

Digital Humanities specialist for FAIR memories project (en)

Project Summary

The FAIR Memories project aims to enhance the long-term value and reusability of analogue oral history archives through the creation of a reusable workflow for their digitisation and FAIR-ification alongside an example dataset based on a unique corpus of interviews about consumer culture in early 20th century Italy. By transforming this legacy corpus into FAIR-compliant digital assets, the project aims to demonstrate how past research efforts can be ethically integrated into the European Open Science Cloud (EOSC) ecosystem and made openly available for future inquiry by researchers and the public alike.

The workflow will be designed for adaptation by researchers, archives, and cultural institutions working with legacy or sensitive oral history collections. It will integrate established Open Science models (TEI/XML, DCMI, CMDI, RDF) and tools (OpenAIRE Amnesia, CLARIN's TranscriptionPortal, INCEpTION) and make use of a curated selection of the corpus, which will be digitized (text and audio), annotated, anonymised, and semantically enriched, as an example of its application to analog oral history collections.

The workflow and dataset will be documented and published on federated repositories connected to the EOSC such as the University of Bologna's AMS Historica and the Italian node of CLARIN, CLARIN-IT, as well as made available via the SSHOC Marketplace. Additional dissemination will be ensured via the creation of an interactive web interface for the dataset, acting as a showcase for how analog oral history media can be opened up to the wider public.

Digital Humanist Role

Within the FAIR Memories project the Digital Humanist will act as project manager and technical specialist, working with two domain experts to fulfill the goals of the project, with a particular focus on the following:

- Providing technical support and expertise for the creation and documentation of the workflow.
- Evaluating and implementing the FAIR-ness of the workflow and example dataset.
- Evaluating the usefulness and appropriateness of Machine Learning tools and systems for use in the workflow and FAIR-ification of the data.
- Selecting existing standards to use for the data model, based on their relevance, usability, and integration with existing EOSC infrastructures.
- Implementation, use, and training for the tools to be used as part of the project, in particular with regards to encoding and semantic enrichment of the transcripts.

- Collaborating with other Unibo teams (including ADlab and AMS Acta) who are providing in-kind support for digitization and dissemination, including evaluation of digitized audio and text and preparation of assets for ingestion in EOSC.
- Supporting domain experts in collaborating with external entities including CLARIN-IT, ASMI (Association for the Study of Modern Italy), and SIS (Society for Italian Studies).
- Implementation of the web interface for the dataset's access by the public, including evaluation of existing solutions such as OMHS and Omeka.
- Supporting domain experts in the production of a final toolkit, delivery of two workshops, and writing of scientific articles.

Skills & Experience

- Tools and platforms for creating and managing Digital Humanities projects such as GitHub, Omeka, Scalar, Jekyll, ArcGIS, Research Space.
- Web dev and programming languages such as HTML, CSS, JavaScript, Python, R, Jupyter Notebooks, tools for working with Machine Learning.
- Audio editing tools such as Adobe Audition.
- Structured data formats and encoding standards such as JSON, XML, TEI, CSV, relational databases.
- Metadata standards such as RDF, DCMI, CMDI, Europeana Data Model.
- FAIR data, Data Modelling, Linked Open Data, and semantic enrichment best practices.
- Oral History best practices and guidelines.
- Project Management tools and best practices.
- Training and instruction of non-technical staff and public workshop delivery.
- Knowledge of current trends and resources in Digital Humanities and Archival Sciences.
- Fluency in English and Italian.

Digital Humanities specialist for FAIR memories project (ita)

Sintesi del progetto

Il progetto FAIR Memories mira a valorizzare e rendere riutilizzabili nel lungo periodo gli archivi analogici di storia orale attraverso la creazione di un workflow riutilizzabile per la loro digitalizzazione e FAIR-ificazione, accompagnato da un dataset di esempio basato su un corpus unico di interviste sulla cultura dei consumi nell'Italia dei primi decenni del XX secolo. Trasformando questo corpus legacy in risorse digitali conformi ai principi FAIR, il progetto intende dimostrare come i risultati di ricerche passate possano essere integrati in modo etico nell'ecosistema della European Open Science Cloud (EOSC) e resi liberamente disponibili per future ricerche, sia da parte della comunità scientifica sia del pubblico.

Il workflow sarà progettato per essere adattabile da ricercatori, archivi e istituzioni culturali che lavorano con collezioni di storia orale legacy o sensibili. Esso integrerà modelli consolidati di Open Science (TEI/XML, DCMI, CMDI, RDF) e strumenti (OpenAIRE Amnesia, TranscriptionPortal di CLARIN, INCEPTION) e farà uso di una selezione curata del corpus, che verrà digitalizzata (testo e audio), annotata, anonimizzata e semanticamente arricchita, come esempio di applicazione alle collezioni di storia orale analogiche.

Il workflow e il dataset saranno documentati e pubblicati su repository federati collegati all'EOSC, come AMS Historica dell'Università di Bologna e il nodo italiano di CLARIN, CLARIN-IT, e resi disponibili anche tramite lo SSHOC Marketplace. Un'ulteriore diffusione sarà garantita dalla creazione di un'interfaccia web interattiva per il dataset, che fungerà da vetrina per mostrare come i media analogici della storia orale possano essere aperti a un pubblico più ampio.

Ruolo del/della Digital Humanist

All'interno del progetto FAIR Memories, il/la Digital Humanist agirà come project manager e specialista tecnico, collaborando con due esperti di dominio per il raggiungimento degli obiettivi del progetto, con particolare attenzione ai seguenti aspetti:

- Fornire supporto tecnico e competenze per la creazione e la documentazione del workflow.
- Valutare e implementare il grado di FAIR-ness del workflow e del dataset di esempio.
- Valutare l'utilità e l'adeguatezza di strumenti e sistemi di Machine Learning da impiegare nel workflow e nella FAIR-ificazione dei dati.
- Selezionare gli standard esistenti da utilizzare per il modello dei dati, in base alla loro rilevanza, usabilità e integrazione con le infrastrutture EOSC esistenti.
- Implementare e/utilizzare strumenti da impiegare nel progetto, in particolare per quanto riguarda la codifica e l'arricchimento semantico delle trascrizioni.
- Collaborare con altri gruppi Unibo (inclusi ADlab e AMS Acta) che forniscono supporto in kind per la digitalizzazione e la disseminazione, inclusa la valutazione di audio e testi digitalizzati e la preparazione delle risorse per l'ingestione nell'EOSC.

- Supportare gli esperti di dominio nella collaborazione con enti esterni, tra cui CLARIN-IT, ASMI (Association for the Study of Modern Italy) e SIS (Society for Italian Studies).
- Implementare l'interfaccia web per l'accesso pubblico al dataset, inclusa la valutazione di soluzioni esistenti come OMHS e Omeka.
- Supportare gli esperti di dominio nella produzione di un toolkit finale, nell'organizzazione di due workshop e nella redazione di articoli scientifici.

Competenze ed esperienze richieste

- Strumenti e piattaforme per la creazione e la gestione di progetti di Digital Humanities, quali GitHub, Omeka, Scalar, Jekyll, ArcGIS, Research Space.
- Web development e linguaggi di programmazione come HTML, CSS, JavaScript, Python, R, Jupyter Notebooks, nonché strumenti per il Machine Learning.
- Strumenti di editing audio come Adobe Audition.
- Formati di dati strutturati e standard di codifica come JSON, XML, TEI, CSV, database relazionali.
- Standard di metadati quali RDF, DCMI, CMDI, Europeana Data Model.
- Dati FAIR, data modelling, Linked Open Data e buone pratiche di arricchimento semantico.
- Buone pratiche e linee guida per la storia orale.
- Strumenti e buone pratiche di project management.
- Formazione e istruzione di personale non tecnico e conduzione di workshop pubblici.
- Conoscenza delle tendenze e delle risorse attuali nelle Digital Humanities e nelle scienze archivistiche.
- Padronanza dell'inglese e dell'italiano.