openEHR Fellowship Programme

Project Brief Proforma

1. Project Title

A clear and concise title for your fellowship project

Development of an Open Source openEHR Archetype Exploration, Mapping and Planning Tool

2. Fellow Information

Full Name:	Martin Andreas Koch
Email Address:	
Contact Number (Optional)	
Current Affiliation/ Institution (if any	CatSalut. Catalan Health Service
Country of Residence	Spain

3. Project Summary/Abstract (max 300 words)

Provide a concise overview of your project.

The challenge of creating openEHR-based data structures lies in the necessity for users to identify existing, published archetypes while also determining which new archetypes need to be developed and proposed to the internationally recognized archetype repository.

The planning of a data structure is difficult, due to the existing relationships between archetypes (parent-child relationships or inclusions/exclusions) and their contents. Furthermore, the search for relevant archetypes is limited by the functionalities of repositories used to date.

To address these issues, a proposed solution is the development of an open online tool designed to enhance the user experience in managing archetypes.

This tool would feature a fast search function for archetypes, enabling users to quickly locate relevant structures. Additionally, it would provide visualization of the relationships between archetypes, such as parent-child relationships, inclusion, or exclusion, thereby facilitating a better understanding of their interconnectivity. The tool would also suggest similar or related archetypes, allowing users to explore options more effectively.

The tool would enable users to create collections of archetypes and plan for new archetypes, generating a report that would help users easily identify needed archetypes.

The expected key outcomes of this project include the creation of an open-source online tool that is accessible to anyone interested in openEHR. The tool would improve the overall efficiency of the archetype development process, which will be showcased by real use cases.

4. Introduction and Background (max 500 words)

Problem Statement

The current process for finding archetypes for a planned openEHR data structure is limited by the main repository used. Even though there is a capable search function in the repository, the search results are not easily parsable and the exploration of the archetype ecosystem is cumbersome.

Users don’t have a comprehensive overview of the existing archetypes, and it is very difficult to create a plan for integrating archetypes and the required data elements into a data structure. Manually identifying and managing dependencies constitute a significant effort for the users.

Also, due to the vast number of archetypes available, it is very difficult for a user to be aware of all existing archetype options. Consequently, users find it sometimes easier to create new archetypes from scratch rather than searching for and integrating those from the official repository.

This diminishes the potential of the official archetype ecosystem by hindering the efficient planning and reuse of archetypes, and limits collaboration and standardization across modelling efforts.

A solution that improves the discoverability and visualization of existing archetypes would empower users to plan more effectively, reuse validated components, and contribute more confidently to the shared resource base.

This project aims to address these challenges by streamlining the search process, improving dependency management, and encouraging broader engagement with the official archetype ecosystem.

Relevance to openEHR

This project directly supports the goals and principles of the openEHR community by being open source and focused on improving the modelling of openEHR archetypes.

By simplifying the discovery and integration of existing archetypes, it promotes reuse and collaboration—key tenets of the openEHR approach.

The tool will help modellers work more efficiently and accurately, reducing redundancy and improving consistency across archetypes.

Ultimately, this contributes to the overall quality of clinical models, reinforcing openEHR's mission to enable high-quality, interoperable health information systems in the service of medicine.

By lowering technical barriers and encouraging broader participation, the project also strengthens the community's capacity to develop and maintain robust, clinically meaningful archetypes.

Existing Work/Context

The project aims to enhance the search for archetypes and develop a comprehensive data structure that integrates these archetypes effectively.

Current efforts focus on search engines like the Clinical Knowledge Manager (CKM) and various mapping tools for openEHR data structures, including template designers and mind mapping applications.

However, this project distinguishes itself by offering rapid search capabilities, visualizing dependencies, and guiding users from an initial checklist to a finalized openEHR data structure design document.

Notably, no existing tool encompasses these features, highlighting the unique value of the proposed project. We hope to significantly improve the efficiency of archetype re-use.

5. Aims and Objectives

Overall Aim

The main objective of this project is to create a searchable, dynamic visualization tool that empowers openEHR modellers to explore archetypes, discover the need for changes or new archetypes, and create a plan for their data structure to be developed.

Specific Objectives

1. Define the specific needs of users/developers/modellers through an initial questionnaire
2. Develop a tool capable of creating a data file from a collection of *.ADL files containing archetype information. This tool will work with all archetypes published in the CKM International, CKM CatSalut, CKM Arjketyper and CKM Highmed.
3. Design and implement a web-based tool featuring search capability and interactive visualization for openEHR archetypes, including all archetypes from the CKM International.
4. Implement a recommendation engine that suggests related or similar archetypes based on semantic tags.
5. Document user feedback on the application of the tool in 2-3 real use cases.

6. Methodology/Approach (max 700 words)

Methodology/Approach

Methodology

Research, Development, and Implementation Strategies

The primary objective of this project is to create a stable working prototype that effectively demonstrates value to the end user. To achieve this, the following methodology for research, development, and implementation is going to be used.

Research

The research phase involves conducting a questionnaire aimed at assessing the need for the proposed project and identifying users' pain points. This step is crucial

as it helps to understand the specific requirements and challenges faced by potential users, ensuring that the final tool is tailored to meet their needs.

Development

The development process is divided into two main parts:

- **Part 1:** This part focuses on creating a tool designed to acquire archetype ADL files from either a local or online repository. The tool will extract relevant data and store it in a data file, facilitating easy access and management of the information.
- **Part 2:** The second part consists in developing an online search, visualization and planning tool. This tool will enable users to explore the previously generated data file, providing a user-friendly interface for data interaction and analysis.

Implementation

The implementation phase includes several key activities:

1. **Extraction of Archetypes:** Archetypes can be extracted from online resources and local zip files. The online repositories include CKM or GitHub repositories. This flexibility ensures that the tool can be applied to a wide range of archetype sources.
2. **Development of the Extraction Tool:** The extraction tool will be developed in Python and compiled into an executable (.exe) file with a graphical user interface. Additionally, a "headless" mode is going to be implemented for an automatic update of the resulting data file. Both the source code and the executable will be published in a GitHub repository under the Apache 2.0 open-source license, making them freely available to the community.
3. **Search, Visualization, and Planning Tool:** This tool will be accessible to any user with an internet connection and a modern browser. HTML and JavaScript are going to be used to develop a user-friendly interface. Data visualization is going to be implemented with d3.js, a widely-used library with comprehensive documentation available at https://d3js.org/. The tool will also be published in a GitHub repository under the Apache 2.0 open-source license.
4. **User Data and User Management:** All data that is used by the user is going to be saved on their local hard drive. No user data is going to be collected or sent to third parties. There are no user accounts or user management planned.
5. **User-Friendly Data Management:** A special effort is going to be made to ensure that data input and output are as user-friendly as possible. Users will be able to upload data in easily manageable formats, such as .xlsx, simplifying the process of data interaction.
6. **Maintenance and Transition:** Both tools will be maintained in a public GitHub repository by Martin Koch from the Catalan Healthcare Service during the duration of the project. Upon completion of the project, there is potential for the content to be transitioned to an official openEHR domain. The future

development and maintenance has to be discussed once the value of the project has been assessed.

Engagement with openEHR Specifications

This project aligns with the openEHR specifications by using and parsing archetype files that comply with the ADL standards.

By adhering to these specifications, we ensure that our tools are compatible with the existing standard and can be effectively utilized within the broader openEHR community.

Furthermore, the tools will utilize REST API calls to various CKM sources for retrieving archetypes.

Specific openEHR Tools, Platforms, or Clinical Modelling Approaches

The primary source of the archetypes integrated into this project will be the international CKM.

The tools to be developed are going to integrate into the common modelling approach, where a list of data elements is given in the beginning, followed by a search for archetypes and planning a layout and finally creating a template. The main focus here lies on the searching and planning.

Quality Assurance

To assure a good quality of this project, the robustness of the data extraction tool will be rigorously tested against all available archetypes in the CKMs, including those from international, Norway, Catalan, and German sources. This comprehensive testing approach will help identify and rectify any issues, ensuring that the tool performs reliably across various scenarios.

Additionally, the quality of the visualization tool will be ensured through user feedback in real use cases. This feedback will be invaluable in refining the tool's functionality and user experience.

7. Scope

In scope

- Development of a data extraction tool for a collection of archetypes in ADL 1.4 format.
- Creation of an online search, visualization, and planning tool.
- Both tool are going to be stable and usable prototypes.
- Online hosting of the project throughout the duration of the fellowship.
- Archetype data is retrieved and generated periodically (e.g. once a day or once a week).
- Only applicable to ADL 1.4

- All user data is saved locally.
- Data extraction in english or the original language if english not available.

Out of Scope

- This project will not result in a finished product.
- There will be no creation of fully “fleshed-out” new openEHR archetypes.
- No development of openEHR templates.
- The project will not provide one comprehensive tool that combines extraction with search, visualization, and planning.
- Hosting and maintenance of the tools is not guaranteed beyond the conclusion of the fellowship.
- No real-time connection to a CKM or other archetype repository.
- No user account management.
- No ADL 2.0 compatibility.
- Does not support multiple languages

8. Deliverables

Specific Objectives

1. Executable and source code of the data extraction tool
2. JS file with the extracted archetype data
3. HTML, JS and CSS files for the search, plan and visualization webpage
4. Full project documentation
5. Initial user feedback
6. Final user feedback and real use cases

9. Timeline/Work Plan (approximate)

Quarter 1 (Months 1-3)

Key Activities: e.g., Literature review, Initial tool setup, Stakeholder interviews (if any), Familiarisation with relevant openEHR specifications	• Design and distribute initial questionnaire. • Conduct stakeholder interviews via questionnaire (developers, clinicians, etc.). • Analyze interview responses to extract needs and pain points. • Define features for both tools (must-have, nice-to-have). • Research existing tools/libraries for parsing ADL files. • Research existing archetype browser o searching tools • Develop initial version of the ADL extraction tool. • Define data model and schema for archetype representation.
Deliverables for this Quarter: e.g., Draft problem definition, Initial list of archetypes to review/develop, Setup of development environment	• Full feature list (categorized: essential / optional). • Stakeholder interview analysis report. • Technical specification document for the extraction tool. • Architecture overview of the extraction tool (inputs/outputs/data flow).
Milestones: e.g., Completion of literature review, First draft of X submitted.	• Stakeholder needs assessment complete. • First working prototype of ADL extractor (alpha).

Quarter 2 (Months 4-6)

Key Activities: e.g., Archetype modelling, Prototyping, Initial data mapping, First round of community feedback solicitation	• Finalize and publish the ADL extraction tool. • Design and begin implementation of the web-based visualization/search/planning tool. • Define UX/UI for visualization/planning tool. • Set up service for updating parsed archetypes data file automatically.
Deliverables for this Quarter: e.g., First set of draft archetypes for review, Basic prototype of Y feature, Report on initial findings and feedback	• Fully functioning archetype extractor. • Documentation for the ADL extractor. • GitHub repository with source code. • Wireframes/mockups for visualization tool. • Basic front-end prototype for searching and visualizing archetypes.
Milestones: e.g., Completion of first archetype set (draft),	• Extraction tool publicly released. • Search tool can manage archetype and project information.

Functional prototype available for internal review	• Technical specification document for the search and visualization tool.

Quarter 3 (Months 7-9) (if applicable)

Key Activities: e.g., Refining models based on feedback, Developing educational materials, Testing application components	• Finalize visualization/planning tool. • Launch test version of the visualization/planning tool. • Recruit pilot user group (developers, clinicians) for user feedback. • Collect use cases for archetype creation and search. • Design and distribute usability questionnaire. •
Deliverables for this Quarter: e.g., Revised archetypes, Draft educational materials, Test plan and initial results	• Test version of the visualization and planning tool. • User guide and instructions. • Use case compilation document. • User feedback report from pilot test.
Milestones: e.g., Archetypes submitted to CKM, Educational materials reviewed	• Searching/visualizarion tool released for testing. • Real-world use cases documented. • User experience feedback collected and analyzed.

Quarter 4 (Months 10-12) (if applicable)

Key Activities: e.g., Finalizing deliverables, Writing final report/paper, Preparing dissemination materials	• Implement final improvements based on user feedback. • Finalize full documentation. • Announce final public release. • Conduct final presentation/demo for stakeholders.
Deliverables for this Quarter: e.g., Final project report, Completed software/tool (if applicable), Published archetypes, Dissemination materials	• Finalized visualization and planning tool. • Complete documentation (installation, user manual, technical guidevisualization). • Summary report of project progress, outcomes, and lessons learned. • Final code repositories.
Milestones: e.g., Project completion, Final presentation delivered, All deliverables submitted	• All user feedback processed. • Tools packaged and documented. • Project completed and reviewed with stakeholders.

10. Resources Required

Software/Tools

Programming/developing:

- Python IDE (have access)
- HTML/Javascript/CSS IDE (have access)

Publication and website hosting:

- Github/Github Pages (Have access, but it is a personal account. Depending on the philosophy of the openEHR ecosystem it might be necessary to grant me access to an openEHR owned account.)

Mentorship/Support

If needed, it would be helpful to have access to a mentorship on the ANTLR4 parsing of ADL files.

Other

NA

11. Expected Outcomes and Impact

For the Fellow

- Deeper familiarity with the openEHR ecosystem, including ADL syntax, and archetypes.
- Understanding of the different user groups involved in archetype modeling (e.g., clinicians, informatics, developers) and their respective workflows and priorities.
- Experience with stakeholder engagement, user research, and requirements gathering.
- Development of practical skills in software design, including user interface design, project management and iterative development.
- Insight into data visualization as a tool for knowledge representation.

For the openEHR Community

- A searchable, visual representation of archetypes that enhances understanding for both novice and experienced modelers.
- A planning tool that supports better coordination and documentation when creating new archetypes, potentially reducing duplication and inconsistencies.
- Open-source tooling that the community can adopt, extend, or integrate into existing modeling environments (e.g., CKM, Archetype Designer).
- Contribution to the knowledge base by collecting real-world use cases, modeling patterns, and user feedback.

Potential Wider Impact

- The presented tools could be used as a training platform for new users (clinicians, students, data modelers) to learn about archetype modeling through exploration and visualization.
- The tools can serve as an example that might be adapted for other health information standards (e.g., HL7 FHIR profiles).

- Improved tooling could accelerate the development of clinical data models, facilitating more consistent and interoperable health IT systems.
- Potential use in research settings for prototyping and sharing domain-specific models (e.g., in rare diseases, genomics).
- Long-term, the tool may contribute to future repositories or platforms that serve as integrated environments for modeling, validating, and managing clinical information models.

12. Project Presentation Plan

1. Open Sharing of Code and Tools

- **Format:** Source code, executables, version history, README documentation, and issues tracker.
- **Location:** GitHub (Public Repository)
- **Audience:** Developers, contributors, and tool integrators.
- **Purpose:** Enable open collaboration, reuse, and community contributions.

2. User-Focused Documentation

- **Format:** User manuals, quick-start guides, video tutorials, FAQs.
- **Location:** Bundled with GitHub releases and hosted online (e.g., GitHub Pages or ReadTheDocs).
- **Audience:** End users, including clinicians, students, and modelers.
- **Purpose:** Ensure tools are accessible and usable by non-technical users.

3. Development Journal, Community Engagement & Dissemination

- **Format:** Structured development log or blog. Periodic updates during the fellowship. Formal project review and demonstration of final tools.
- **Location:** openEHR discourse.
- **Audience:** Fellow researchers, funders, openEHR community.
- **Purpose:** Share project rationale, progress, challenges, design decisions, and reflections.

13. Ethical Considerations

If your project involves human subjects, patient data, or has other ethical implications, please describe these and how you plan to address them (e.g., ethics approval, data anonymization). If not applicable, state "Not applicable."

Questionnaire:

This project involves the use of a voluntary and anonymous questionnaires to gather feedback from users. The purpose of the questionnaires is to inform the design, content, and functionality of the tools developed in this project.

- **Voluntary Participation:** All participants engage on a completely voluntary basis, with the option to withdraw at any time.
- **Anonymity and Privacy:** No personally identifiable information will be collected. Responses are anonymous and cannot be traced back to individual participants.
- **Data Usage:** Data will be analyzed in aggregate form only and used exclusively for research and development within the scope of this project.

Web-Based Tool:

The web application developed as part of this project adheres to strong ethical principles regarding user privacy and data protection.

- **No Personal Data Collection:** The application does not collect, process, or store any personal data.
- **No Tracking Technologies:** The website does not use cookies, analytics scripts, or third-party tracking services.
- **Local Storage Only:** All information related to the use of the tool is stored locally in the user's browser. This data remains entirely on the user's device and is never transmitted to any server.
- **User Autonomy and Transparency:** The design prioritizes user control, transparency, and minimal data exposure to ensure a secure and privacy-respecting experience.

By implementing these safeguards, the project maintains a high standard of ethical responsibility and respects the rights and privacy of all users.

14. Supervisor/Mentor Information

(if applicable and known at time of application)

Name of Primary Supervisor/ Mentor:	Heather Leslie
Affiliation:	openEHR International, Atomica Informatics
Email:	
Add more if multiple supervisors/ mentors	

15. Declaration

I, Martin Andreas Koch, declare that the information provided in this project brief is true and accurate to the best of my knowledge. I understand the goals of the openEHR Fellowship Programme and commit to undertaking this project with diligence and to sharing the outcomes with the openEHR community.

Signature: _____

Name: Martin Andreas Koch

Date: 10/07/2025