



CLIOX: CULTURAL HERITAGE AND ARCHIVES ECOSYSTEM

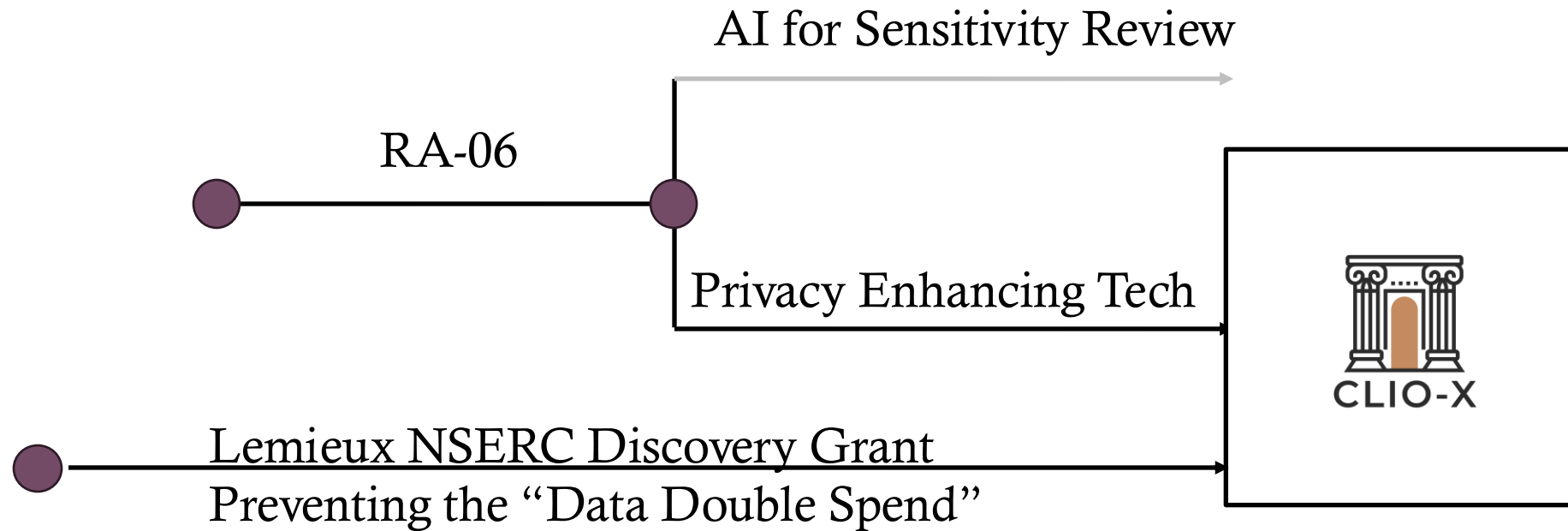
Prof. Victoria Lemieux

School of Information, The
University of British
Columbia

LUSTRE/GLOW Workshop

Manchester, July 2022,
2025

PROVENANCE OF CLIO-X RESEARCH



THE RESEARCH TEAM

- **UBC**

- Dr. Victoria Lemieux, Prof. Archival Science, UBC
- Dr. Zehua (David) Wang, Adjunct Prof in ECE, UBC
- Chris Zhou, MASc Student in ECE, UBC
- Gontse Faith Molaosiwa, MLIS Student, UBC
- Brandon Li, BSc Statistics, UBC, MASc in CS, SFU

- **U. Lleida**

- Dr. Roberto Garcia, Associate Professor of Computer Engineering and Digital Design
- Dr. Rosa Gil, Associate Professor of Computer Engineering and Digital Design

- **UNED**

- Luis de la Torre Cubillo, Prof. Titular de Universidad, ETSI. Informática



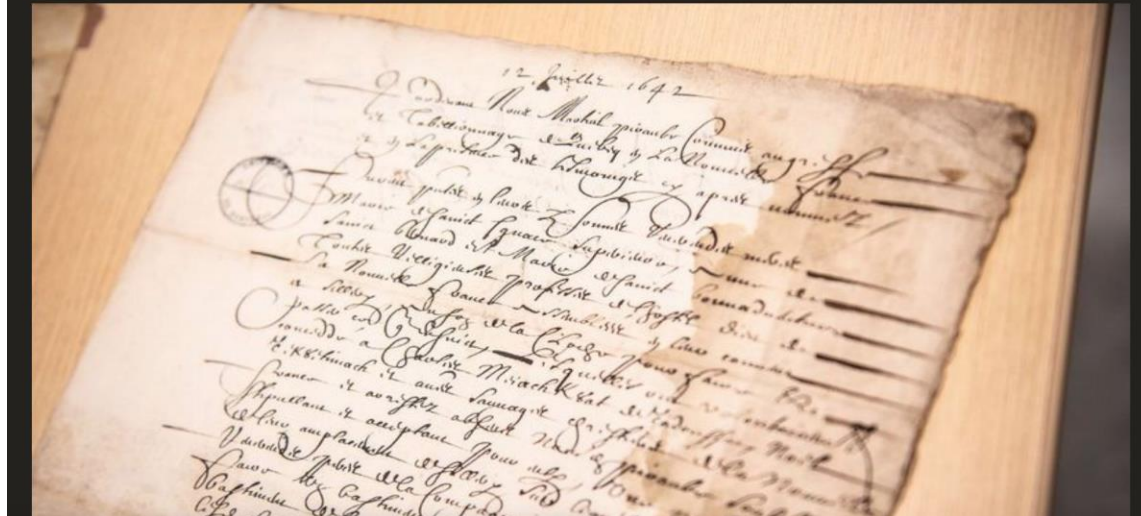


CLIO-X: MOTIVATION

Archives are increasingly using AI

Using AI to decipher historical manuscripts

[UDEMNOUVELLES](https://nouvelles.umontreal.ca/en/article/2023/11/08/using-ai-to-decipher-historical-manuscripts/) | 11/08/2023 | MARTIN LASALLE



Source: <https://nouvelles.umontreal.ca/en/article/2023/11/08/using-ai-to-decipher-historical-manuscripts/>

The Broad Vision is to build on the Infrastructure proposed in . . .

TOWARDS A PRIVACY-PRESERVING COMPUTE INFRASTRUCTURE FOR ARCHIVES

Victoria Lemieux

InterPARES Trust AI 9th Symposium

Vancouver, October 28, 2023

OPPORTUNITY #1, CONT'D

Privacy enhancing technologies:

➤ An emerging class of technologies and methods meant to “harness the power of data in a manner that protects privacy and intellectual property, enabling cross-border and cross-sector collaboration to solve shared challenges”

The White House. 2021. US and UK to Partner on Prize Challenges to Advance Privacy-Enhancing Technologies.

1



Protecting Privacy in Digital Records: The Potential of Privacy Enhancing Technologies

VICTORIA L. LEMIEUX* and JOHN WERNER*, School of Information, University of British Columbia, Canada

With increased concerns about data protection and privacy over the past several years, and concomitant introduction of regulations restricting access to personally information (PI), archivists in many jurisdictions now must undertake 'sensitivity reviews' of archival documents to determine if they can make those documents accessible to researchers. Such reviews are onerous, given increasing volume of records, and complex, due to how difficult it can be for archivists to identify whether records contain personal information (PI) under the provisions of various laws. Despite research into the application of tools and techniques to automate sensitivity reviews, effective solutions remain elusive. Not yet explored as a solution to the challenge of enabling access to archival holdings subject to privacy restrictions is the application of privacy enhancing technologies (PETs) - a class of emerging technologies that rest on the assumption that a body of documents is confidential or private and must remain so. While seemingly being counter-intuitive to apply PETs to making archives more accessible, we argue that PETs could provide an opportunity to protect PI in archival holdings whilst still enabling research on those holdings. In this paper, to lay a foundation for archival experimentation with use of PETs, we contribute an overview of these technologies based on a scoping review and discuss possible use cases and future research directions.

CCS Concepts: • **Security and privacy** → **Domain-specific security and privacy architectures**; *Information accountability and usage control*; *Management and querying of encrypted data*.

Additional Key Words and Phrases: Privacy Enhancing Technologies, Homomorphic Encryption, Trusted Execution Environments, Secure Multiparty Computation, Personal Data Stores, Privacy Preserving Machine Learning, Synthetic Data

ACM Reference Format:

Victoria L. Lemieux and John Werner. 2023. Protecting Privacy in Digital Records: The Potential of Privacy Enhancing Technologies. *ACM J. Comput. Cult. Herit.* 1, 1, Article 1 (January 2023), 19 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

A new era of access to archival documents is upon us. It has opened increased experimentation with using artificial intelligence (AI) tools to make archives more accessible and is enabling novel methods of exploring archival holdings (see, e.g., [70], [26], and [45]). Recent programs exemplifying this trend include AEOLIAN (Artificial Intelligence for Cultural Organizations) [45], Computational Archival Science (CAS) [58], and InterPARES AI

InterPARES 'WE_CAN'!

Phase 1 (2024-2025)

- discuss and finalize the solution requirements and architecture & identify initial test data (e.g., newspaper holdings, X (formerly Twitter) holdings).
- refactor the WB project and integrate it & MyCNTRL code to meet initial distant reading requirements (e.g., for sentiment analysis)
- implement with one strategic partner initially, with a view to scaling in Phase 2

Phase 2 (2025-2026)

- scale the solution to additional strategic partners
- identify additional AI functionality to support distant reading
- identify additional PETs & other technical functionality
- implement additional functionality into the solution

Beyond

- implement [Decentralized Autonomous Organization](#) to govern the shared infrastructure for participating archives & return generated revenue to maintain infrastructure and benefit archives



Explore archival data securely. Build knowledge collectively.

ClioX is a new kind of privacy-first platform—built by and for researchers, archivists, and cultural institutions. It's designed to help you explore, share, and collaborate on sensitive archival material with AI without compromising on privacy or ownership.

[Get Started](#)[Learn More](#)



CLIO-X OVERVIEW VIDEO

A COMPUTE TO DATA FRAMEWORK ENABLES DATA SOVEREIGNTY & ETHICAL AI



CLIO-X

TRUSTED EXECUTION ENVIRONMENTS ENABLE A SECURE & FEDERATED DATA ECOSYSTEM



CLIO-X



PRIVACY PRESERVING MACHINE LEARNING



CLIO-X

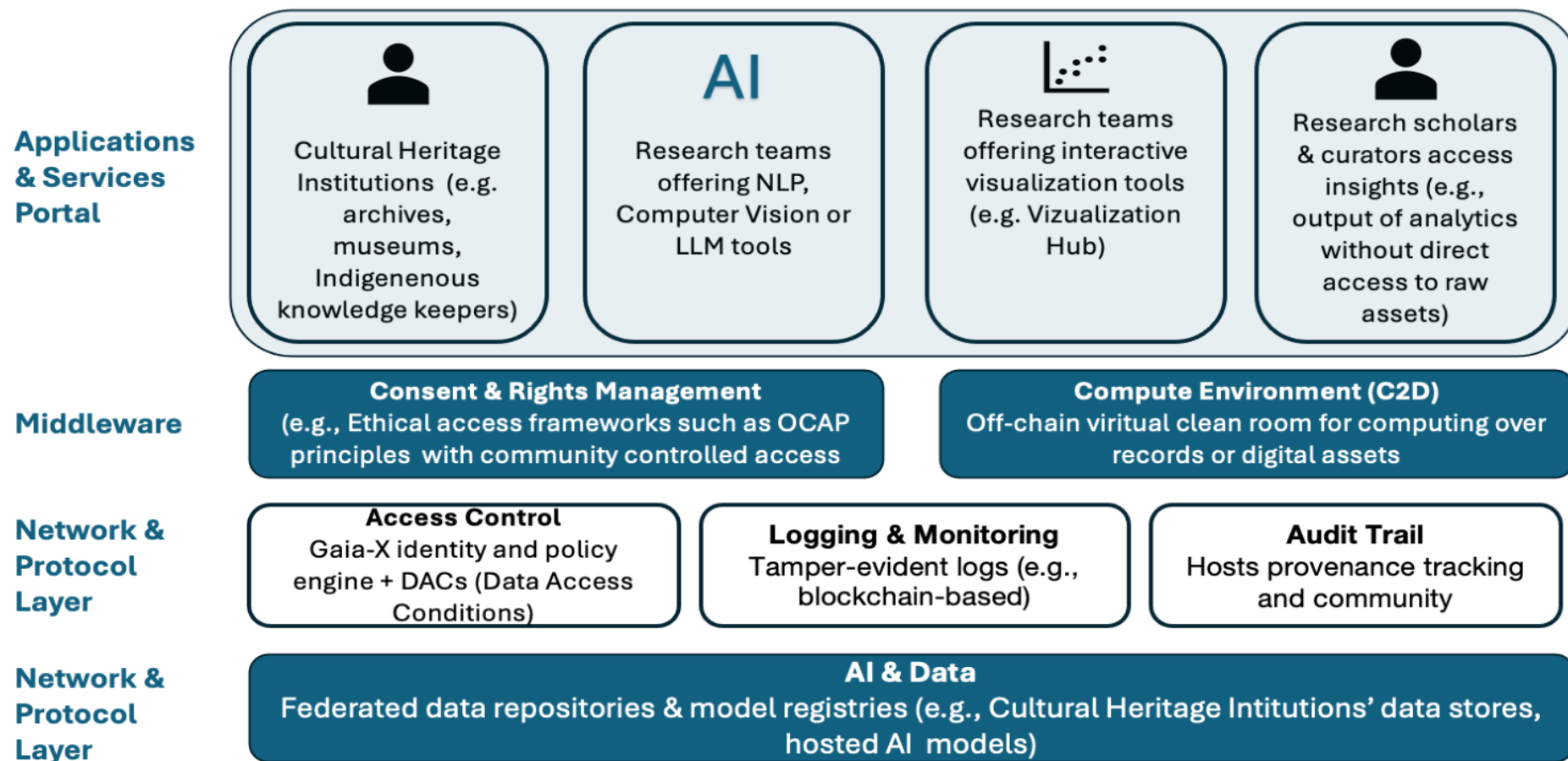


FIG. 1. OVERVIEW OF CLIO-X SOLUTION ARCHITECTURE (SOURCE: AUTHORS' OWN ADAPTED FROM PONTUS-X)

MANUFACTURING

● Live

Flex4Res

Co-funded by the
European Union



Flex4Res

MANUFACTURING

● Live

AIRBUS

Airbus

For those who make this world a safer
place. We pioneer defense and space for
their mission success.

GO TO PORTAL ↗

GENERAL

● Live



Pontus-X Main Portal

AGRICULTURE

● Live



AgrospAI

AI

● Live



beebucket

RESEARCH

● Live



Clio-X

INDUSTRY 4.0

● Live



Future Engineering Network

LANGUAGE MODELS

● Live



Berlin State Library

Explore archival data securely. Build knowledge collectively.

ClioX is a new kind of privacy-first platform—built by and for researchers, archivists, and cultural institutions. It's designed to help you explore, share, and collaborate on sensitive archival material with AI without compromising on privacy or ownership.

[Get Started](#)[Learn More](#)

Setup MetaMask

To interact with the portal, you need to install and create an account with MetaMask

Click on the button to download MetaMask from the official website.

Once you've set up an account, you are ready to move to the next step.

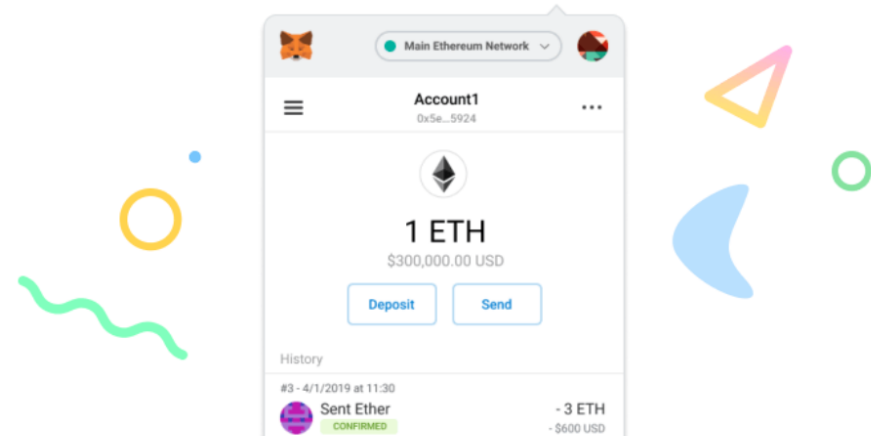
DOWNLOAD METAMASK

Chrome

iOS

Android

Install MetaMask for your browser



NEXT STEP

Choose Your Role

Select the path that best describes you to see your next steps.



The Archivist

Publish and protect your holdings
with tools built for ethical AI
stewardship and collaboration.



The Researcher

Explore archival datasets with AI to
gain data-driven insights.



Ecosystem Partner

Ready to join a global values-aligned
Web3 community?

Data NFT* ⓘ



PX Data NFT — PX-NFT
Data NFTs are unique digital assets that represent the intellectual property of your digital services.

Asset Type*

Dataset

Algorithm

Title*

Enron Test Dataset 2

Description* ⓘ

Dataset comprised of Enron Emails.

Service Credential ⓘ

e.g. <https://file.com/service-credential.json> Validate

THE ARCHIVIST

- Publish and protect your holdings with tools built for ethical AI stewardship and collaboration.

Access Type * ⓘ

Download

Compute

Access Controller URL * ⓘ

https://provider.test.pontus-x.eu

✓ File confirmed

File *

URL

API

SAAS

GRAPHQL

IPFS

File * ⓘ

https://raw.githubusercontent.com/ciferresearch/ClioX/refs/heads/core-components/core-components/data/enron/enron_subset.csv

✓ File confirmed 348.7 KB plain

Sample file

URL

File ⓘ

Personally Identifiable Information *

☒ The dataset offered contains PII

Data Controller * ⓘ

Legal Basis * ⓘ

e.g. GDPR2016:6.1.a

Purposes * ⓘ

e.g. ServiceOptimization, UserInterfacePersonalization

Data Protection Contact Point *

Consent Withdrawal Contact Point *

Terms & Conditions *

☒ I agree to the Terms and Conditions

ARCHIVISTS CAN PUBLISH METADATA ABOUT THEIR ARCHIVAL DOCUMENTS TO MAKE THEM AVAILABLE

Metadata

Access

Policies

4 Pricing

5 Preview

6 Submit

PRICED

FREE

Set your price for accessing this dataset. The datatoken for this dataset will be worth the entered amount of the selected base token.

Price

EUR0e

0

Must be more or equal to 1

Pontus-X Community Fee

0.1

Marketplace Fee

0

Earnings after fees

0

Back

Continue

id *

Address

8aB96B7CbB55A6E1c1054678137bA09780acA

Add

Address

8aB96B7CbB55A6E1c1054678137bA09780acA

Add

Conditions

/file.com/tandc.md

Validate

Conditions under which this service can be accessed. A resolvable link to the T&C document is expected. If no Terms are given the default terms of this portal will be

ARCHIVISTS CAN SPECIFY SPECIAL TERMS OF ACCESS AND WHETHER THEY WILL CHARGE FOR ACCESS



CataloguePublishVerifyPartners▼Resources▼



ConnectWallet



Search for service offerings

279 results

Filters 2

Clear filters

Service Type

☒ datasets

☐ algorithms

☐ saas

Access Type

☐ download

☒ compute

Sort

Type

☐ Relevance

☒ Published

☐ Sales

☐ Price

Direction

☐ ↑ Ascending

☒ ↓ Descending

COMPUTE | DATASET | CEP-IMGS

Pig pen images sequence for animal well-being assessment

Centre d'Estudis Porcins (CEP)

A sequence of images from video surveillance of one of the pens in the Centre of Swine Studies of Catalonia (CEP), an experimental pig farm managed by a cons...

2.5 EUROe

1 sale

Pontus-X Devnet

COMPUTE | DATASET | GXAT

[test] enron 100

University of British Columbia

for testing only with only 100 data lines

Free

1 sale

Pontus-X Devnet

COMPUTE | DATASET | GXAT

Final Data for Automated Energy Consumption Benchmarking for S...

Craftworks

Overview This data offering was generated within the EuProGigant Gaia-X Lighthouse project context. It can be used by an AI-Algorithm that enables metal part ...

Free

1 sale

Pontus-X Devnet

COMPUTE | DATASET | GXAT

DATA TEST 2 - Automated Energy Consumption Benchmarking for S...

Craftworks

DATA TEST 2 - Automated Energy Consumption Benchmarking for Sustainable Manufacturing - EuProGigant

Free

2 sales

Pontus-X Devnet

COMPUTE | DATASET | SYMBOL

Nautilus-Example: Compute Dataset Name

deltaDAO AG

Nautilus-Example Description This asset has been published using the nautilus-examples repository.

Free

Pontus-X Devnet

COMPUTE | DATASET | GXAT

[test] tokenized enron subset

University of British Columbia

cloud storage version

Free

1 sale

Pontus-X Devnet

THE RESEARCHER

- Explore archival datasets with AI to gain data-driven insights.



CLIO-X



SELECTING DATA AND AI ALGORITHMS
FOR COMPUTATION

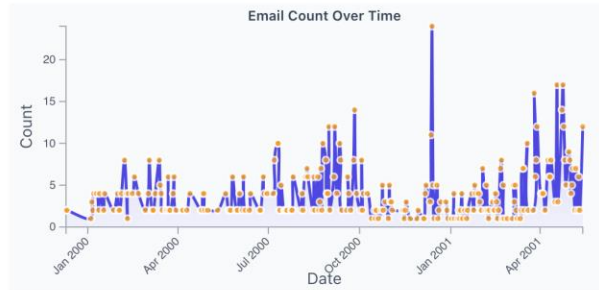
Data Distribution on Email Counts

Shows the distribution of email counts over time

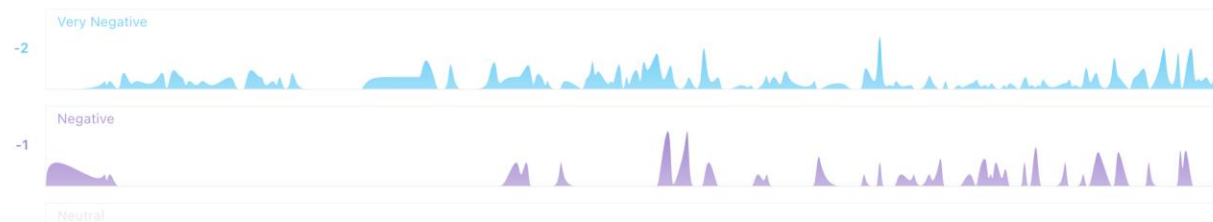


Data Distribution on Date

Shows the distribution of emails by date



Sentiment Analysis by Category

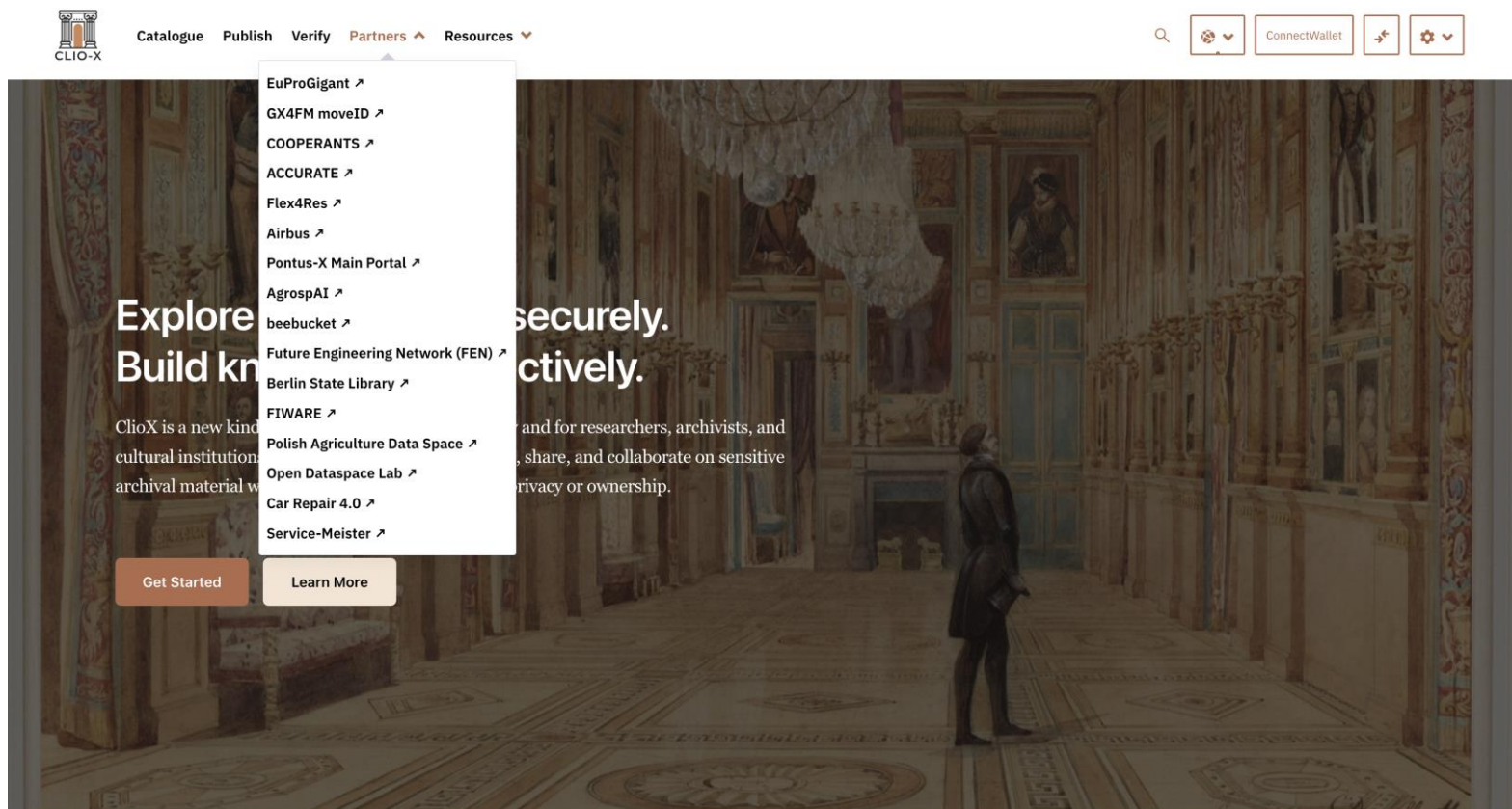


THE RESEARCHER

- Explore archival datasets with AI to gain data-driven insights.

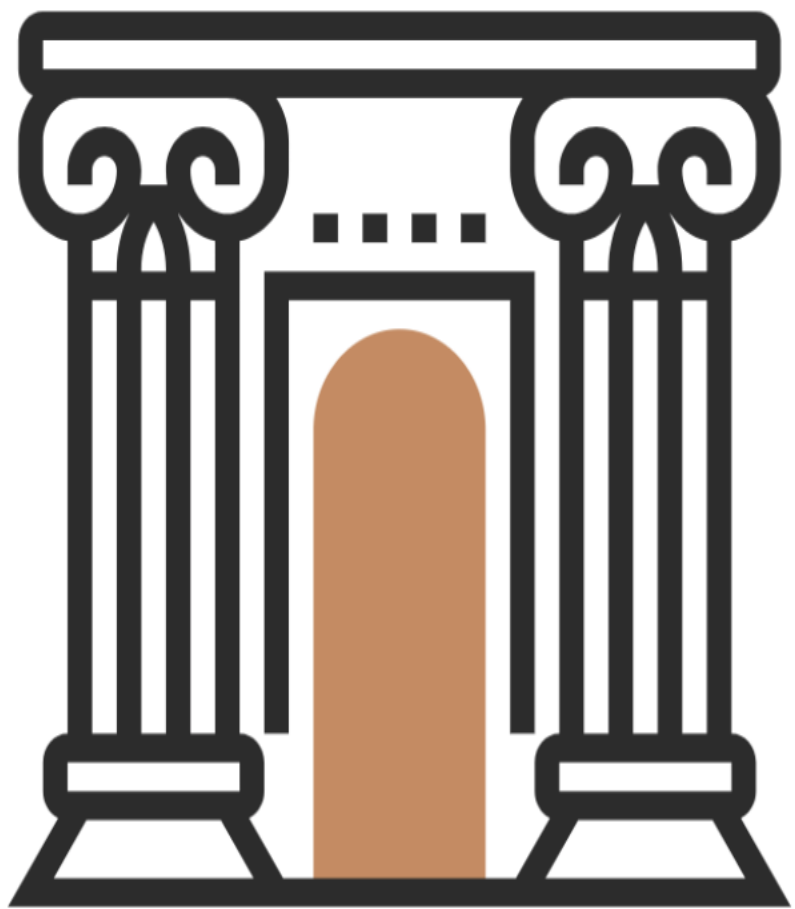


SELECTING DATA AND AI ALGORITHMS FOR COMPUTATION



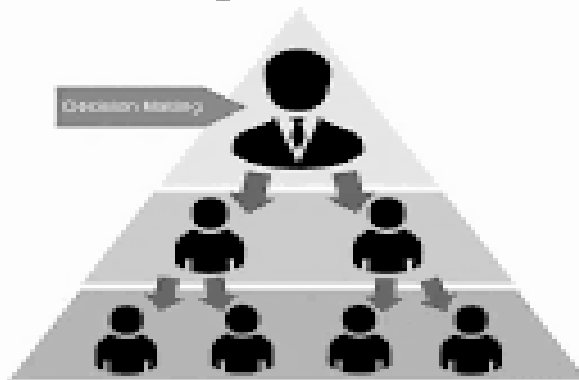
THE ECOSYSTEM PARTNER

- Ready to join a global values-aligned Web3 community?

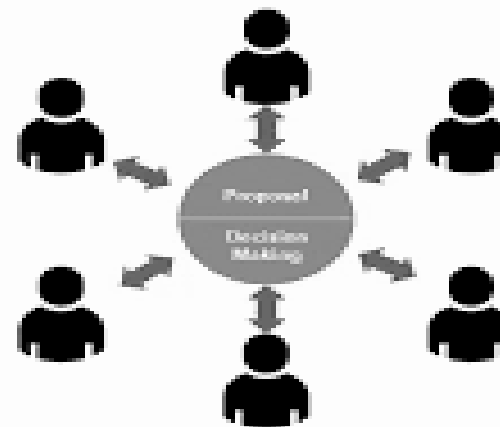


CLIO-X

Traditional Organization



DAO



Explore archival data securely. Build knowledge collectively.

ClioX is a new kind of privacy-first platform—built by and for researchers, archivists, and cultural institutions. It's designed to help you explore, share, and collaborate on sensitive archival material with AI without compromising on privacy or ownership.

If you are interested in trying
Clio-X, visit
<https://www.cliox.org/>
or to explore becoming a
partner, email
vlemieux@ubc.ca

Thank you!

Get Started

Learn More

Launch in June
2025 during IAW25



Onboarding first
partners



Blockathon
(hackathon) to
develop new
analytic tools July
2025