

**INSA Rennes recruits  
a post-doctoral researcher / engineer in computer science**

**Analysis systems for serial sources in collections of historical image documents**

**IRISA - Intuidoc**

IRISA is a joint research center for Informatics, including Robotics and Image and Signal Processing. 850 people, 40 teams, explore the world of digital sciences to find applications in healthcare, ecology-environment, cyber-security, transportation, multimedia, and industry... INSA Rennes is one of the 8 trustees of IRISA.

The Intuidoc team (<https://www.irisa.fr/intuidoc>) conducts researches on the topic of document image recognition. Since many years, the team proposes a system, called DMOS-PI method, for document structure analysis of documents. This DMOS-PI method is used for document recognition, or field extraction in archive documents, handwritten contents damaged documents (musical scores, archives, newspapers, letters, electronic schema, ...).

**EURHISFIRM project**

EURHISFIRM European project aims at developing a research infrastructure to connect, collect, collate, align, and share reliable long-run company-level data for Europe to enable researchers, policymakers and other stakeholders to analyze, develop, and evaluate effective strategies to promote investment and economic growth. To achieve this goal, EURHISFIRM develops innovative tools to spark a “Big data” revolution in the historical social sciences and to open access to cultural heritage.

EURHISFIRM is a project funded by the European Commission within the Infrastructure Development Program of Horizon 2020. The first phase of the Infrastructure Development Program lasts for three years. It aims at developing an in-depth design study of the Research Infrastructure. After this phase, Development and Consolidation Phases follow if further applications will be successful. EURHISFIRM brings together eleven research institutions in economics, history, information technologies and data science from seven European countries.

**Position to be filled**

Position: Post-doctoral fellow or engineer

Time commitment: Full-time

Duration of the contract: Up to 18 months (March 1st 2020 – August 31st 2021)

Supervisors: Bertrand Coüasnon and Aurélie Lemaitre

Indicative salary: Up to €36 000 gross annual salary (according to experience),  
with social security benefits

Location: IRISA - Rennes

**Missions**

The post-doctoral fellow will join a team working on EURHISFIRM workflow. The goal is to extract information from images of financial documents from the 20<sup>th</sup> century. We mainly focus on two collections: yearbooks, which describes the companies and their administrators, and prices lists, which are newspapers showing the daily stock prices.

Due to the large variety of those documents, it requires a flexible and easy-to-adapt document recognition system. It is based on a modelling of knowledge not only at the page level but also at the collection level in interaction with experts of the historical sources.

A first system has been designed, using DMOS-PI method. It uses a grammatical language, EPF (Enhanced Position Formalism), to describe a general page layout, with perceptive vision mechanisms, and an iterative analysis. The system also combines structural method with Deep Learning. For new collections, an adapted description of the document layout must be developed. This has to be done on a large range of structure levels: from very structured pages like table structures from stock exchange lists, up to a paragraph-oriented structures from yearbooks.

Some first experiments have been led on the recognition of “price lists” documents, on a specific French corpus. The objective of the work will be to generalize this system to other collections of price lists, from other countries. This requires to identify which are the common parts of price lists from each stock exchanges and countries, and how to make the system able to be adapted to a new collection in an easy way.

### Main Skills

PhD or Master degree or Engineering degree in computer science.

Experience in document recognition, statistical analysis or deep learning.

Fluent English

Skills in grammars and languages and/or logical programming are nice-to-have.

Candidates should contact via email: [bertrand.couasnon@irisa.fr](mailto:bertrand.couasnon@irisa.fr), [aurelie.lemaitre@irisa.fr](mailto:aurelie.lemaitre@irisa.fr)