



Job Description

Research Data Scientist: Creating The Human Organ Atlas

Department: UCL Mechanical Engineering

Grade: 7/8

Fixed term: 3 years; opportunity for a full time permanent position within ARC, subject to suitable performance evaluation.

Location: UCL Bloomsbury Campus

Reports to

Dr Claire Walsh and Prof. Peter Lee (Mech. Eng.), and Jonathan Cooper - Acting Head of Data Science (ARC)

About the Project

Would you like to explore, manage, infer, validate, analyse, and visualise some of world's newest and most complex bioimaging data? The Human Organ Atlas project and UCL Advanced Research Computing (ARC) are looking for a Research Data Scientist/Machine Learning Specialist to work with high resolution 3D imaging datasets from the world's newest bioimaging imaging technique - Hierarchical Phase-Contrast Tomography (HiP-CT).

HiP-CT is a multiscale Synchrotron X-ray imaging technique developed in 2020 through a collaboration between UCL, the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), and Clinicians in Germany. It enables whole human organs to be scanned hierarchically: the whole organ at ca. 10 $\mu\text{m}/\text{voxel}$ down to 1 $\mu\text{m}/\text{voxel}$ in local regions (see bit.ly/HiP-CT-PW). This unprecedented imaging data is being used to create The Human Organ Atlas, <https://human-organ-atlas.esrf.eu>, which aims to become the highest resolution 3D interactive map of the human body. (bit.ly/HiP-CT-videos).

The datasets created by HiP-CT are large (currently up to 16TB, but will become even larger) and require advanced computing to manipulate, visualise and share. The successful candidate will work jointly with UCL's Advanced Research Computing (ARC) Centre and the Human Organ Atlas Project team to provide data science expertise to curate, visualise, share and extract insights from these datasets.

At the end of the project with the Human Organ Atlas there will be the opportunity to move to a full time permanent position within ARC, subject to suitable performance evaluation. Further information about

working in ARC can be found on our website www.ucl.ac.uk/arc.

About the Collaborators

The post involves working directly with core groups of the project: i) The Human-Organ-Atlas team (UCL), ii) ARC (UCL), and iii) the computational infrastructure team (ESRF). In addition there will be opportunity for diverse collaboration within UCL with e.g. UCL Centre for Medical Image Computing CMIC, UCL Centre for Advanced Biomedical Imaging (CABI) and the MSM@H group, as well as contribution to the wide portfolio of ARC projects.

International collaboration is a core part of the Human Organ Atlas project, collaborating with other imaging teams and with clinical academics seeking to address fundamental biomedical challenges with HiP-CT.

All collaborators agree to HiP-CT data being ingested into the Human Organ Atlas and shared to broaden accessibility and impact. Given this ambitious scope, collaborating with partners in large image data handling e.g. Viren Jain and Jeremy Maitin-Shepard (Google), Nicholas Sofroniew (CZI-napari team), and the eBrains initiative (<https://ebrains.eu/>), will be within the scope of the role.

Context

Funded by the Chan Zuckerberg Initiative, you will be part of a fast growing International Collaboration to utilise Hierarchical Phase-Contrast Tomography (HiP-CT), to create the world's highest resolution interactive map of the human body.

The larger project will: i) develop HiP-CT, ii) populate 'The-Human-Organ-Atlas', and iii) use the data to answer biomedical questions e.g. understanding COVID-19 injury patterns or providing ground truth for the human connectome project. You will be part of a multidisciplinary international team of X-ray physicists, computer scientists, medics and computational modellers. Dr Claire Walsh, the UCL Director of the project, and Prof Peter Lee, the project PI, will be your local line managers with additional

supervision and mentorship provided by the Head of Data Science at ARC.

ARC is UCL's institute for infrastructure and innovation in digital research – the supercomputers, datasets, software and people that make computational science and digital scholarship possible. We are both a professional services department that delivers reliable and secure infrastructure and services to UCL research groups, and a laboratory for research and innovation in the application of advanced computational and data intensive research methods. Our positioning as a hybrid of a research institute and service centre means these activities will be synergistic.

The post holder will benefit from associate membership of ARC, being part of our professional community, and working within the scaled agile people-activity allocation process, you will agree with line management between one and three concurrent research projects, service changes, or capability enhancements to which you will contribute. We will work together to keep these appropriate to your skills and development needs, providing variety and opportunities for growth.

Main purpose of the job

- To develop robust pipelines for large image data handling, (e.g. compression, transfer, upload, data format conversion) from the servers at ESRF through to data ingestion into the Human Organ Atlas.
- Develop, work with and expand existing software platforms (e.g. neuroglancer) and in-group scripts to implement data visualisation and interaction in robust and scalable ways.
- To understand the data handling needs of the wider team and seek to deploy pipelines and training to address these.
- Work with research infrastructure developers at UCL and ESRF to deploy these to production.

Duties and responsibilities

The post is available as either a Research Data Scientist (Grade 7) or a Senior Research Data Scientist (Grade 8), depending on experience.

Both Grades:

- To develop robust pipelines for HiP-CT image data handling, from the ESRF local servers through to data ingestion into the Human-Organ-Atlas.
- To develop our image data visualisation platform (currently based in Neuroglancer), developing it to increase the ease of use and amount of data interaction that can be performed.
- To collaborate with ESRF Data Scientists to ensure the data curation and metadata pipelines

they develop are fit for purpose, and compatible with UCL infrastructure.

- To work with PDRA and PhDs developing deep-learning based image registration and segmentation pipelines, helping them deploy their pipelines on the large datasets, both on HiP-CT and other projects within the group.
- To participate and help manage the preparation high quality ground truth datasets for machine learning.
- Design and delivery of teaching and training to the wider team, through providing demonstration or seminars, or delivering individual training or teaching sessions, or authoring materials.
- To engage with the wide-range of stake-holders including biologists, clinicians and other image analysts, to understand the data manipulation needs of the group.
- To maintain an awareness of research literature that is pertinent to the project. You will take the time to grow within your fields of expertise, trying new things and building experience and knowledge.
- Contribute to research and innovation dissemination, through authoring sections of technical white papers, formal research papers, blog posts.
- To disseminate research findings at appropriate meetings, workshops and conferences.
- To prepare progress reports on research for funding bodies as required.
- To contribute to the overall activities of the research team and Department as required.

For Grade 8, Senior Research Data Scientist, you will be responsible and accountable for all of the above, and in addition at least two of the following additional responsibilities. These duties will vary over time according to discussion with your line management.

- *Opportunity development and research leadership:* building and delivering the growth of research, teaching and innovation income for the group through building collaborations, leading or contributing to the preparation and drafting of research bids, grant applications and proposals.
- *Technical leadership:* Leading design, architecture and implementation for one or more technical aspects of research projects. This may include supervising the technical work of individual contributors.

- *Delivery ownership/project management:* Coordinating and leading a research activity, ensuring timely progress.
- *Design and delivery of teaching and training courses:* including lecturing on UCL modules or training courses, and authoring materials.
- *Co-supervision of MSc and/or PhD students* in the Human Organ Atlas Team.

Person Specification

Criteria	Essential or Desirable	Assessment method (Application/Interview)
Essential Experience/Skills		
<ul style="list-style-type: none"> PhD Degree in computer science or equivalent combination of degrees and professional experience 	Essential	Application
<ul style="list-style-type: none"> Demonstrable proficiency in scientific computing in a Linux/UNIX environment, with experience working in a programmatic framework such as R or Python 	Essential	Application, Interview
<ul style="list-style-type: none"> Experience with image analysis/image processing 	Essential	Application, Interview
<ul style="list-style-type: none"> Experience in high performance and/or cloud computing 	Essential	Application, Interview
<ul style="list-style-type: none"> Production experience with the implementation of data visualisation packages 	Essential	Application, Interview
<ul style="list-style-type: none"> Knowledge of and commitment to technology best practices that enable the delivery of reliable, sustainable and trustworthy outputs, e.g. version control, CI/CD 	Essential	Application, Interview
<ul style="list-style-type: none"> Experience authoring technical documents to a high standard, for example, research papers, technical white-papers, documentation describing a technology solution, or design, strategy or policy papers in a digital research space 	Essential	Application, Interview
For Grade 8 appointment as Senior Research Data Scientist		
<ul style="list-style-type: none"> <i>Experience mentoring or leading other research technologists or students, either through formal line management or through guiding the work of more junior colleagues</i> 	<i>Essential</i>	<i>Application, Interview</i>
<ul style="list-style-type: none"> <i>Experience in a number of successfully completed complex technical projects, having had a leading role in bringing about those successes (either technically or managerially)</i> 	<i>Essential</i>	<i>Application, Interview</i>
<ul style="list-style-type: none"> <i>Experience contributing to the development of research proposals, business cases, or tenders for infrastructure submissions</i> 	<i>Essential</i>	<i>Application, Interview</i>
<ul style="list-style-type: none"> <i>Demonstrable leadership in one or more research or technology communities, such as through membership of standards bodies or editorial boards, organisation of community events, or leadership role for open-source tools or packages</i> 	<i>Desirable</i>	<i>Application, Interview</i>
Preferred Experience/Skills		
<ul style="list-style-type: none"> Experience with large (>1TB) imaging data. 	Desirable	Application, Interview
<ul style="list-style-type: none"> Experience with or working knowledge of data curation tools 	Desirable	Application, Interview
<ul style="list-style-type: none"> Experience with or working knowledge of deployment of containerized application 	Desirable	Application, Interview
For Grade 8 appointment as Senior Research Data Scientist		
<ul style="list-style-type: none"> <i>Process or project management expertise</i> 	<i>Desirable</i>	<i>Application, Interview</i>
<ul style="list-style-type: none"> <i>Experience supporting live, running technology services</i> 	<i>Desirable</i>	<i>Application, Interview</i>
Personal Attributes		
<ul style="list-style-type: none"> Ability to work under own initiative 	Essential	Application, Interview
<ul style="list-style-type: none"> Commitment to high-quality, interdisciplinary research 	Essential	Application, Interview
<ul style="list-style-type: none"> Commitment to UCL's policy of equal opportunity and the ability to work harmoniously with colleagues and students of all cultures and backgrounds 	Essential	Interview
<ul style="list-style-type: none"> Ability to work collaboratively and as part of a team, both in-person and using remote collaboration technologies 	Essential	Application, Interview

Criteria	Essential or Desirable	Assessment method (Application/Interview)
<i>For Grade 8 appointment as Senior Research Data Scientist</i>		
<ul style="list-style-type: none"> • <i>Skilful at facilitating and enabling group collaboration</i> 	<i>Essential</i>	<i>Application, Interview</i>
<ul style="list-style-type: none"> • <i>Proven ability to manage multiple concurrent tasks and activities, working to deadlines and prioritising as appropriate</i> 	<i>Desirable</i>	<i>Application, Interview</i>