Speaker’s Notes: Examining sports history through digitised & born digital resources

Helena Byrne, Sporting Irish Lives, Ulster University, 11/11/2022

Helena.Byrne@bl.uk

**Slide 1 – Title slide**

* Information about where and when the workshop is taking place with contact information for the facilitator.

**Slide 2 – Agenda**

1. Warm Up Activity
2. Digital Resources
3. Digitised Newspapers
4. Web Archives
5. Hackathon - Preserve Irish sporting heritage online.
6. Wrap Up Activity

**Slide 3 - Warm Up Activity**

* A link to an online poll with the following questions. This can be done using an online polling tool or by a show of hands.
1. Do you consent that the anonymised data collected in this activity can be used in future publications? Yes/No answers
2. What department are you affiliated to? Free text answers
3. What sport(s) do you study? Free text answers
4. Do you understand the difference between the terms digitised and born digital? Yes/No/Not sure answers
5. Have you ever used a web archive? Yes/No/Not sure answers
6. Have you have ever used the UK Web

**Slide 4 - Search Tips for Digital Resources**

* This slide gives a very basic definition of the various types of digital resources researchers may come across.
* Digital Archive/Digital Library - Contains a mix of digitised and born digital objects. Examples include correspondence, newspapers, digitised books, eBooks, academic journals etc.
* Web Archive - Contains born digital objects published online only. Examples include websites, subsections of websites, individual web pages and some social media.

**Slide 5 - Search Tips for Digital Resources**

* Searching digital resources can potentially bring up thousands if not millions and billions of results. Such large results are impossible to review effectively.
* Using Boolean Operators, Phrase Searching, Truncation and Wildcards can help you refine your search results. These methods are generally universal and should work on most digital resources.
* Most platforms provide ‘Search Tips’ so if there is anything unique for that particular platform it will be highlighted there.
* The image on Slide 4 illustrates how using the Boolean Operators – And/Or/Not refine the type of results you get when searching sports terms.[[1]](#footnote-1)
* There is also a link from Ulster University Library, which explains how to use Phrase Searching, Truncation and Wildcards search techniques.

**Slide 6 - Search Tips for Digital Resources**

* Phrase Searching example – Gaelic Football search in Trove on 01/11/2022 brings back **5,332** results.
* This is a huge number of results that would be quite difficult to review all of them for relevance to a research question.

**Slide 7 - Search Tips for Digital Resources**

* Phrase Searching example – “Gaelic Football” in quotation marks search in Trove on 01/11/2022 brings back **938** results.
* Compared to the first figure this is a lot more manageable to review. Most digital resources also have facets that can help refine your search results even further.

**Slide 8 - Digitised Newspapers**

* Using digitised newspapers was covered extensively in the previous session at this workshop so I will not spend much time on this topic.
* The two main platforms for researching Irish sport using digitised newspapers are the Irish Newspaper Archive (INA) and the British Newspaper Archive (BNA).
* These are subscription services that you might have free access to through your academic library or you may have to pay for access yourself.
* Trove is an example of a digitised newspaper archive that is open access and although the publications are from Australia, the previous slides have shown that it may contain content relevant to studying Irish sports history.
* Although the newspaper archives have huge amounts of content in them, they also have large gaps in their collections.
* In the article, ‘Where are we now? A review of research on the history of women's soccer in Ireland’ I give an overview of some of the pitfalls of working with digitised newspapers.[[2]](#footnote-2)
* One way to identify gaps is to search for a particular event using a short date range. This way you will see if your search gives you any results and if the key newspapers that generally publish about your chosen sport are represented.
* If the newspapers you want to review do not come up in this search then you will need to visit a library that holds physical copies of the newspaper whether that be in the original format, microfilm or microfiche.
* If you are interested in researching women’s football of any code in Ireland then the crowdsourced ‘Mapping Irish Football’ dataset is a resource that will be useful to your study. This is held in the British Library Shared Research Repository: <https://doi.org/10.23636/ndrx-sf60>

**Slide 9 - Using Web Archives in Your Research**

* You can use web archive content in your research in two ways.
* A resource that you use alongside other traditional research resources.
* A point of study.
* Most citation guidance does not mention archived websites as a source so until there is specific guidance it is best to follow the guidance they give for the live web if possible.

**Slide 10 - Data Management**

* Lots of Copies Keep Stuff Safe (LOCKSS) applies to everything. You have probably been told to keep backup copies of your thesis before. The same also applies to any digital resources you might be using that are not managed by a trusted repository, Library or Archive.

**Slide 11 - Data Management**

* Link rot is the term used to describe when links on a website are broken. This is when you would usually get a 404 Error Code, instead of the content you wanted to see.

**Slide 12 - Data Management**

* Content drift is the term used to describe where content on a web page has changed from when it was first published.

**Slide 13 - Data Management**

* There are a number of studies done to measure the prevalence and impact of link rot and content drift. The blog post link on this slide gives a good overview of this issue and how it can be resolved.

**Slide 14 – Web Archives**

* Time Travel is a union catalogue of web archives.
* Time Travel helps you find and view versions of web pages that existed at some time in the past. It is linked to all the open access web archives that use the Mementos Protocol.
* Many web archives have onsite only archives so the URL you are looking for might have been preserved in a closed access web archive and will not come up in this search.

**Slide 15 - Web Archiving in the UK and Ireland**

* This slide list all six of the public web archives available to researchers in the UK and Ireland.
* This slide gives an overview of the web archive name, the organisation(s) running the web archive, a link to their website and the access status of the web archive.
* Some universities and cultural organisations run small-scale web archiving programmes.
* With the exception of the UK Web Archive, all of the web archives are accessible online due to the Non-Print Legal Deposit Regulations 2013. Access to the UK Web Archive is at one of the six UK Legal Deposit Libraries unless permission has been given by the website owner to make the content open access. The title of all content is included in search results on the UK Web Archive website so users can review what content related to their research area is in the archive before visiting a Legal Deposit Library.
* Suggested reading: Jane Winters, ‘Giving with one click, taking with the other: e-legal deposit, web archives and researcher access’

<https://sas-space.sas.ac.uk/9439/1/Giving%20with%20one%20click%2C%20taking%20with%20the%20other.pdf>

**Slide 16 - Sports in the UK Web Archive**

* Sport has always been an important part of the UK Web Archive collection since it launched in 2005.
* Current sports collections set up in late 2017.
* Three collections dedicated to sport – Sports Collection, Sports: Football, Sports: International Events.
* The suggested reading links on this slide gives an general overview of our sports collection policy and most blog posts published on the UK Web Archive blog will be tagged with the ‘sports’ tag. However, many older blog posts might not have any tags and need to be searched manually.

**Slide 17 - Current subsections (Handout 1)**

* A breakdown of the three collections dedicated to sport – Sports Collection, Sports: Football, Sports: International Events.
* Some sections have subsections depending on the sport/classification.
* The general sports collection subject headings were based on the Universal Decimal Classification page about sport (from PD 1000 – 2003 UDC Abridged Edition). We used this general taxonomy of sports so that the collection can easily adapt to new sporting trends that emerge in the future.
* The Ball Sports section excludes football, as there is already a dedicated collection on this subject. Ball sports is probably the most versatile section and this has an additional five subsections.

**Slide 18 - UK Web Archive Derived Data**

* In the suggested reading listed here ‘Reviewing Soccer History in the UK Web Archive’ I explain what SHINE and the UK Web Archive Open Data are and give examples of how you could use them in potential research projects related to soccer. These approaches could easily be applied to any sport.
* In the article, ‘Where are we now? A review of research on the history of women's soccer in Ireland’ I also give an overview of using SHINE as a research resource.

**Slide 19 - DIY Web Archiving**

* The two examples Conifer and Webrecorder allow anyone to archive content on the web. This is a manual approach so can capture a higher fidelity copy of the website than the Heratrix Crawlers used by bigger intuitions as these crawlers can only follow clickable links and usually cannot capture embedded content hosted outside the main domain like YouTube and social media.
* Using these tools you have to manage the archived data yourself either through a cloud storage account or on a local device.
* The UK Web Archive, the Internet Archive Wayback Machine and Arquivo.pt from Portugal all have a way to self-archive web content without having the responsibility of managing the archived content.
* The UK Web Archive will only archive content from the UK but the others will archive content from anywhere.
* If you are nominating content, you should remember LOCKSS and nominate the content to as many web archives as possible.

**Slide 20 – UK Web Archive**

* One the UK Web Archive website there is a ‘Save a UK Website’ tab on the home page.
* It brings you to a page where anyone can nominate UK published online content within scope under the Non-Print Legal Deposit Regulations 2013.
* The key points of what is in and out of scope is highlighted on this page.
* An email address is only required if you own the copyright to the content you are nominating as they would like to send a request for open access after it has been added to the archive.
* Web content that is going to change or go offline should be nominated with advance warning, as the nomination process is manual. Meaning, there could be a few days between when content is nominated and when it is archived.
* Nominations can be included in multiple collections. If you are nominating content that fits into a specific collection(s) then please add a note in the ‘Any other information box’.
* Content tagged into a curated collection is available to view quicker than content added to the general collection. This is due to the sheer volume of data available to search so there is a lag on when ‘Frequently Crawled’ and ‘Domain Crawled’ content becomes available to view.

**Slide 21 – Internet Archive Wayback Machine**

* You can see on the Internet Archive Wayback Machine page that there is a ‘Save Page Now’ section.
* Add a URL to the box and push the ‘Save Page’ tab. The URL you have entered will archive in real time and the archived version is available to view straight away.
* If you want to archive a full website then you have to follow the same process for each individual page.
* Even if the Internet Archive has copies of the web content you are citing, it is still best practice to use the ‘Save Page’ option when you are viewing the content, as they might not have that page in their collection.

**Slide 22 – Arquivo.pt**

* When you enter a URL into the Arquivo.pt search bar and press the ‘Gravar’ tab you can browse through the website and everything within that domain that you browse through will be archived.
* Any external links on your chosen URL will have to be added individually.
* Embedded content from outside that domain is challenging to capture. I added the Sporting Irish Lives website to this archive and when I reviewed the archived copy, the programme for the conference was not visible because it was embedded from a Google Drive. I tried to archive the Google Document separately but it did not work because Google would not give the archive access to the document.
* The content is archived in real time, but the process of adding these captures to the archive is manual. The nominations are usually added to the archive once a week so there could be a short lag on when you can review the content you have added to Arquivo.pt.

**Slide 23 - Hackathon**

* Use some of the DIY web archiving strategies to preserve Irish sporting heritage.
* There are two options for this activity.
1. Add online content that you have used in your research to the relevant web archives.
2. Review what web content has already been preserved from your area of study in the UK Web Archive Sports Collections. Then select online content from the web to nominate to the UK Web Archive.
* Online content includes – a full website, a subsection of a website, an individual page from a website, an event page and Twitter accounts. In the last few months, Twitter has changed the way they publish their content making it more difficult to archive but at the moment we are still scoping Twitter content into the UK Web Archive collection. Other social media platforms although mostly in scope are too difficult to capture.
* The suggested reading on this slide ‘What content should I nominate on the UEFA Women’s Euro to the UK Web Archive?’ gives an overview of the type of content the UK Web Archive collects and although this example focuses on the UEFA Women’s Euro tournament it could be applied to any subject area.

**Slide 24 - Wrap Up Activity**

* A link to an online poll with the following questions. This can be done using an online polling tool or by a show of hands.
1. How likely are you to use a web archive as a resource for your research? Yes/No/Not sure
2. How likely are you to save content you view online in a web archive? Yes/No/Not sure

**Slide 25 – End of slide deck**

* Repeat contact information.
1. Image credit from UCD Library: <https://libguides.ucd.ie/litreview/searching> [↑](#footnote-ref-1)
2. <https://doi.org/10.1080/17460263.2019.1604422> [↑](#footnote-ref-2)