

Simple Sunsetting with GitHub Pages & The Internet Archive

John Pilbeam, Sainsbury Library, University of Oxford

✉ john.pilbeam@sbs.ox.ac.uk 🐦 @jpilbeam



Background

In 2020, we began a project to sunset our decade-old research repository. There was no longer a need for it due to institutional changes in research management, but there were no resources or budget to sunset it. Therefore, we required a cost-free and simple way of sunsetting the project. This poster aims to show how to archive and preserve a project without a budget.

Objectives of the project

- Gracefully shutdown (sunset) the EPrints repository.
- Preserve links.
- Provide access to archived versions of the content.
- Achieve results without a project budget.
- Ensure the longterm preservation of and access to the content.

Methods

- Adapted the Internet Archive's 404 handler to provide a redirect from the original URLs to a 404 page with a link to the archived version in our university's Archive-It service.¹
- Followed established principles for retiring digital projects.²
- Used readily available and free resources.
- Used our institution's web archiving service (Archive-It) to ensure longterm access and preservation.
- This solution would also work with content only available via the Internet Archive.
- Made the code repository available on GitHub.³

Results

- The repository domain name now points to an archive site hosted on GitHub Pages. This has maximised the chance of users finding any repository content that is still linked to elsewhere.
- All repository content is now preserved in the institutional web archive.
- All links to repository content elsewhere on the Internet or in print, now point to a custom 404 page with links to the archived content.

Conclusions

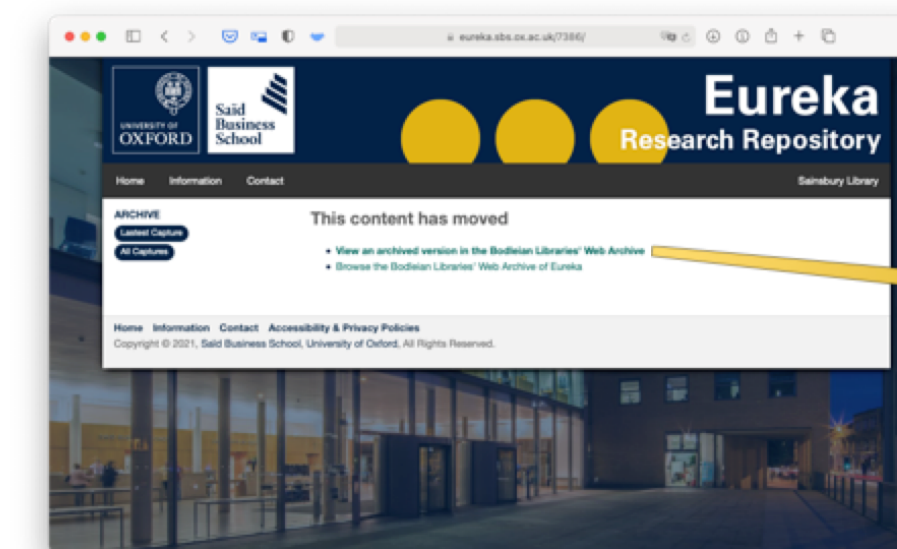
- Successful sunsetting project: all of the objectives were achieved.
- One minor downside is the use of a 404 page for what should really be a 301 redirect, but the benefits outweigh this issue.
- The approach is reproducible so we can use it again to sunset other projects.
- Other libraries are welcome to build on this solution and we would recommend exploring sunsetting options when closing down websites and services.

1) Home page and project pages



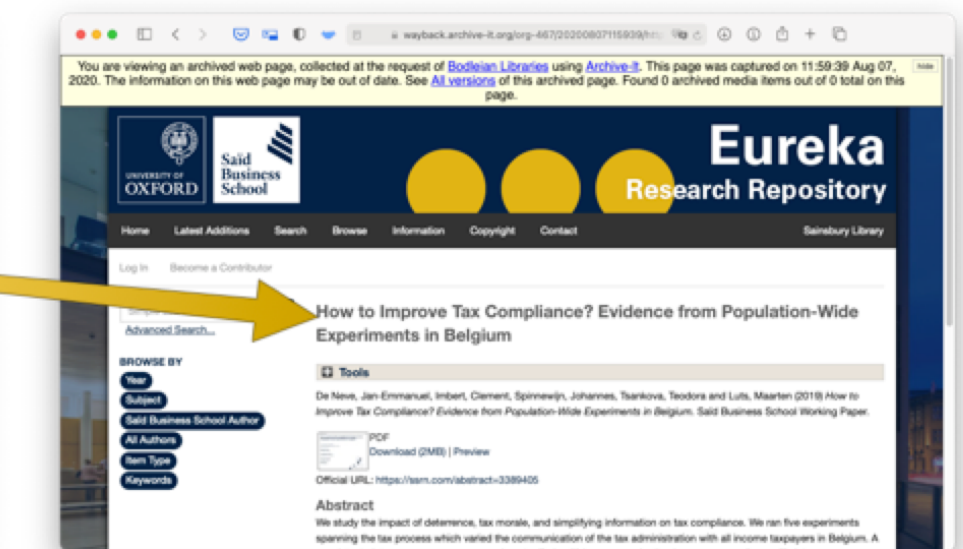
- Static website hosted on GitHub Pages.
- Foundation responsive front-end framework (based on the theme used by the original EPrints repository).
- The home and project pages show who worked on the project, describes the project and provides contact details for any legacy issues.

2) 404 page



- GitHub pages custom 404 page served instead of original web page.
- Redirect links created by a version of Internet Archive's Wayback Machine 404 script.
- Script modified to use most recent snapshot of the site in the institution's Archive-it. e.g. the page <https://eureka.sbs.ox.ac.uk/7386/> provides a link to the archived version at <https://wayback.archive-it.org/org-467/20200807110009/https://eureka.sbs.ox.ac.uk/7386/>

3) Archived content



- Preserved version of the page on Archive-It.
- Archive includes the original full-text where available.
- For libraries without access to an institutional Archive-It, the original Wayback Machine 404 script could be used to point to a copy on the Internet Archive.

Links: 1. <http://blog.archive.org/2013/10/24/web-archive-404-handler-for-webmasters/> | 2. <https://ronallo.com/sunsetting-book/> | 3. <https://github.com/sainsburylibrary/eureka>