

## **Web Continuity Service – Glossary of Terms**

### **Archived Content**

Web content which has been captured, quality assured and made available for online access in the NRS Web Archive.

### **Capture**

The process of copying digital information from the web to a repository for storage and archival purposes.

### **Crawler**

Software that explores a website by following hyperlinks within the site, finding and copying web content as it goes.

### **Dynamic content**

Parts of a website which are generated by software on the web server, normally in response to an action by the user (e.g. typing in a search term).

### **External links**

Links within a website which point to content that is hosted on a different website.

### **Hyperlink**

A reference to web content that an online user can directly access either by clicking or by hovering. Hyperlinks point to whole documents or to a specific element within a document that is hosted within a website.

### **Instance**

A specific capture of a website in an archive, either as a sole capture, or one in a series of captures.

### **Live Website**

The current online version of a website, as opposed to historic web content which has been archived and made accessible in the NRS Web Archive.

### **NRS Web Archive**

The collection of archived instances of websites which have been captured, quality assured and made publicly available by the NRS Web Continuity Service. This collection is publicly accessible at <http://webarchive.nrscotland.gov.uk/>.

### **NRS Web Continuity Service**

The Service that enables NRS to archive the websites of our stakeholder bodies, and offer the opportunity for these stakeholders to enable web continuity redirection on their live websites.

### **Quality Assurance (QA)**

The process of checking the completeness of captured web content, on factors such as availability, content, and navigation.

### **robots.txt protocol**

A convention to control which parts of a website are accessible to web crawling software, comprising a set of rules specified in a 'robots.txt' file located in the top-level directory of a website.

### **Seed**

A URL that acts as the starting point from which a web crawler uses to explore and capture content from a website.

**Target Website**

Websites which are selected for archiving by the NRS Web Continuity Service.

**Uniform Resource Location (URL)**

A type of URI which identifies both the resource and its location. It therefore acts as an address for networked resources such as web content.

**WARC (Web ARChive) file**

The ISO file format for combining and preserving multiple digital resources into an aggregate archival file, together with related information. These types of file are traditionally used to store web archive instances, as sequences of content captured and archived from the World Wide Web.

**Web archiving**

The process of capturing, preserving and making available web content in the long term.

**Web browser**

A software application for retrieving, presenting and traversing information resources on the World Wide Web.

**Web continuity**

A redirection service that can be enabled on a live website server to take users from broken links on that site into the NRS Web Archive, where a search for an archived version of the missing page will be made, and served if found. This means users see many fewer '404 page not found' error messages when visiting these live sites.

**Web server**

A computer programme which receives HTTP requests from clients (usually web browsers), and 'serves' the requested web content to them. The term may also be applied to the computer on which the web server software is running.

**Website**

A website is a collection of related web resources, usually as grouped by some common addressing.