

ARCH

Products

Archive-It

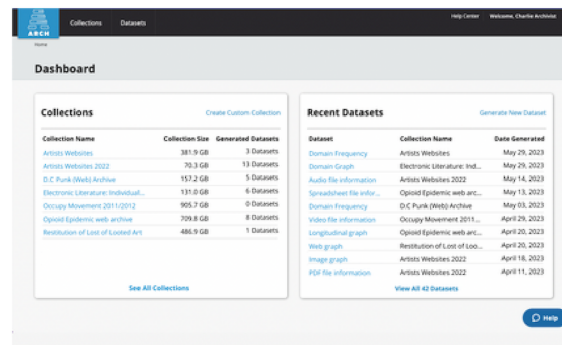
ARCH

Vault

Web Services

Overview

ARCH (Archives Research Compute Hub) is a research and education service that helps users easily build, access, and analyze digital collections computationally at scale. With ARCH's intuitive interface, users can build custom research collections relevant to a wide range of subjects, generate and access research-ready datasets from collections, and analyze those datasets. In line with best practices in reproducibility, ARCH supports open publication and preservation of user-generated datasets. ARCH is a key asset to research, classroom instruction, and collection curation. Optimized for web, text, and image collections, ARCH gives users the power to study and understand digital collections in new ways.



Collection Name	Collection Size	Generated Datasets
Artists Websites	381.9 GB	3 Datasets
Artists Websites 2022	70.3 GB	13 Datasets
D.C. Punk (Web) Archive	157.2 GB	5 Datasets
Electronic Literature: Individual...	131.0 GB	6 Datasets
Occupy Movement 2011-2012	905.5 GB	0 Datasets
Opusculum: web archive	370.8 GB	8 Datasets
Restitution of Lost of Losted Art	486.9 GB	1 Datasets

Dataset	Collection Name	Date Generated
Domain-Frequency	Artists Websites	May 29, 2023
Domain-Graph	Electronic Literature: Ind...	May 29, 2023
Audio file information	Artists Websites 2022	May 14, 2023
Spreadsheet file info...	Opusculum: web arc...	May 13, 2023
Domain-Frequency	D.C. Punk (Web) Archive	May 03, 2023
Video file information	Occupy Movement 2011...	April 29, 2023
Longitudinal graph	Opusculum: web arc...	April 20, 2023
Web graph	Restitution of Lost of Loo...	April 20, 2023
Image graph	Artists Websites 2022	April 18, 2023
PDF file information	Artists Websites 2022	April 11, 2023

ARCH Dashboard.

Features

Build

Leverage the ARCH platform's capabilities to build custom collections that are well-scoped for specific research and education purposes.

Access

Generate more than a dozen different dataset types (e.g., full text, images, pdfs, graph data, and more) from digital collections with the click of a button. Download generated datasets directly in-browser or via API.

Analyze

Easily work with research-ready datasets in interactive computational environments and applications like Jupyter Notebooks, Google CoLab, Gephi, and Voyant and produce in-browser visualizations.

Publish & Preserve

Openly publish datasets in line with best practices in reproducible research. All published datasets will be preserved in perpetuity.

Support

Make use of synchronous and asynchronous technical support, online training, and extensive help center documentation.

Pricing

ARCH is available for institutional and individual use and supports flexible access for researchers, librarians, archivists, museum professionals, journalists, and more. Reach out to learn more about pricing and availability.

Join Us

Contact us to learn more about ARCH.

Email: arch@archive.org

Internet Archive Archiving & Data Services: <https://webservices.archive.org/>

Internet Archive: archive.org