

THE UNIVERSITY OF MANCHESTER
PARTICULARS OF APPOINTMENT
FACULTY OF SCIENCE & ENGINEERING
SCHOOL OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE
RESEARCH FELLOW (HEALTH INFORMATICS)
VACANCY REF: SAE-029430

Salary: Grade 7 £47,389 to £58,225 per annum, depending on relevant experience

Hours: Full Time

Duration: Fixed term until 31 March 2027

Location: Oxford Road, Manchester

Enquiries about the vacancy, shortlisting and interviews:

Name: Shoaib Sufi

Email: shoaib.sufi@manchester.ac.uk

BACKGROUND

This role sits in the [eScience Lab](#), the research group led by [Professor Carole Goble](#) and Dr Stian Soiland-Reyes at the department of Computer Science, University of Manchester in the UK. The Lab is focused on research and development of tools designed for data driven and computational research. The eScience Lab has a long history of working with research use cases and creating tools to support them. The main tools and platform the team are currently working on are systems for sharing experiments ([FAIRDOM SEEK](#)), a registry of scientific workflows ([WorkflowHub](#)), a way to package and describe Research Objects ([RO-Crate](#)) and an online 'best practice guide' for data management, focused on the life sciences ([RDMKit](#)). The group is currently involved in over 20 funded projects, with funding from UKRI, Horizon 2020, Horizon Europe / Innovate UK and subcontracted research grants. While the group is focused on life science and health use cases, it works cross-domain in areas such as astronomy, social science and chemistry, as well as cross-cutting concerns such as reproducible research, provenance, community building, scholarly communication and software sustainability.

TREvolution is a [DARE UK](#) project that seeks to advance Trusted Research Environments (TREs) by addressing challenges in scaling their capabilities for complex, multi-site data analysis, especially with AI. It aims to develop a federated network of TREs with common standards,

enhanced security through semi-automated disclosure control, and interoperable computational environments. TRevolution will build on existing DARE UK projects, creating a unified reference implementation and engaging with stakeholders, including the public, to ensure trustworthiness and adoption. Ultimately, the goal is to enable more efficient and secure research using sensitive data, benefiting the public through faster scientific advancements and informed policy. This post is funded by the TRevolution project.

The postholder will work closely with the DARE UK early adopters:

- FOCUS-5: FEDERATING OPERATIONS AND COLLABORATIONS USING THE FIVE SAFES
- FRIDGE: FEDERATED RESEARCH INFRASTRUCTURE BY DATA GOVERNANCE EXTENSION
- PANORAMA: PAN-NORTH DATA RESEARCH ADVANCEMENTS FOR MULTI-DOMAIN ACCESS AND ANALYSIS
- SAFER: SAIL ADVANCING FEDERATED EXPLORATION AND READINESS
- SOAR: SEMI-AUTOMATED OUTPUT ASSESSMENT AND REVIEW
- VISTA: VIABLE IMPLEMENTATION OF SATRE, SACRO AND TOOLS FOR AI

More information about the DARE UK early adopters can be found at - <https://dareuk.org.uk/how-we-work/ongoing-activities/dare-uk-early-adopters/>

The postholder will focus on research on RO-Crate metadata based approaches for information representation, evaluation and enhancement for the TRE early adopters. They will collaborate with other research groups and bring the areas of metadata, synthetic data, data architecture and natural language processing together to investigate the state of the art in this space.

In addition they will act as the senior product owner for the RO-Crate metadata related to the TRE early adopters guiding the work plans of the RO-Create technical team and the ways of working of the six TRE early adopter teams and associated three product team to increase the understanding, efficacy and transparency of metadata driven federated analytics aspired to by the TRE early adopters and the DARE programme.

The postholder will be working closely with, and be line managed by the Manchester Principal Investigator Dr Stian Soiland-Reyes and be part of the eScience Lab team based in the Department of Computer Science.

Overall Purpose of the Job

Experiment with and evaluate RO-Crate driver metadata mechanisms to represent information flow within and between TRE early adopters. You will look deeply into validation mechanisms and develop and evaluate a related query dataset. You will further investigate the use of natural language processing / text mining and synthetic data. As a leader you will collaborate with other related research groups and investigate approaches which could bring value to the TRE early adopters. You will lead on the establishment of research networks and consortia to support

knowledge sharing and partners for seeking research funds to take the work further. The role will include expected activities of a research fellow such as writing academic publications, presenting at conferences, supervising students and contributing to/leading on funding proposals.

You will bring clarity and certainty to the needs and adoption of RO-Crate to the early adopters of the DARE UK. Moving the current conceptual usage forward to practical usage and guiding the implementation. You will lead the metadata architecture of RO-Crate that meets the needs of the early adopters. You will act as a senior product owner setting the priorities of the work for both the RO-Crate technical team and influencing the ways of working of the six TRE early adopter team and three product teams.

Key Responsibilities, Accountabilities or Duties

Research into health informatics

- Experiment with and develop evaluation mechanisms for RO-Crate metadata based profiles to represent information within the TRE early adopters and exchange between them
- Examine the needs of TRE early adopters and use these to inform further development of metadata standards to capture relevant information, e.g., developing new RO-Crate profiles.
- Investigate validation mechanisms for semantic data structures, such as SHACL and SPARQL, and their combination in formalised profiles/schemas.
- Build up a set of knowledge representation queries to help evaluate schema profile instances for the individual TRE early adopters both for internal and external data.
- Formalisation of evidence tracking profiles for TRE activities, extending existing literature on semantic provenance standards
- Investigate the use of Natural Language processing in TRE early adopters
- Investigate the use and development of good synthetic data with the associated text mining based evaluations for creating safe and portable testing environments representing TRE architectures and practice.
- Collaboration with other research groups within and outside the university in areas of interest (e.g. text mining and digital health).
- Horizon scanning and new approaches which will aid TRE early adopters data, analytics, transparency and sharing needs; this will focus on applied areas which can bring value early.
- Take a leadership role in establishing an associated research network in an European context, bridging between metadata, text mining, large language models and European infrastructure such as [ELIXIR](#).
- Drafting and publishing academic articles on these topics, e.g. for conferences
- Helping funding proposals to gain further funds
- Co-supervision of PhD/MSc students on related work

Research consultancy/leadership:

- Leading the engagement with the Trusted Research Environments (TRE) early adopters on the DARE Programme (<https://dareuk.org.uk/how-we-work/ongoing-activities/dare-uk-early-adopters/>) on the DARE programme which consists of six early adopter teams and three product teams. This will involve developing a productive and efficient way of working with them and bridging between them and the RO-Crate technical team at Manchester.
- Understand the metadata, data and analytics requirements in the context of the RO-Crate specification and tooling; this will include research consultancy on how the technology fits into their architectural needs.
- Represent the early adopters needs as the main product owner guiding RO-Crate implementation to match their needs (this will involve working with and directing the RO-Crate technical team of the eScience Lab). It is envisaged that this will require creating the RO-Crate profile and helping specify new tooling to match their use cases.
- The postholder will also be required to understand the training needs of the early adopters and work with the RO-Crate technical team to develop and deliver the associated upskilling workshops
- The postholder will develop the metadata architecture to cover the needs of the different TRE early adopters.
- Guide the health use case through the FAIR (Findable, Accessible, Interoperable and Reusable). They will highlight the usefulness and also do this in a sympathetic way, understanding the data sharing requirements and privacy needs as a priority and understand other legitimate barriers to adoption.
 - Findability needs to be subject to access requirements
 - Reusable - TREs are often opaque in how they work and capturing this and being able to repurpose for another TRE is a very useful area of further work - the postholder will use RO-Crate technologies to allow this to become more transparent and allow capturing the process, making it transparent and allowing carry over.

Product ownership/leadership

- The postholder will be the Senior Product Owner of RO-Crate & Metadata needs for the DARE UK early adopter community.
 - They will lead on technology adoption
 - They will curate an updated backlog of issues and guide the sprint planning to update the work plans of the RO-Crate technical team members
 - They will influence the work plans of the early adopters to cover
 - Clarity of their work plans and ways of working
 - Specification changes
 - Establish agreed upon mechanisms for how change is communicated
 - By understanding the technology maturity levels and bring them up to a minimal useful metadata level.
 - Support changes to communication between TRE early adopters.

- Lead on RO-Crate as the glue between different product teams and adopter teams, e.g. to be much more specific about what they are communicating internally and what they are communicating externally.

Note

The job description represents a guide of duties required and is not intended to be definitive. It may be subject to variation from time to time, following consultation with the post holder.

PERSON SPECIFICATION

Essential:

- BSc (2.1 or higher grade or a higher degree) in a course with a significant life computing aspect, or equivalent (e.g. on the job) experience.
- A PhD with a significant life science/computing aspect, or equivalent (e.g. on the job) experience.
- A proven track record of productivity, including evidence of contributing to manuscripts for publication in high-quality peer-reviewed journals as first/corresponding author
- Experience of developing research ideas and designing detailed analytical strategies independently, and progressing those ideas to completion in a short timeframe
- Experience of working with health data from multiple sources, particularly within environments such as SDEs / TREs (e.g. within the NHS).
- Detailed understanding and experience of applying advanced metadata, ontology and natural language processing / text mining techniques, including work at the intersection of these areas.
- Experience developing metadata standards / ontologies in live industry settings, where responding to stakeholder needs is essential in short timeframes.
- High level of proficiency in Python, R or equivalent programming language for data processing
- Experience of productive stakeholder engagement (e.g. with collaborators, practitioners and software engineers)
- Ability to explain complex ideas and analyses to non-technical audiences
- Experience of identifying and critically appraising relevant academic and policy literature
- Ability to work independently and as part of a multidisciplinary and multi-site teams
- Excellent time management and organisation skills, with the ability to prioritise workload and manage time to meet deadlines and objectives
- Clear oral and written presentation skills, with the ability to communicate with both academic and policy audiences
- Excellent interpersonal communication skills
- A willingness to travel in the UK, Europe or internationally when required (maximum 20%).

Desirable:

- EU project experience
- An understanding of [FAIR issues](#) related to data, software, training or workflows.
- Product ownership experience
- Understanding of the digital health programmes in England
- Broad knowledge of modern research methods used in digital health
- Experience of disseminating research findings to non-academic audiences
- Experience of contributing to successful research funding applications
- Experience of supervising junior colleagues and/or research students.