

AGIR Project Proposal

2025-2026

Title (titre):

Provide a title of your project / Donnez le titre de votre projet

Natural Language Querying of Historical Military Records via Knowledge Graphs and RAG

Partner/Client (partenaire/client):

Provide the name of the partner/client of the project (association, company, community, research lab, department, etc.) / Donnez le nom du partenaire/client de projet (association, entreprise, communauté, laboratoire de recherche, département, etc.)

Diana Nurbakova, LIRIS diana.nurbakova@insa-lyon.fr

Killian Barrere, LIRIS killian.barrere@insa-lyon.fr

Dimensions (dimensions):

TECHNIQUE / SOCIETALE / SCIENTIFIQUE / METHODOLOGIQUE / ARTISTIQUE / AUTRE

Highlight the dimensions of your project. Several dimensions can be specified. If you select "Other" (AUTRE), please specify / Soulignez les dimensions de votre projet. Plusieurs dimensions peuvent être spécifiées. Si vous sélectionnez "Autre", veuillez préciser

Summary including Objectives and Deliverables (résumé y compris les objectifs et les livrables):

Provide a short summary of your project clarifying the main objectives and deliverables / Ecrivez un court résumé du projet, en précisant les objectifs et les livrables attendus

This project develops an intelligent information retrieval system for 18th-century French military registers, enabling researchers to query genealogical, demographic, and career data using natural language. We address the challenge of validating such systems before costly manual data extraction by generating synthetic datasets that replicate the structure and relationships found in historical military documents (birth records, parentage, enlistments, promotions, migrations).

The technical approach combines: (1) rule-based synthetic data generation with historical constraints, (2) knowledge graph construction capturing entities (persons, places, military units) and relationships (parentage, service records, geographic movements), and (3) a hybrid query system that translates natural language to graph queries (Cypher) while leveraging RAG techniques for complex exploratory queries.

This framework enables early validation of architecture choices and query translation accuracy before investing in large-scale handwritten text recognition and data extraction.

References (références):

(If known) Provide references including articles, websites, software, etc. / Donnez la liste bibliographique y compris les articles, sites web, logiciels, etc.

Project Team (groupe-projet):

(If known) List people involved in the project / Donnez les noms de personnes impliquées dans le projet.

Provide their affiliation / Donnez leur affiliation.

Give the name of the team coordinator (contact person) and specify if the team is open for new members and if yes, how many / Donnez le nom de personne qui porte le projet et précisez si l'équipe est ouverte à accueillir de nouveaux membres et si oui, combien.