

France - UK: complementary views on web archiving

 britishlibrary.typepad.co.uk/webarchive/2015/03/france-uk-complementary-views-on-web-archiving.html



Considering the nature of the web, it is fairly impossible to archive all of it, and choices have to be made. Usually two strategies are combined. The first one aims at being representative, by collecting a sample of everything without discrimination. The second one selects websites in order to build a collection, the way libraries are used to do with more traditional material. UK and France both combine the two methods.

UK has recently changed its legislation (6th April 2013) to embrace non-print resources in the legal deposit scope, including web sites. France had already done that shift in 2006.

Both national libraries use robots to broadly crawl the national web every year. In UK the crawling is done by the British Library. The National Archive also collects web sites related to government ([UK Government web archive](#)), but this comes under another regulation, the Public Records Act. In France, [INA](#) (Institut National de l'Audiovisuel) archives all the websites related to radio and television, while [BnF](#) (Bibliothèque nationale de France) is in charge of all the rest.

To complete this broad harvesting both countries create collections on specific topics, made of websites collected by curators in their area of expertise. To do so, national libraries may be helped by partners: researchers, associations but mostly other libraries. In UK five other libraries are in charge of legal deposit and participate in web archiving. In France a similar partnership goes on with the network of regional libraries, also contributing to legal deposit.

In BnF, the Digital Legal Deposit Department coordinates a network of correspondents in each department, where specific policies have been developed through the years. What's happening now is that the global BnF's selection policy is being updated and will include websites, considering they are not different from any other material, which makes sense.

Breadth vs openness

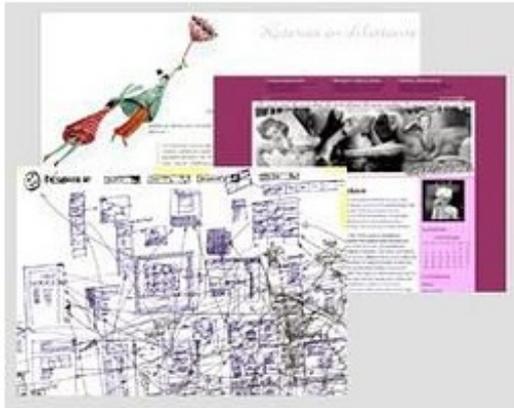
The websites collected for legal deposit purpose can only be consulted in the libraries reading rooms, for copyright reasons. But while all the websites collected by BnF are only accessible in the reading rooms dedicated to researchers, British Library gives access on the [UK Web Archive](#) to a part of its collections. This showcases websites for which permission has been obtained. This process is of course very time consuming and frustrating, for only 30% of the permission requests receive a positive answer and the vast majority receive no answer at all.

Exploring the collections

BnF proposes a research through URL and a guiding approach through specific topics, in order to give an overview of the collections. For example, one of its remarkable selections is related to private diaries on the web. Others may concern elections, sustainable development, science, and many others

themes.

(S')écrire en ligne : journaux personnels et littéraires



L'internet s'affirme de plus en plus comme le champ d'une façon nouvelle de concevoir l'écriture et l'expression autobiographique. Ce parcours guidé, essentiellement centré sur l'écrit (désigné comme « littéraire » ou pas, car ces frontières bougent avec l'Internet) souhaite rendre compte des conséquences de ce passage du papier à l'écran, et montrer comment la forme du blog a transformé le journal intime, l'écriture et la critique littéraires.

Parcours guidé – (S') écrire en ligne : journaux personnels et littéraires

© BnF

It's similar in the open UK web archive where you can browse the archive by special collection (Queen Jubilee, Northern Ireland...). As in France, the choice of a topic is often related to current affairs. At the moment, a collection about [Magna Carta](#) is being developed regarding the exhibition to come, as well as one concerning the next General Election.

Openness seems to be a good goal for highlighting the collection. The Open UK Web Archive is promoted via British Library's website, this blog, Twitter... It provides fine visualisation tools and most importantly pretty good research functionalities. They're based on title, URL and dates. There's a full text index too for the massive legal deposit crawl and this is quite remarkable. (To give an idea of the magnitude of the task, it will take about six months to generate the 2014 crawling's index). Then, when you type a research, you sometimes get really a lot of results and it can be far from easy to go through them, but this is [another issue](#).

Browse by Special Collection

Each Special Collection is a group of websites, usually more than fifty and less than four hundred, brought together on a particular theme, either events-based (e.g. The Olympic & Paralympic Games 2012), topical (e.g. The Credit Crunch Collection), or subject-oriented (e.g. The British Countryside Collections). These have been especially compiled by curators and other subject specialist to make useful and interesting Special Collections.

Remember you can use the Search tips



19th Century English...



Blogs



British Countryside



Cambridge Network



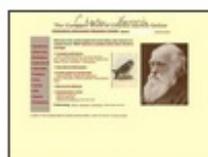
Commonwealth Games G...



Cornwall



Credit Crunch



Darwin 200



Energy



E-publishing Trends