



Data Engineer (Research and Real World Data)

Candidate Information

September 2023

The Institute of Cancer Research (ICR)

About our organisation

We are one of the world's most influential cancer research institutes with an outstanding record of achievement dating back more than 100 years. We are world leaders in identifying cancer genes, discovering cancer drugs and developing precision radiotherapy. Together with our hospital partner The Royal Marsden, we are rated in the top four centres for cancer research and treatment worldwide. As well as being a world-class institute, we are a college of the University of London.

We came second in the league table of university research quality compiled from the Research Excellence Framework (REF 2021). We have charitable status and rely on support from partner organisations, charities, donors and the general public.

We have charitable status and rely on support from partner organisations, charities, donors and the general public.

We have more than 1000 staff and postgraduate students across three sites – in Chelsea and Sutton.

About the Research Teams and the Generations Study

The Data Engineer will work with the Integrative Epidemiology Team (led by Professor Montserrat Garcia- Closas) and Clinical Epidemiology Team (led by Professor Amy Berrington) at the ICR Division of Genetics and Epidemiology. The Division is internationally renowned for its pioneering work in understanding the underlying genetic and environmental causes of cancer risk. High-quality laboratory, epidemiological and clinical research within the division is driven by energetic, innovative leadership and complemented by participation in national and international research consortiums, clinical collaborations and technological partnerships.

Data Engineer (Research and Real World Data)

Candidate Information

At the Integrative Epidemiology Team we use integrative analyses of large-scale data in epidemiological studies to investigate the causes of cancer, understand carcinogenic processes and improve risk assessment for precision prevention. At the Clinical Epidemiology Team we use real world data to investigate the late-effects of cancer treatments, cancer survival and cancer risks from other medications. Our work informs prevention and public health strategies at both the population and individual levels to reduce the burden of cancer.

We have a program of research based on the ongoing Generations Study, a national study of over 110,000 women from the UK established. Women in the study have provided blood samples and detailed questionnaire information at recruitment, and in repeat follow-up questionnaires. Data includes self-reported risk factor information, hormone levels, genetics, and artificial intelligence (AI) analyses of tissue images from breast tumours, benign breast disease and mammography images. We also access their medical records to collect information on cancer screening and treatments. The scientific staff comprise epidemiologists, statisticians and data scientist who collaborate with researchers around the world.

We are part of the newly formed Cancer Epidemiology and Prevention Research Unit [https://www.icr.ac.uk/our-research/centres-and-collaborations/strategic-collaborations/the-cancer-epidemiology-and-prevention-research-unit-\(cepru\)](https://www.icr.ac.uk/our-research/centres-and-collaborations/strategic-collaborations/the-cancer-epidemiology-and-prevention-research-unit-(cepru)), a research partnership between The ICR and Imperial College London to establish collaborations in research, training and knowledge dissemination in cancer epidemiology and prevention.

We are seeking to appoint an experienced **Data Engineer** to join our dynamic and forefront research group using epidemiological and real-world data-driven approaches to understand the causes of cancer and how to prevent it.

The Job Role and Requirements

As a Data Engineer you will be responsible for designing, developing, and maintaining the data infrastructure that supports our research initiatives, with a particular focus on handling sensitive data. You will collaborate with multidisciplinary teams of epidemiologists, statisticians, data scientists and data managers to ensure that data is available, reliable, and secure for various research projects.

This is an exciting opportunity to play a crucial role in the creation of end-to-end data management and processing solutions, according to FAIR (Findable, Accesible, Interoperable and Re-usable) principles to support efficient and secure research data reuse to advance science.

Data Engineer (Research and Real World Data) Candidate Information

You will be working in the ICR campus based in Sutton, London under the direction of Professor Garcia-Closas. The post will be full time for 2 years in the first instance with potential for renewal thereafter. The starting date would be as soon as practical, by mutual agreement. Salary will be on the scale £47,500 to £57,680 per annum according to qualifications and experience. This is an on-site job; however flexible, hybrid ways of working can be considered with a minimum mutually agreed days working on-site to enable the collaboration and contact with other members of the team.

For further information, potential applicants can contact Professor Garcia-Closas (E-mail: montse.garcia-closas01@icr.ac.uk). More information about the Institute of Cancer Research can be found on our website (www.icr.ac.uk).

Data Engineer (Research and Real World Data)

Candidate Information

Our values

The ICR has a highly skilled and committed workforce, with a wide variety of roles, each requiring different skills. But whether you work as a researcher, or work as part of our corporate team, your work and behaviour is underpinned by these six values. They are what bring us together as one team - as 'One ICR'.



Pursuing excellence

We aspire to excellence in everything we do, and aim to be leaders in our field.



Acting with Integrity

We promote an open and honest environment that gives credit and acknowledges mistakes, so that our actions stand up to scrutiny.



Valuing all our people

We value the contribution of all our people, help them reach their full potential, and treat everyone with kindness and respect.



Working together

We collaborate with colleagues and partners to bring together different skills, resources and perspectives.



Leading innovation

We do things differently in ways that no one else has done before, and share the expertise and learning we gain.



Making a difference

We all play our part, doing a little bit more, a little bit better, to help improve the lives of people with cancer.



Our values set out how each of us at the ICR, works together to meet our mission – to make the discoveries that defeat cancer.

They summarise our desired behaviours, attitudes and culture – how we value one another and how we take pride in the work we do, to deliver impact for people with cancer and their loved ones.”

Professor Kristian Helin
Chief Executive

Data Engineer (Research and Real World Data)

Candidate Information

Job description

Department / division: Genetics and Epidemiology Division

Pay grade / staff group: Analytical Scientist 3 (SP03)

Hours / duration: Full time (35 hours per week), Monday to Friday. Fixed term contract for 2 years (with potential for renewal thereafter)

Reports to: Montserrat Garcia-Closas, Professor of Epidemiology

Accountable to: Professor Garcia-Closas

Main purpose of the job: To design, implement and document data architecture and data modelling solutions for analyses of large-scale epidemiological and real-world data.

Duties and responsibilities:

The post holder will be expected, under supervision, to work on the following, as required:

Key Duties and Responsibilities

Data Architecture and Modeling:

- Evaluate existing systems to design and implement data architectures that facilitate efficient storage, retrieval, and analysis of existing and new research data.
- Develop data models that support data integration, normalization, and transformation
- Reduce data redundancy and streamline data movements

Data Integration:

- Build ETL (Extract, Transform, Load) pipelines to integrate data from various sources into a central repository.
- Ensure data quality and consistency by implementing data validation and cleansing processes.

Data Security and Compliance:

- Implement robust data security measures to protect sensitive research data from unauthorized access or breaches.
- Ensure compliance with data protection regulations and research ethics.

Interoperability, scalability and Performance:

- Optimize data systems for interoperability, scalability, performance, and reliability to handle large volumes of research data.

Data Documentation and Metadata Management:

- Maintain detailed documentation of data sources, transformations, and schemas.
- Manage metadata to facilitate data discovery and lineage tracking.

Collaboration:

Data Engineer (Research and Real World Data)

Candidate Information

- Collaborate with cross-functional teams to understand data requirements and ensure data solutions align with research goals.

Data Backup and Recovery:

- Develop and implement data backup and recovery strategies to safeguard against data loss.

Data Governance:

- Establish and enforce data governance practices, including access controls, data retention policies, and auditing.

Technology Evaluation:

- Stay up-to-date with emerging data technologies and make recommendations for technology adoption when relevant.

Any other duties which may be required, that are consistent with the nature and grade of the post

This job description is a reflection of the present position and is subject to review and alteration in detail and emphasis in the light of future changes or development.

General

All staff must ensure that they familiarise themselves with and adhere to any ICR policies that are relevant to their work and that all personal and sensitive personal data is treated with the utmost confidentiality and in line with the General Data Protection Regulations

Any other duties that are consistent with the nature and grade of the post that may be required.

To work in accordance with the ICR's Values.

To promote a safe, healthy and fair environment for people to work, where bullying and harassment will not be tolerated.

This job description is a reflection of the present position and is subject to review and alteration in detail and emphasis in the light of future changes or development.

Data Engineer (Research and Real World Data) Candidate Information

Person specification

Education and Knowledge

Bachelor's degree in computer science, data science, information technology or related discipline	Essential
Master's degree in computer science, data science, information technology or related discipline	Desirable

Experience

Hands-on experience in data modelling, including designing, building, and testing data products based on feeds from multiple systems, using a range of different storage technologies and access methods.	Essential
Experience in delivering data solutions in accordance with agreed organisational standards that ensure services are resilient, scalable, and future proof.	Essential
Experience in producing, maintaining, and updating relevant data models for an organisation's specific needs.	Essential
Experience in designing, coding, testing, correcting and documenting programmes or scripts.	Essential
Expertise in database technologies (SQL, NoSQL), data storage, and data manipulation	Essential
Experience in communicating complex issues and data concepts to non-technical staff.	Essential
Experience with data visualization tools and reporting platforms.	Desirable
Experience with team management.	Desirable

Skills/Knowledge

Strong programming skills in languages such as SQL, Python and JavaScript.	Essential
Knowledge of best practices in software engineering, including code documentation, testing, and optimisation.	Essential
Knowledge of version control systems (e.g. GitHub)	Essential
Knowledge of metadata management and related tools	Essential
Knowledge of data security best practices and compliance (e.g., GDPR, HIPAA).	Essential
Proven understanding of Research Data Management (RDM), Open Science, FAIR and/or digital curation communities.	Essential
Strong problem-solving skills and ability to work independently and collaboratively in a fast-paced environment.	Essential
Strong attention to detail	Essential

Data Engineer (Research and Real World Data) Candidate Information

Excellent communication skills to interact with multidisciplinary teams and convey complex technical concepts to non-technical stakeholders.	Essential
Familiarity with cloud platforms (e.g., AWS, Azure, GCP) and containerization (Docker, Kubernetes).	Desirable
Knowledge of common data models in the medical field such as Observational Medical Outcomes Partnership (OMOP), and vocabularies such SNOWMED-CT, LOINC, and RxNorm / dm+d.	Desirable
Knowledge of OHDSI software open-source tools to support data-analytics in health data sciences.	Desirable
Familiarity with data modeling software such as Microsoft Visio, erwin Data Modeler, or SAP PowerDesigner.	Desirable
Familiarity with mathematical foundations and statistical analysis.	Desirable
Familiarity with data manipulation and analyses software such as R, SAS or STATA	Desirable

Data Engineer (Research and Real World Data)

Candidate Information

Benefits

We offer a fantastic working environment, great opportunities for career development and the chance to make a real difference to defeat cancer. We aim to recruit and develop the best – the most outstanding scientists and clinicians, and the most talented professional and administrative staff.

The annual leave entitlement for full time employees is 28 days per annum on joining. This will increase by a further day after 2 years' and 5 years' service. All staff receive an additional three days at Christmas.

Staff membership to the Universities Superannuation Scheme (USS) is available. The USS is a defined benefit scheme and provides a highly competitive pension scheme with robust benefits. The rate of contributions is determined by USS and details of the costs and benefits of this scheme can be found on their website. If staff are transferring from the NHS, they can opt to remain members of the NHS Pension Scheme.

We offer a range of family friendly benefits such as flexible working, a parents' group, and a maternity mentoring scheme. Other great benefits include interest free loans for discounted season tickets for travel and bicycle purchases, access to the NHS discounts website, a free and confidential Employee Assistance Programme which offers a range of well-being, financial and legal advice services, two staff restaurants, and access to a gym and sporting facilities at our Sutton site.

Further information

You may contact Prof. Montserrat Garcia-Closas for further information by emailing Montse.garcia-closas01@icr.ac.uk. This job description is a reflection of the current position and is subject to review and alteration in detail and emphasis in the light of future changes or development.