# ReStore

A Sustainable Web Resources Repository

Version: Release V1.2

Guidelines for sustainable online resources

Sustainability principles for ESRC-funded online resources

ESRC National Centre for

Research Methods

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL
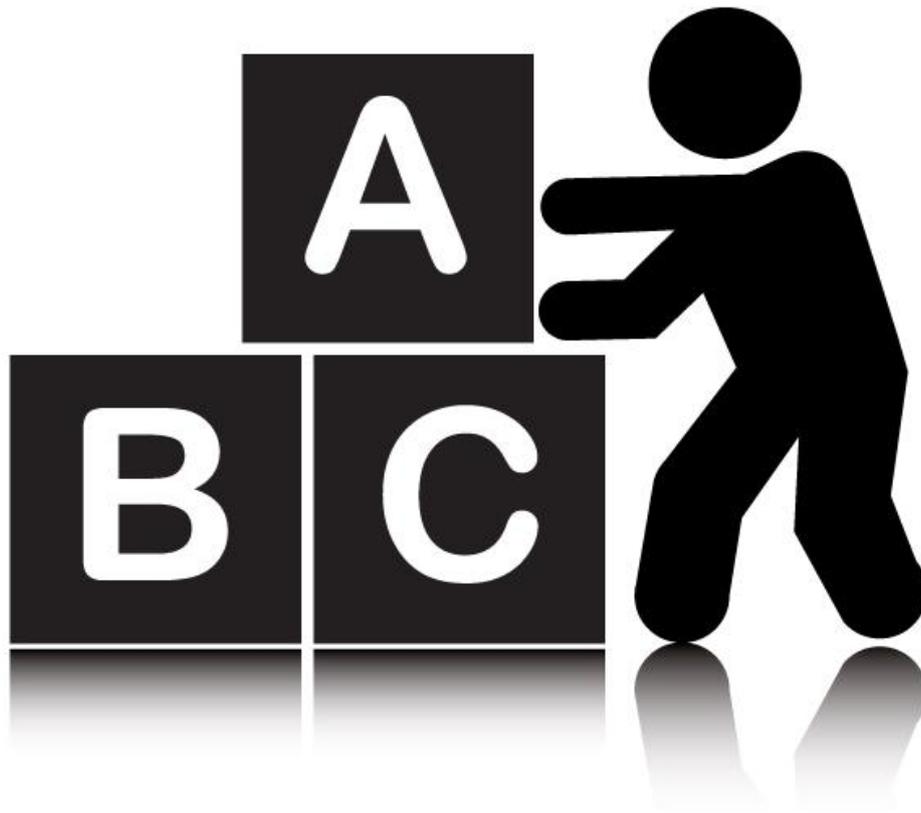
Conventions used in this document

These guidelines materials 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 on the left of the box)

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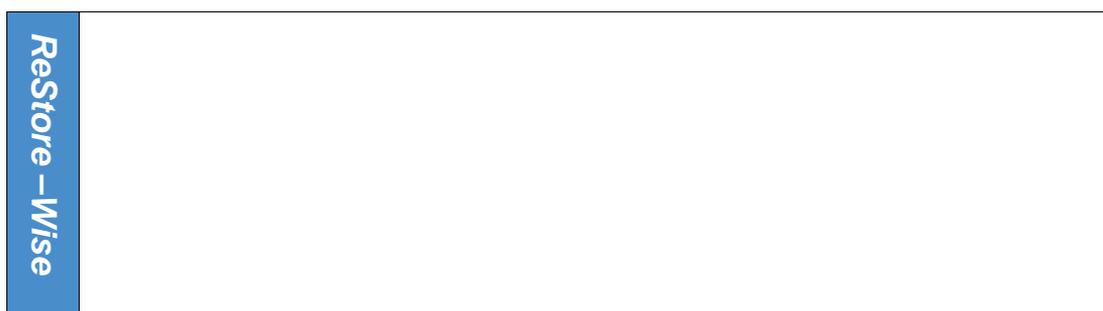
(b) Provide important information largely based on the specifics of the ReStore project (Horizontal blue header on the top of the box)

Information

Any discussion of this type 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, for example by using FAQs or exploring a tag cloud.

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).

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[1]. 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 websites, but not all content is worth such effort. An important question which arises is therefore how to discern the quality of web content when there is no straightforward litmus test. In particular, how should we determine what needs to be preserved for future use? Associated with this question would be issues relating to content ownership and legislation which aims to protect the rights of the creator or owner.

Alongside the enormous potential of the web as an information resource, the relative scarcity of funding for research and teaching demands that research funders spend extremely wisely when investing monetary resources 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,

---

[1] **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).

particularly by considerations of sustainability from the very outset of the 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:

1. 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.
2. 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 and/or owner and aspects of quality will all be considered here. These guidelines, primarily aimed at the creation of sustainable online resources, address 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.

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. Such researchers may 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 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

Chapter 2: Why sustain web resources?

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 online resources 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 researchers, including postgraduate students, mid-career academics and experienced research staff in the commercial and public sectors. Online 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. 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.

Figure 1

End of project funding

Utility

User awareness

A

Quality of resource

Time from project start
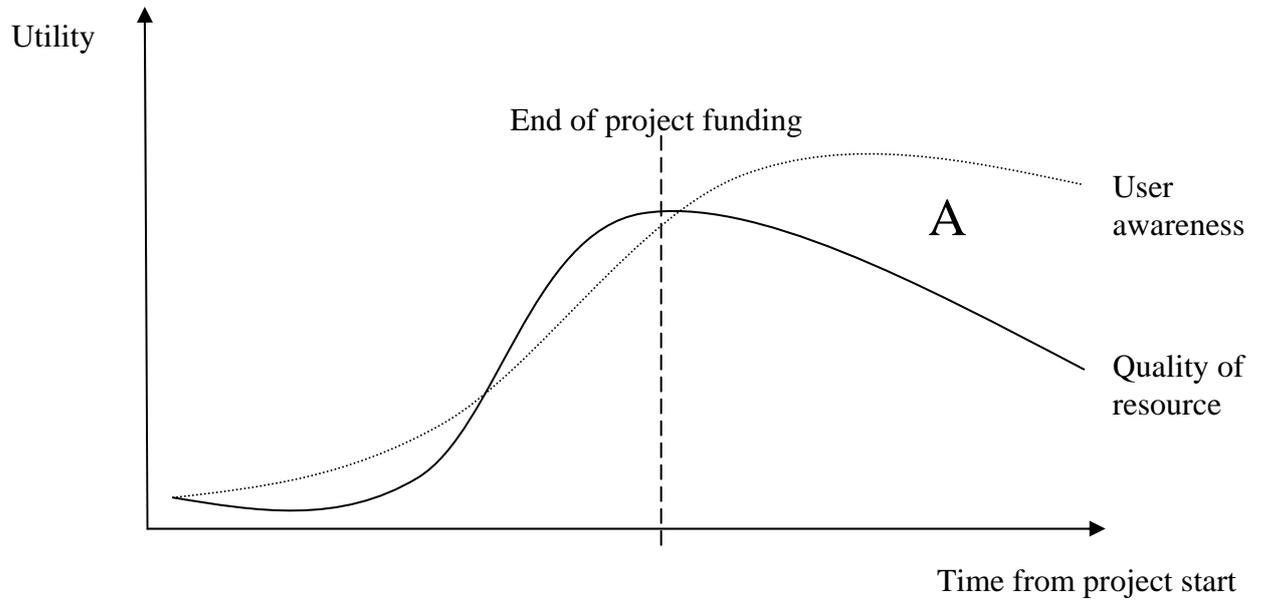
What are "Sustainable web resources" all about?

Sustainability generally refers to something having the quality of lasting for longer due to its resilience and robustness. A web resource in this sense is termed sustainable if it stands the test of time and maintains its original shape with the least possible human intervention. The preservation and upgrading of a typical web resource ideally needs to be addressed before its contents are outdated, its links become non functional or it ceases to be presented appropriately on 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 three 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 longevity and quality 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[2]. 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

---

[2] http://www.dcc.ac.uk

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[3]. 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[4]).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 3rd party consent documents, software licenses and ad hoc contract documents
- Gathering details on 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 whiles the core contents remain valuable to the wider research community. At this stage there is still an opportunity to keep the resource updated and maintained by transferring into a suitable repository such as ReStore, thus preserving the effectiveness of the resource. ReStore adhere to the policy of restoring the content, appearance and behaviour of a typical web resource. We have developed a set of questions, attached at Appendix 1, which help the original author of the resource to assess the readiness of their material for ongoing preservation in this way.

---

[3] http://www.restore.ac.uk
[4] http://www.archive.org/web/web.php

Some practical issues: experience from ReStore
Experience with the ESRC ReStore project allows identification of types of online resource that are particularly requiring of investment in sustainability:

i.   Resources that will not 'work' unless updated and maintained. For these resources a lack of up-dating means that some or all of the investment would be lost. Typically examples are where web resources relate to a particular analysis package that routinely changes, eg SPSS, STATA, NVIVO. In some cases exemplars will no longer run if users are not working with the same software version as the web resource. In other cases the software now features a new development and resource pages which do not reflect this become unhelpful.

ii.   Resources that provide a coherent and integrated web-site with a large amount of material where a failure to up-date would quickly result in poorer quality and consequence fall-off in value. Many resources require links to other pages that, if no longer available, will quickly reduce the value of the resource. 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 question of access control and that of for how long a resource should be maintained. It is recommended that wherever possible, standalone resources should not be restricted to certain users, for example by using federated access management or localised 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 user management. 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. It is generally only practical to guarantee the active maintenance of online resources for fixed periods of time – it is recommended that ongoing maintenance be reviewed after no more than three years to ensure that the resource is still serving a user base 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 of the resource; 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 that may be an ideal time to consider its transfer into a static web archive.

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/PING), audio (mp3, mp2, wav) and video files (mpeg, dat) etc. are 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 – 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 computer (e.g., Internet Explorer 6, 7, 8, Netscape Navigator4, 5, Opera, Firefox, Google chrome, Flock 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 the 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 by the funding body for inclusion in their own website e.g. ESRC Society Today (http://www.esrcsocietytoday.ac.uk). Although useful during the lifetime of the project, these sites contain limited academic content and the key information is likely to be retained elsewhere on the web. Such a site may be a potential target for an archiving initiative for various purposes but the direct benefits to the end user are limited. Such web site would normally fail to meet the criteria set out for the ReStore repository in Appendix A.

Reference sites

These may contain all the information from a simple project site but also contain a library of some 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 UK National Archive [5] for major ESRC-funded datasets or UKDA Store[6] for self-archiving of smaller collections). 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 ESRC require the inclusion of publications arising from the projects they fund within ESRC SocietyToday. In addition to publishers' own websites and library catalogues, publication details are becoming increasingly searchable and retrievable by multiple routes. 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 the most prestigious outputs cannot readily be shared 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. Essentially, ReStore aims to provide a well-recognised, maintained archive for precisely this type of online material.

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

---

[5] http://www.nationalarchives.gov.uk
[6] http://www.data-archive.ac.uk

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[7].

Such web site besides having attributes and components of other web sites, maintain an active connection with a remote or local database for displaying contents as per users' inputs. In other words a web site is called a deep web site when some of its contents are generated dynamically after specific requests have been submitted to it by users such as searching for a product, datasets or book title. A typical example would be UKDA-store[8] where large collections of datasets and other digital objects are kept which can be accessed through web pages having an active connection with that database.

ReStore –Wise	The ReStore search engine is a typical example of a dynamic web site. We store into our MySQL database all web links, keywords, titles harvested from various web resources on the ReStore site which in turn improves precision in search results.

---

[7] http://www.digitalpreservationeurope.eu/publications/briefs/preservartion_blogs.pdf
[8] http://www.data-archive.ac.uk

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 intuitional 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.

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. (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[9]. 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 UK Web archive, UK Data Archive and others focuses mainly upon preservation and Curation of web contents and after having preserved the snapshot of the web resource, they almost forget about the creators and/or authors of the web resource. ReStore makes a difference in this regard by ensuring collaboration with the authors before and after the ReStoration in order to keep the web resource up to date and fully functional. This extra efficiency surely 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 contents over it with the collaboration of the original creator.

Proposer vs. funding provider

Current financial pressures, more than ever before, force various Governmental and non Governmental funding bodies to think carefully and stringently before releasing funds to investigators. In most cases, a proposer has to articulate issues of impact and value of the outcome. The impact could either be assessed in tangible and/or intangible ways which may involve a finished web resource, document repository, knowledge, experience gained etc. when it comes to web resources, creating them has become far easier in today's world of social networking and web communities. How to preserve valuable contents and how to evolve a strategy to ensure what is produced

---

[9] http://www.academy.gcal.ac.uk/cd-lor/learningcommunitiesreport.pdf

will be ultimately preserved selectively, is however the most daunting task discovered so far.

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 would not only engender a sense of responsibility among investigators, developers and managers but also maximise the chance of creating sustainable resources. An online resource being created, with preservation in mind at the start of the project, would save time and money currently required to restore a web resource into ReStore repository after the cessation of initial funds. This in return will augment the impact of the finished web product at a very low cost thus giving a very high return on investment to fund provider like ESRC. The aim of including this section into this document is to stress the importance of web resource preservation before even the resource is created in the first place.

Consistent collaboration with authors

Initiatives like UK Web archive, UK Data Archive and others focuses mainly upon preservation and curation of web contents and after having preserved the snapshot of the web resource, they almost forget about the creators and/or authors of the web resource. ReStore makes a difference in this regard by ensuring collaboration with the authors before and after the ReStoration in order to keep the web resource up to date and fully functional. This extra efficiency comes at a cost but fulfils the basic criteria of web resource preservation in its true spirit. It also helps in maintaining the resource by publishing up to date contents over it with the collaboration of the original creator.

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 doesn't necessarily mean that all indexed web pages would be accessible to users after a few months. This type of preservation is normally termed as shallow web preservation and doesn't take into account long term sustained access to web resources. In other words, once a web resource decays and a large number of files in it become inaccessible due to various reasons (discussed in previous chapters), these search engines stop indexing those particular web pages. These factors cause the resource to appear substandard in the eyes of, users who stop visiting it gradually as time passes.

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?

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. These reviews are all implemented online. The process begins with a review request to the original author who is encouraged to reflect n 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/guidance/downloads/Author_Reviewer_Form.pdf . 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/guidance/downloads/Academic_Reviewer_Form.pdf. 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.

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 concerning 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 is 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 academic publication standards for conventional academic outputs in the equivalent 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, such 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 affinity with the user of the materials. Because web resources are not subject to a regular framework of academic review and associated indication of quality that comes from publication in a recognised book series or highly-cited journal, the web resource 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. but 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 (please 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 enthusiasm fades away when it comes to the more detailed pages. The primary reason perhaps may be the idea that users always stake 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.

- 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.

Frequency of content updates

Frequently updating content on a web resource site (e.g. every few weeks or months) keep 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[10], a free tool for monitoring all links in your site, will be of tremendous help to carry out this kind of updates.
- 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

---

[10] http://www.google.com/webmasters

browsers. Such testing is necessary because browser software is regularly updated by their developers, which in some cases may result in changing the look and feel of web pages.
- Updating text, images and other content if identified as necessary by regular review of the 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 seek and remove all superfluous materials and to ensure that everything presented 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. 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 menus backgrounds and foreground colours etc.
- Rhythm e.g. the structural arrangement of the various artefacts in a page and overall web site

Web resource usage statistics

The collection of web resource usage statistics is a key to understanding users' activity on a web site. Such statistics form a key component of any consideration 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 respective web resource creation projects:

- Google Analytics
- Google Webmaster & Google Sitemaps
- AWStats

Google Analytics
Google analytics[11] will provide insight into user activity and web site traffic on daily, weekly, monthly and yearly basis which is quite 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[12] tool is in many ways an extension to Google analytics as it can aid in improving web resource site traffic which will be reflected in 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 of activity, Google Sitemap[13] file 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 for crawling by search engines, which ultimately raises the usage ranking of a web resource site thereby increasing its visibility to potential users. Once a Google Sitemap has been manually created, it is submitted to the Google Webmaster tool.

AWStats

---

[11] http://www.google.com/analytics
[12] http://www.google.com/support/webmasters
[13] http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=156184

Unlike the Google-based site statistics tools, AWStats[14] 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.

[14] http://awstats.sourceforge.net

Chapter 6: Intellectual property rights (IPR)

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 IPR and there is a wealth of existing guidance available, including through the UK Intellectual Property Office website[15]. 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 giving 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, trade marks and designs. The rights of the owner and/or creator in 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 [16]. 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

---

[15] http://www.ipo.gov.uk/types.htm
[16] http://www.opsi.gov.uk/acts/acts1988/UKpga_19880048_en_1.htm

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 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 the copyright owner is, can be important as it is the copyright owner who 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 and where 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. So in terms of a website an example would be that the concept of a website hosting and maintaining research methods project websites would not be protected but someone couldn't copy the material or content. Using the same selection of project websites, organising them in a similar way and having similar look and feel would in all likelihood bring it into the realms of copyright infringement. In 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 you may have breached terms of the licence if you reverse engineer to find a work around).

Trademark
Most web resources will also include trademarks. These are brands and/or logos used to distinguish a product or service, and even an organisation supplying the same, and are covered by the Trademark Act 1994[17]. 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

---

[17] http://www.opsi.gov.uk/acts/acts1994/Ukpga_19940026_en_1.htm

publications (including websites). This would include permission from the funder and any collaborating partners as most will 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. Here is another stage when you should be careful as 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 something to consider when putting a resource together.

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 this sorted up front.

Below is some 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 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 can be 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 OK 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). 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[18], 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 site you link to 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

---

[18] http://www.sherpa.ac.uk/romeo

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 subject. 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 the 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[19], 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).

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

---

[19] http://www.sherpa.ac.uk/romeo

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[20].

Software
This 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 the place 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 posting the same. 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 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 want to

---

[20] http://www.microsoft.com/About/Legal/EN/US/IntellectualProperty/Permissions/Default.aspx

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 generally 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 they would 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. 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 taken about using these, however.

If considering using these, 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. IPR issues aside, which are substantial, 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 putting in a coherent structure could involve considerable time and effort (i.e. a project in itself) and may even be something they charge for.

Some other issues can be that 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
Running workshops or conferences often forms part of an overall project and if you do this 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[21] that the depositing institution would be able to use to get the retrospective permissions we would need.
Also be cautious if you decide to video the proceedings of 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

---

[21] http://www.restore.ac.uk/guidance/Contributor_permission_form.doc

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.

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 unless the site you link to agrees otherwise. However this can be frustrating and the user may not get to the item to which you are specifically wanting to guide them. 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 to maximise their usefulness to researchers. 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.

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.

Managing IPR in a resource

The importance of managing IPR in your 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 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 start at the outset as trying to cover all this retrospectively 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 than at the time when you are 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 organisations may 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 it to another organisation at the moment do try to think beyond the lifetime of the project and your current horizon. Potentially you, or one of your collaborating researchers, might move institutions; or perhaps you are delivering a service that the funder will want to put out to competitive call 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 rights included 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 first by you independently or even before the date the complaining party can demonstrate they generated it. Clearly in that instance it would demonstrate their work was not original. 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 weight to the value of your project outputs if you can demonstrate they are used into 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 only getting 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[22] which can be used as a template for this sort of thing. 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

---

[22]http://www.restore.ac.uk/guidance/IP%20Register%20Form%20Web%20Resource%20record.xls

and it is quite possible that your institution has developed their own tool, database or other system you can use.

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 discussed at regular 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 to ReStore the University of Southampton would be vulnerable for law suits from that copyright infringement being levied at them as they would now be hosting and making available the infringing material. This 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 you to work collaboratively with us 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[23].
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 & 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.

---

[23] http://www.restore.ac.uk/guidance/Author_Reviewer_Form.doc

The alternative for doing 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 discussions so far and 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 University of Southampton as hosting organisation for ReStore. 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 University of Southampton, who will be the new hosts, 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[24] is available online.

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[25] are available online.

---

[24] http://www.restore.ac.uk/takedownpolicy.php
[25] http://www.restore.ac.uk/terms.php

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 foul of this and is more publishing, distributing and disseminating than use for your own research purposes in any event. 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 whether it is or is not for commercial purposes and just because an educational institution may be non profit making, this does not mean it would not be considered a commercial purpose 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.

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 ReStore/University of Southampton. 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 who will be very 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 where this is being combined with authoring the academic content which is naturally the key focus of attention. Getting this right may not appear to be a distraction from an academic project, but getting it wrong may well prevent users from staying on a site long enough to 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) which can be created using a wide variety of software editors.

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 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. If the design is such that a developer or author has to go to every single page to insert the code for 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 learn 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 site users, and may make the difference between someone telling another user about the site or not. The URL of a web resource should not be longer than 78 characters to avoid wrapping across a line feed inside an html editor, email message or browser. Shorter URLs are easier to spell and people often directly type them into their browser rather than accessing them from bookmarks or searches.

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 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.

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, 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 available for download free of cost and can be installed with minimum features to conserve disk space usage. Academic 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. Since this runs on the server rather than the browser, the probability of serving an incomplete page will be substantially reduced.

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[26].

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. 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,

---

[26] http://www.w3.org/TR/html4/charset.html

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 licensing, security, maintenance and support implications of different 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 and feel style of 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 at W3C website[27] 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

---

[27] http://www.w3.org/standards/webdesign/htmlcss

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 different sections (e.g. header, body, footer, etc.) of web pages 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 encourage web resource authors to follow. 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). An SHTML web page easily lets you "INCLUDE" other pieces into HTML using special directives such as "<!--#include virtual="filename"-->". For further details on server side includes, please see Wikipedia entry on Server Side Include[28].

File naming

Frequently the naming of files during web resource creation is undertaken without any overall guiding conventions being adopted. It would be good practice to consider the following questions:
- 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 enhancing the maintenance and sustainability of the site.

---

[28] http://en.wikipedia.org/wiki/Server_side_include

	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 name 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.
ReStore –Wise	

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 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. This could be easily done by placing a simple attribute "target" in the hyperlink directive inside a web page.

Linking to specific sections inside a web page

In order to direct readers to a specific section within a longer web page, it is good practice to annotate specific areas of the page and tag each section by giving it a proper name for direct hyperlinking. This is also referred to as hash tagging of a web page. Hash tagging involves two steps: a) naming sections using <a name="name of tag"> and b) calling the tagged section through hyperlinks e.g. <a href="#name of tag">.

During hash tagging consideration should always be given to lowercase and uppercase tags. For example if a section has been tagged with <a name="filename">, the calling tag must exactly match <a href="filename">. If the cases of the tag and link do not match (e.g. filename, Filename), navigation will not work inside the page. You can read more on this topic from html tutorials on the web.[29, 30]

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 major 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 of far more importance 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

---

[29] http://www.w3schools.com/HTML/html_links.asp
[30] http://www.w3schools.com/HTML/tryit.asp?filename=tryhtml_link_locations

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 can be made once and is immediately 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. By accessing all images, logos, etc in a single CSS file and including the CSS file in the web page (as opposed to placing links to these items individually in each page), a developer not only makes life easier for themselves but also for anyone who may come to maintain the site at a later date.

Title
The title is of obvious importance because it informs users about the content of the page in a few words. Authors should review page content and take care to choose appropriate titles. Almost all search engines after indexing a web page, identify it by its title. The more meaningful the title of the page, the greater the likelihood that it will be highly ranked in search results. In the absence of explicitly defined metadata in a web page, the title is treated as an important descriptor of page content which further underscores its importance.

Metadata
Metadata is typically described as "data about data", and is a central element in the creation of a web page. It gives the user information about the page, e.g. its location, author, date of creation, copyright, intended purpose etc. There are various metadata standards which could be adopted at the time of web

resource creation which immensely increase the visibility of a particular web resource in user searches. The most commonly used metadata standard is called Dublin Core[31]. The simple Dublin Core Metadata Elements Set (DCMES) consists of 15 metadata elements, as follows:

> Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights

Each of these elements is optional and a web resource author is not required to provide descriptions for all 15 elements. Any element may be repeated if there are multiple relevant descriptive details. For example, the Contributor element can be included more than once if there is more than one contributor to page content. To create metadata for a simple web page, a Dublin Core Metadata Editor is available at UKOLN[32].

Most similar sites offer automatic metadata extraction from a web page, which makes embedding metadata into the respective pages quite easy and less technical. The Dublin Core standard is a widely used metadata standard for web resources because it offers all the metadata elements which may be required by current and future search engines in order to index, preserve and retrieve a particular web page. Good metadata also facilitates the long term digital preservation of pages using approaches such as Open Archive Initiative (OAI) harvesting.

Content (body)
The central part of a web page contains the actual content. This should always be placed between opening and closing body tags (i.e. <body> content </body>). In other words, if the body section is left blank, nothing will be seen on the user's screen. Attention must be given to the use of various special characters when formatting text within the body.

For example the "©" copyright symbol in an HTML file may not be interpreted correctly by browsers other than Internet Explorer. In the case of dynamic pages, web servers may not interpret the special symbol correctly. To get around this problem the HTML name for the symbol "&copy" should be used. Similarly, instead of using the symbols for less than "<" or greater than ">", the names "&lt" and "&gt" could be used to ensure that browsers render the page correctly. Almost all HTML editors offer access to these functions in their user interface menus. For further details on special character encoding please see the sections on Character Encoding.

Web programming scripts
In addition to the above, a static HTML web page may also contain JavaScript code which is interpreted by the user's browser rather than the web server. For example in many static web pages, JavaScript is included to create drop down effects when user hovers their cursor over buttons or tabs within the page. Another common use is the creation of a chained menu, created

---

[31] http://dublincore.org
[32] http://www.ukoln.ac.uk/metadata/dcdot

following the user's selection from the available options. All such code is referred to as client side scripting. The following are examples of situations in which a web resource author needs to include client side script in their pages:

- Using Google Analytics[33] for site usage statistics
- Generating a dynamic menu in a site e.g. drop down links menu such as that at Geo-Refer[34] online resource
- Creating special effects with clicks, hovering cursor, opening/closing current files
- Online form validation
- Dynamically managing the display of text within a page, e.g. showing/hiding specific content

Footer

The footer, as the name suggests, comes after everything else in a web page document. The footer of a web page may contain links to "Accessibility", "Contact", "Copyright statement" and "Disclaimer" pages. Like the header, the footer should ideally be kept in a separate file accessible through an "Include" directive. In the case of dynamic web pages, such as PHP the scripting language includes script would readily serve the purpose.

---

[33] http://www.google.com/analytics/
[34] http://www.restore.ac.uk/geo-refer