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.

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

Build
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.

Access
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.

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

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

Support

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