Documenting Web Archiving Quality Assurance

New York Art Resources Consortium Web Archiving Fellowship

Olivia Meehan | omeehan@pratt.edu

NYARC

The New York Art Resources Consortium (NYARC) is stationed at the Frick Art Reference Library and includes the libraries and archives of the Frick Collection, the Brooklyn Museum, and the Museum of Modern Art. The NYARC Web Archiving program regularly archives the websites of all three museums along with the websites of art galleries, artists, and art resource organizations based in New York City. Over **9 terabytes** of website data have been collected by NYARC since 2010.

Web Archiving Basics

NYARC's archived websites are collected and managed using **Archive-It**, a platform developed by the **Internet Archive**. Websites are captured using a web-crawler that automatically navigates through a website and saves every file it encounters. Archived websites are saved as **WARC** (Web ARChive) files. Websites can also be manually crawled using **Conifer**. WARC files saved from Conifer can be uploaded directly to Archive-It collections. Archived websites in Archive-It can be viewed with the **Wayback Machine**.
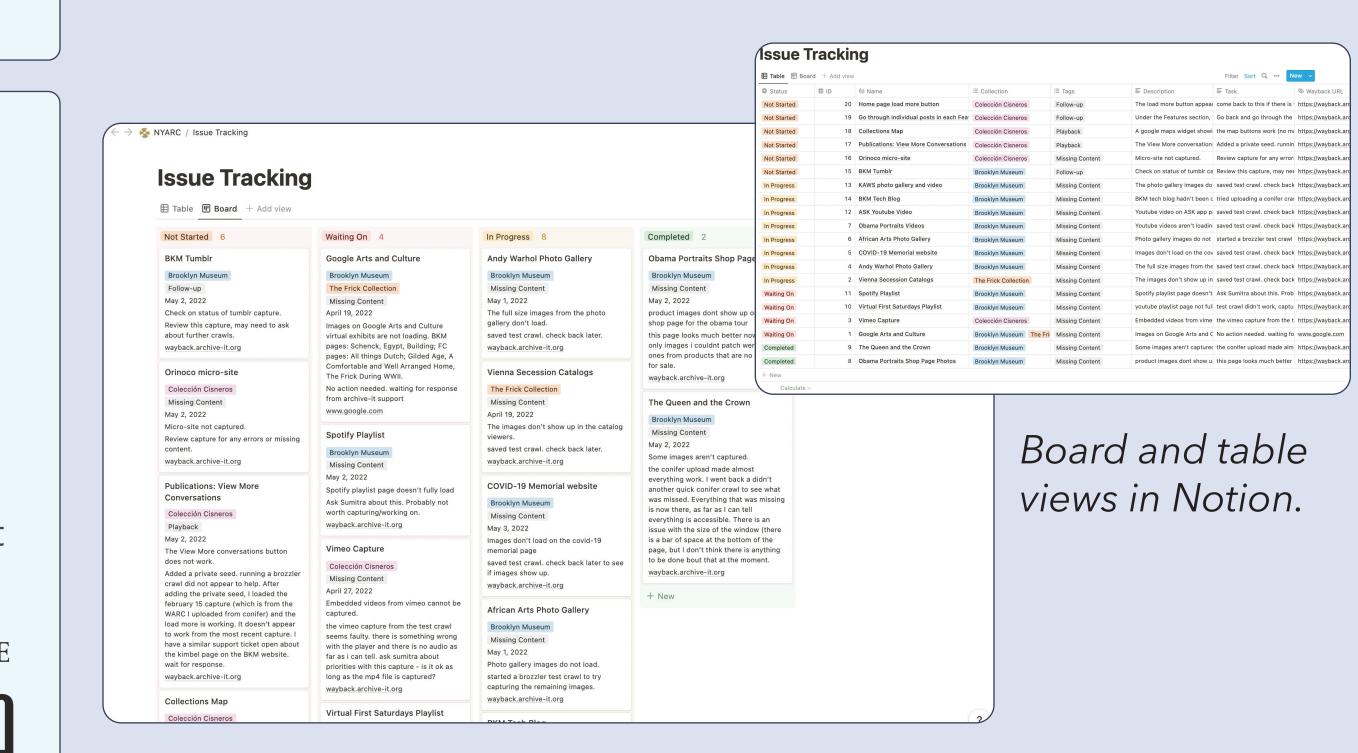
Quality Assurance

Newly archived websites should be reviewed for quality and functionality. Missing files (photos, videos, links, etc.) can sometimes be patched in to make the archive more accurately represent the live site. For art-focused websites, this process is especially important. With exhibit images, portfolios, virtual exhibits, and digital artwork, these sites rely heavily on visual structure and functionality. Making sure things look and work correctly directly improves the accessibility of the archived information and artwork now and in the future.

Documentation

Reviewing an archived website for quality and functionality is an inherently repetitive task. It requires the thorough review of every link and interactive element on each page, ideally in comparison to the live website. Documenting issues and solutions is an important part of maintaining progress and developing a record of recommendations that can be referred to for future quality assurance.

Board and table views in Notion.

Documentation Workflows

An effective documentation workflow should be easy to maintain and easy to understand. Too much manual note-taking is difficult to keep up with when you are trying to efficiently work through a large website. Short, undetailed notes may not be helpful to yourself or to others. Productivity and task management tools like **Notion** can help expedite the documentation process while retaining the detail and flexibility of a spreadsheet. In Notion, for example, you can easily build a small database with relevant properties. That database can then be displayed and easily edited as a project board, table, list, calendar, and more.
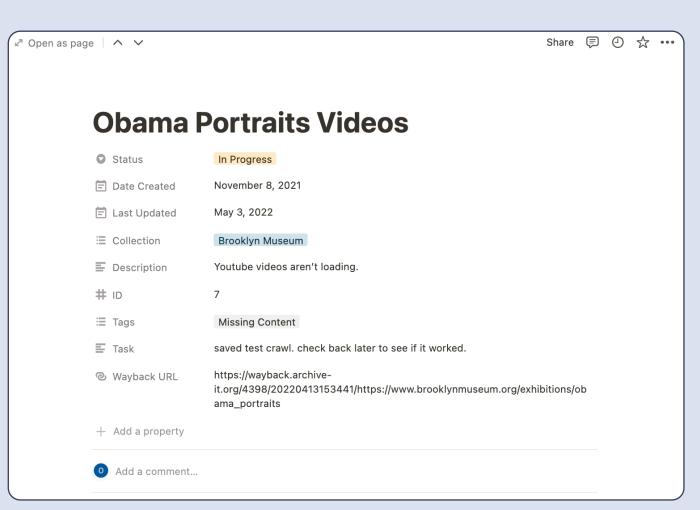
What to Document

Useful properties that can be included in a documentation workflow include:

1. Title
2. Status
3. Date opened
4. Date last updated
5. Collection or website name
6. Brief description
7. ID number for easy reference
8. Tags indicating type of issue
9. Next task required
10. Link to the archived page
11. Record of previous QA tasks for future reference.

Data input page in Notion.

Acknowledgements

Thank you to Sumitra Duncan, the Frick Art Reference Library and the Pratt School of Information for this fellowship opportunity.