinks enable search engines to crawl and index web pages effectively, thus enhancing the overall probability that they will be highly placed in the results displayed following a user's search. Well documented web pages leave a good impression on users, thus boosting retention and decreasing deflection.

Alerting users that clicking on a hyperlink will open a new window enhances user's browsing experience helping them keep track of all opened browser windows.

## Search Engine Optimisation

Search engine optimisation (SEO) refers to the strategies that can be adopted to ensure that web pages will be listed by search engines such as Google, Yahoo, etc. These search engines routinely “crawl” the web, examining pages and indexing their characteristics in order that they may be presented to users in response to searches. Those pages which best fit the specified search will appear higher in the list of results. At the time of writing, this ReStore guidance was appearing as the third result in a Google search for “online resources guidance”, however the similar search “online resources author guidance” did not produce any reference to ReStore on the first page of results. This is because the phrase “author guidance” is strongly associated with publishers’ guidance for authors of journal articles and therefore the results are primarily the author guidance pages of highly-read journals. Authors of online resources should therefore give careful thought to the terms which might be used by people who want to find out about the subject of their materials.

SEO is not a one-off task: when a site is first created, care should be taken to ensure that pages are provided with titles and metadata which most appropriately describe their content – this information will then be picked up by the search engines. However, tools such as Google Webmaster and Google Analytics can be used to examine the extent to which a site is being found and the search terms which have been used by visitors. Monitoring of this information can inform further improvements in page titles and metadata, thereby further increasing visibility and usage. Further options such as submitting a sitemap may increase the visibility of pages deeper within the site.

An exception to these comments is paid advertising, which ensures that the author’s content will appear prominently on the search results page when certain keywords appear in the search. There may be circumstances under which this is considered to be an appropriate strategy for an academic project although great care needs to be given to the intended audience and likely keywords. It is more often used, for example, in the promotion of a specialist Masters programme which brings an income stream into the institution than for advertising the content of online research support materials.





The ReStore team  
National Centre for Research Methods (NCRM)  
Faculty of Social and Human Sciences  
Building 58, University of Southampton  
Southampton SO17 1BJ  
United Kingdom

Web. [www.restore.ac.uk](http://www.restore.ac.uk)  
Email. [info@restore.ac.uk](mailto:info@restore.ac.uk)

The logo for the National Centre for Research Methods (NCRM). It features the letters 'NCRM' in a bold, sans-serif font. The 'N' is dark blue, and the 'i' is a lighter blue with a yellow dot. The 'C', 'R', and 'M' are a medium blue. Below the logo, the full name 'National Centre for Research Methods' is written in a smaller, dark blue, sans-serif font.

NCRM  
National Centre for  
Research Methods