Swansea University
Prifysgol Abertawe

Job Description: Research Officer

Faculty:	**FMHLS**
Department/Subject:	**Health Data Science**
Salary:	**Grade 8: £39,355 to £45,413 per annum together with USS pension benefits**
Hours of work:	**Full time**
Number of positions:	**1**
Contract:	**This is a fixed term position for 12-months**
Location:	**This position will be based at the Singleton Campus**

Main Purpose of Post	1. To design, architect, and deliver VELA (Versatile Extraction & Logic Abstraction): a production-grade Python infrastructure that enables researchers to query billions of health records across multiple UK Trusted Research Environments (TREs) through a unified interface, while automatically integrating domain expertise and generating quality reports to ensure research reproducibility and trustworthiness. 2. Serve as the lead technical architect, transforming a functional R prototype into a robust, multi-component system comprising intelligent query orchestration, knowledge management, and automated quality reporting capabilities. 3. Maintain full technical ownership while working in close collaboration with the project founder (Dr. Marcos Del Pozo Banos) and our network of clinical and data science collaborators.
	4. Design and implement a multi-level abstraction layer that enables researchers to query multiple TREs through a unified Python interface. 5. Build intelligent query orchestration that translates high-level research questions into optimized database operations adapted to different database dialects and execution engines. 6. Abstract away the complexity of heterogeneous data schemas and secure environments, allowing researchers to focus on their science rather than data wrangling. 7. Exhibit technical independence in determining architectural direction and making pragmatic design decisions that balance extensibility with pragmatic delivery. 8. Design and implement a knowledge base system that captures domain expertise from clinical and data science collaborators, including data quality considerations, measurement limitations, and best-practice guidance. 9. Develop logic to match user queries with relevant domain expertise and quality metadata. 10. Engineer automated report generation that assembles contextual quality documentation alongside extracted data, transforming distributed expert knowledge into actionable research guidance. 11. Create workflows that enable domain experts to contribute structured knowledge to the system. 12. Develop robust, production-grade pipelines for interacting with billion-record relational databases, ensuring absolute precision in data extraction, transformation, and harmonization. 13. Implement comprehensive error handling and validation for queries executed in air-gapped environments (secure research environments with no or limited external connectivity). 14. Manage complex interactions with massive relational databases across multiple TREs, ensuring data integrity and reproducibility. 15. Maintain a production-grade codebase adhering to SOLID principles, comprehensive unit testing, and automated documentation. 16. Prioritize First-Time Quality. 17. Prepare comprehensive technical documentation describing developed tools, both for internal secure use and for external open-source publication. 18. Act as the lead technical authority on implementation, using initiative to determine suitable approaches to tackle architectural challenges. 19. Interact professionally with collaborators within the Faculty of Medicine, Health and Life Sciences, national partners (e.g., DATAMIND, UK LLC, DataLoch, NCSR), and industry stakeholders to ensure VELA

	meets cross-institutional requirements. 20. Work in close collaboration with project founder Dr. Marcos Del Pozo Banos to align technical implementation with research vision. 21. Engage with domain experts to formalize knowledge into structured formats that the system can utilize. 22. Act as a representative on technical committees and working groups as required. 23. Keep informed of developments in Trusted Research Environments (e.g., SAIL), Federated Analytics, Common Data Models (e.g., OMOP), and Big Data platforms (e.g., Databricks) to ensure VELA remains state-of-the-art. 24. Proactively contribute to external funding applications, providing technical specifications and architectural roadmaps necessary to support the long-term growth and sustainability of VELA. 25. Maintain all development records and project logbooks to the highest institutional standards. 26. Contribute to the broader research software engineering community through knowledge sharing and best practice dissemination.
General Duties	27. To promote equality and diversity in working practices and maintain positive working relationships. 28. To conduct the job role and all activities in accordance with safety, health and sustainability policies and management systems, in order to reduce risks and impacts arising from the work activity. 29. To ensure that risk management is an integral part of any decision-making process, by ensuring compliance with the University's Risk Management Policy. 30. Any other duties as agreed by the Faculty of Medicine, Health and Life Sciences.
Person Specification	**Essential criteria:** 1. 3+ years of professional software engineering experience with a proven track record of delivering complex technical projects where precision and reliability were paramount. 2. Evidence of taking ownership of software implementation, moving projects from prototype to production-grade systems. 3. Expert-level Python proficiency including object-oriented design, library development (not just scripting), and API development. 4. Deep experience with large-scale relational databases including query optimization, data modelling, and working with SQL databases. 5. System design capabilities: ability to architect multi-component systems with clear separation of concerns. 6. Strong software engineering fundamentals: SOLID principles, comprehensive testing strategies, and documentation practices. 7. Demonstrated ability to work independently and take initiative on complex engineering tasks without constant direction. 8. Ability to work collaboratively with domain experts (e.g., clinicians, data scientists) who may not have programming expertise. 9. Evidence of code documentation and technical writing (e.g., README files, API docs, technical blog posts, or open-source contributions). 10. Self-motivated with pragmatic approach to balancing quality with delivery timelines, comfortable in fast-paced research environments. **Desirable Criteria** 11. Postgraduate qualification (MSc or PhD) in Software Engineering, Computer Science, Data Engineering, or a related quantitative/technical field. 12. Experience with knowledge representation systems, metadata management, or automated documentation generation. 13. Experience working within secure research environments (TREs) or with health data governance frameworks (e.g., SAIL, UK LLC). 14. Familiarity with Common Data Models (e.g., OMOP CDM) or health data standardization. 15. Experience with template-based report generation (e.g., Jinja2, LaTeX) or Big Data platforms (e.g.,

	Spark, Databricks). 16. Background working with domain experts to formalize knowledge into structured formats. 17. Track record of open-source contributions or collaborative software development. 18. Understanding of data protection regulations (GDPR) or health research governance.
Welsh Language Level	Level 1 – 'a little' - pronounce Welsh words. Able to answer the phone in Welsh (good morning / afternoon). Able to use very basic every-day words and phrases (thank you, please etc.). Level 1 can be reached by completing a one-hour training course. For more information about the Welsh Language Levels please refer to the Welsh Language Skills Assessment web page, which is available here.
Additional Information	Informal enquiries: Marcos Del Pozo Banos (m.delpozobanos@swansea.ac.uk) and Kenvil Souza (k.c.souza@swansea.ac.uk)