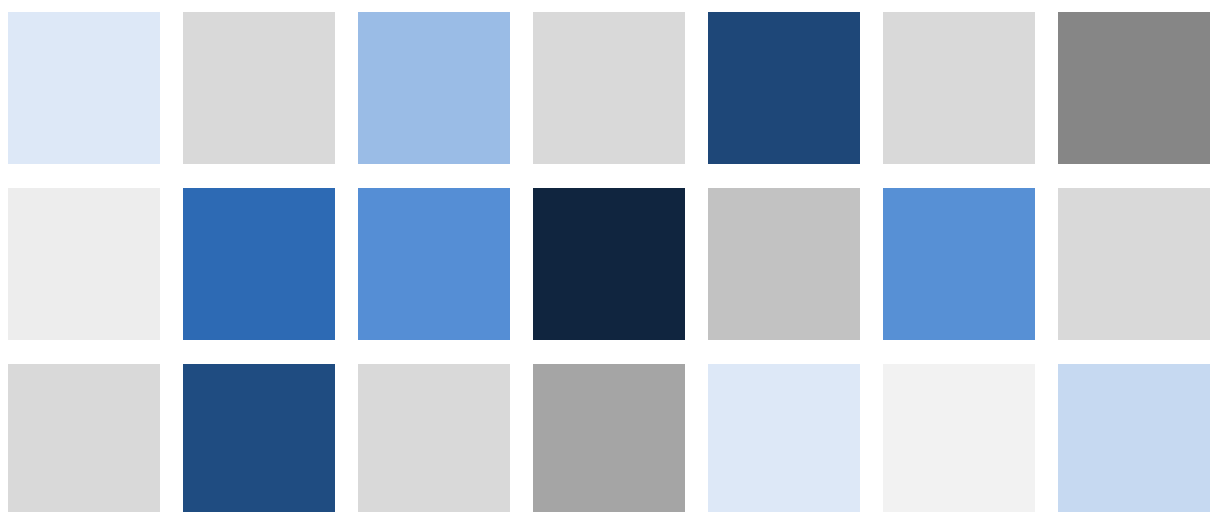


Long-term data for Europe

EURHISFIRM

M1.2: Position paper (midterm)



AUTHORS:

Jan ANNAERT (UNIVERSITEIT ANTWERPEN)

Wolfgang KÖNIG (JOHANN WOLFGANG GOETHE-UNIVERSITÄT FRANKFURT AM MAIN)

Angelo RIVA (ÉCOLE D'ÉCONOMIE DE PARIS)

Lana YOO (ÉCOLE D'ÉCONOMIE DE PARIS)

APPROVED IN 2019 BY:

Jan ANNAERT (UNIVERSITEIT ANTWERPEN)

Wolfgang KÖNIG (JOHANN WOLFGANG GOETHE-UNIVERSITÄT FRANKFURT AM MAIN)

Angelo RIVA (ÉCOLE D'ÉCONOMIE DE PARIS)

Table of contents

I. Project objectives and vision 4

II. Work accomplished thus far to realise the objectives and vision 5

III. Conclusions and the way forward 7



I. Project objectives and vision

The EURHISFIRM project designs a sustainable and accessible Research Infrastructure (RI) to deliver reliable, comprehensive, and standardised long-term historical European financial data based on open science principles. The project envisions designing an RI assured in its **long-term self-sustainability; scientific excellence;** and the **promotion of European communities in science, research technologies, and research infrastructures.**

To accomplish these objectives, the project incorporates:

- 1) *Research excellence* in Pan-European economic and financial history to ensure scientific reliability and accuracy of the database and data interaction designs;
- 2) *The application of state-of-the-art technologies and modelling* to enable the transformation, standardisation, and storage of historical paper sources into databases based on advanced artificial intelligence (AI) and self-adaptation techniques to accommodate current and future heterogeneous data, as well as FAIR data principles (findability, accessibility, interoperability, reusability);
- 3) *A sound business, cultural valorisation, and legal plan* for the RI's long-term financial and practical sustainability and to design an RI that will best serve future users' interests (academics, policy makers, firms and financial intermediaries, the general public/citizens);
- 4) *Active involvement in European research infrastructure developments* in order to integrate the latest European research technologies and to exploit increasing network effects in Europe;
- 5) *Building a scientific community based on open science principles* to encourage cross-collaboration among researchers in order to enrich scientific contributions and advancements in the social sciences.

The motivations behind these objectives are three-fold. Firstly, the lack of usable historical European financial data (which currently mostly exist in archived, paper-bound books) substantially hinders the creation of long-term empirical economic and financial models. Such models are essential to an in-depth understanding of the modern European economy, taking duly into account its rich diversity and the important acknowledgements of the occurrence of rare but significant events. The importance of incorporating these elements—which current models mostly lack—has demonstrated itself not long ago with the arrival of the Great Recession following the 2008 financial crisis. The lack of long-term empirical economic models creates significant barriers to understanding the causes of and solutions to such dramatic events. To further illustrate, the 2008 Financial Crisis began in the mortgage-backed security (MBS) sector, where risk models went astray because they were calibrated on only five to ten years of historical data, taken from a very benign period. Moreover, macro-financial models failed to take into account several frictions: in the period over which these models were estimated/calibrated, such frictions had not manifested themselves. Secondly, even if adequate data are made available, scientific research with a European scope cannot make notable advancements within isolated silos; cross-collaboration among Pan-European research communities based on open science principles is crucial. Also, network effects will stimulate national teams to contribute to the RI and to share data with the research community. Thirdly, superior research technologies and infrastructures serve as catalysts for scientific

advancements by optimizing the necessary tools for the creation, cleaning, distribution, and storage of scientific data.

II. Work accomplished thus far to realise the objectives and vision

For further details on the progress accomplished so far, please refer to the project Deliverable 1.5: Project midterm report.

1) *Research excellence*

To ensure that the developed technology properly accommodates the data that it will handle, Work Package 4, along with the project's other economic historians, worked on cataloguing EURHISFIRM's data sources and established a common standard of documentation for current and future data. In Work Package 7, the economic historians work with the technical developers to ensure that the tools developed can process the data elements correctly and can take into account historical nuances specific to each country and/or time period. This procedure includes verification of the test results from the tools developed and producing reference documents with regards to the sources being treated. Six sources in total (price lists from three markets (Belgium, France, Germany) and yearbooks from three markets (France, Germany, Spain)) have been chosen as the representative samples for this work based on the existing or potential capacities for digitalisation in the sources' current states.

2) *State-of-the-art technologies and modelling*

The technical Work Packages of the project are Work Packages 5, 6, 7, and 9, working on common data modelling, data connecting and matching, data extraction and enrichment system, and infrastructure policy and architecture, respectively. The official deliverable deadlines for most of these Work Packages begin on or after the midterm, but the work for these tasks have been in progress. Work Package 5 has studied the models of existing database models and describes key design elements to be considered for the EURHISFIRM common data model; it also established a methodology for common data model evaluation. It coordinates the Working Group on Identification and Standardization (WGIS) to increase the communication and collaboration between the different work packages to facilitate the implementation of technical standards in the project. Work Package 6 has also recently begun its work in data connection and linkage among the independent databases sources, in line with the project timeline. This work will logically use the consortium's most advanced databases (the Brussels and Paris financial databases) as a first test case by linking them and testing their abilities to retrieve the queried data in a synchronised and accurate way. Work Package 6 has already started discussions with other relevant institutions regarding further test cases. Work Package 7 has begun developing artificial intelligence-based systems to recognise the tabular structures and the texts of the data sources through collaboration with the consortium's economic historians. It has also begun to conduct tests on the tools developed in order to validate and improve the results, and web linking tools (to connect the company entities in the databases to the data existing in the wider world wide web) have begun to be evaluated under task 7.4. Work Package 9—while it is officially scheduled to start in the third quarter of 2019—has begun to investigate the architectural framework of the project by



investigating materials from existing European research infrastructures and participating in related technical discussions with other Work Packages.

3) *A sound business, cultural valorisation, and legal plan*

Work Package 3 has begun its work concerning ownership and property rights, especially in the context of open science, which will address the overall legal recommendations, issues, constraints, and solutions concerning the EURHISFIRM infrastructure design in all of its aspects (including but not limited to data rights, user and access rights, privacy laws). Work Package 8, in charge of understanding the infrastructure's potential interaction with users and identifying the target users and their needs, has conducted quantitative and qualitative studies with relevant subject pools. Work Package 10's core work in business model and governance will begin in the third quarter of 2019, but it has nonetheless already made progress in defining alternative business model concepts, as well as conducting a preliminary assessment of business and governance model alternatives. Work Package 11's work in evaluating alternative strategies to use digitized material (from historical sources) for the promotion of European cultural heritage will also begin in the third quarter of 2019, and preparatory work has been in progress to ensure a solid start to these tasks.

4) *Active involvement in European research infrastructure developments*

In order to stay current with the European research infrastructure technologies and platforms, EURHISFIRM has formed partnerships with various organisations in the European RI ecosystem and envisions further collaborations with relevant institutions in the future. Huma-Num (<https://www.huma-num.fr/>, the French services provider of Digital Research Infrastructure for the Arts and Humanities (DARIAH)), collaborates with EURHISFIRM by hosting the project website and the project tools such as secure internal file sharing, and video-conferencing and instant messaging tools. EURHISFIRM has also joined the Social Sciences and Humanities Open Cloud (SSHOC) as part of the European Open Science Cloud (EOSC) programme (in which Huma-Num also participates). Indeed, EURHISFIRM was invited to join SSHOC by CESSDA (Consortium of European Social Science Data Archives), an incoming new member of the EURHISFIRM consortium. This partnership will further solidify EURHISFIRM's dedication to FAIR data principles and ensure mutual benefit from the progress made within the European data community. The Executive Committee, along with the Steering Committee and Work Package 1, drive these partnerships as well as the project strategy and direction (including compliance to open science frameworks and FAIR data protocols through a continuously updated Data Management Plan maintained in Work Package 1).

5) *Building a scientific community based on open science principles*

Work Package 2 leads the project's communication and outreach tasks, including the maintenance of the project communication tools (website, social media), the project identity and logo, and organisation of the project's annual General Assembly meetings. The community building task, concerning Work Packages 1 and 2, is indeed a project-wide effort from all consortium members and a key part of EURHISFIRM's overall ambitions. EURHISFIRM further aims to concretise and expand upon these community-building efforts by applying for the Horizon 2020 INFRAIA-02-2020 call (Integrating Activities for Starting communities) to further expand the community systematically. In the spirit of open science, EURHISFIRM publishes all of its public deliverables on OpenAIRE

(https://explore.openaire.eu/search/project?projectId=corda_h2020::612830f55f1f92d36a5477538163d4e5) and on the project website (<https://eurhisfirm.eu/>).

III. Conclusions and the way forward

The EURHISFIRM project has reached its midterm period. During the first 1,5 years of the project, the project has completed its tasks according to the project timeline: creating an inventory of the source data, deciding on metadata standards, studying the existing common data models and the potential models for EURHISFIRM, starting the development of the tools for the data extraction and verifying the test results with economic historians, beginning the research on the legal issues and business plan models, establishing a communication plan and project identity and a website, and creating preliminary models for logistical and project management foundations. EURHISFIRM has also advanced in its efforts to become more actively involved in European research infrastructure developments and in starting to build a scientific community based on open science principles in the research infrastructure topic, especially in the social sciences and humanities. For the second term, EURHISFIRM will continue developing and will complete these tasks according to the project timeline.

Although EURHISFIRM is a design study and is in its preliminary stages, it is necessary to keep in mind some key questions on the project's practical issues in order to ensure that the design is consistent with the project's long-term sustainability:

- ▶ What will be the expected final costs of operating the infrastructure once the comprehensive system design is completed?
- ▶ How shall the steps in the data processing (data acquisition, collation, standardisation, sharing, storing) be organised at the national and European levels? What will be the expected costs per step and per level?
- ▶ How shall the future EURHISFIRM system be operationally organised (governing bodies; operational, maintenance, and research personnel)?

While these questions cannot be answered in the current stage, they will serve as important guideposts towards important issues that will be considered as the project progresses.