

Joe Heffer

Senior Research Data Engineer

Sheffield, UK

heffer.info

github.com/Joe-Heffer-Shef

linkedin.com/in/joe-heffer-50323743

orcid.org/0000-0001-8733-1117

I build reliable research software and data pipelines that help researchers turn messy data into clear answers. I'm a physicist turned research data engineer, embedded with research teams across disciplines at the University of Sheffield, working across the whole data lifecycle — from sensitive NHS health records to environmental sensor streams and genomic sequencing data.

EXPERIENCE

Senior Research Data Engineer Jan 2020 – Present

University of Sheffield — IT Services

Embedded with research teams across disciplines, designing research software and building data pipelines. Lead engineer on the CUREd+ research database, linking NHS hospital, mental health and ambulance service records to reveal how patients move through the healthcare system. Mentor engineers and contribute to the Research Software Engineering community.

Python SQL Data Pipelines Secure Data Environments

Analytics Developer 2017 – 2020

Financial Services

Delivered data analytics and software in the financial services industry, before moving into research computing.

Analytics Software Development

PhD, Physics 2013 – 2017

University of Liverpool

Doctoral research in atom interferometry, as part of precision gravimetry and searches for dark contents of the vacuum.

MPhys, Physics and Astronomy 2008 – 2013

The University of Manchester

SELECTED WORK

SORT Online 2026

Software engineering and data analytics for a platform built with NHS England that lets healthcare organisations self-assess their research readiness — secure data handling, consent pathways, and progress-tracking dashboards.

FAIR² for Research Software 2023

A training course teaching researchers to document their software, developed with the University of Sheffield's Research Software Engineering team.

Urban Flows Observatory 2021

Data engineering for the University of Sheffield's citywide sensor network — servers for field data-loggers and Python pipelines ingesting public air-quality and river-level data.

AIRBODS airborne-infection research 2021

Data engineering for an EPSRC-funded study of airborne SARS-CoV-2 transmission — ingesting IoT air-sensor time-series into secure storage, orchestrated with Apache Airflow.

CUREd+ Research Database 2020 – Present

A linked healthcare research database bringing together NHS hospital, mental health and ambulance service records.

COG-UK genomics pipeline 2020

Software-engineering support for the University of Sheffield's COVID-19 Genomics UK effort — architecture, containerisation and automated testing for a Nanopore sequencing pipeline.

TOOLKIT

Python R SQL Linux Git
AI / Machine Learning Data Pipelines
HPC

AT A GLANCE

Based in Sheffield, UK

Focus Data Engineering & Applied AI for Research

Status Open to collaboration

AI & RESEARCH COMMUNITY

PROGRAMME LEAD

AI for Research Day 2026

Centre for Machine Intelligence's annual conference, leading the Programme.

PARTICIPANT

Data Director Agentic AI Blueprint

Research Data Alliance working group on agentic AI tools for FAIR data management.

ORGANISER

AI Solutions for Research Support

Workshop on practical AI tools and approaches for research support.

SESSION LEAD

Research Agents Reading Group

Fortnightly reading group on agentic AI tools and research.

MEMBER

Shef.AI Community

Regular attendee of Sheffield's AI community meetings and Centre for Machine Intelligence town halls.