

Joseph David Ruff

+44 7948 835 684

ruffmaestro.jr@gmail.com

East Sheen, London

GitHub: github.com/JosephRuff

Website: JosephRuff.github.io

Personal Profile

Research Software Engineer and Data Infrastructure Specialist with expertise in scientific computing, high-dimensional biological data, and standards-based data integration.

As Lead Developer on the AGENT Project at the Royal Botanic Gardens, Kew, I designed and delivered FAIR-compliant, analytics-ready digital infrastructure supporting the integration and interoperability of global plant genetic resource data.

My work focuses on relational data modelling, ETL pipeline architecture, RESTful API development (BrAPI), and quantitative software systems that prioritise data quality, reproducibility, and cross-institutional interoperability.

Technical Skills

Core Languages: Python, SQL, R, C++, Java

Data Infrastructure & Engineering: Relational data modelling; ETL pipeline architecture; data cleaning, deduplication, and validation; RESTful API design (BrAPI); FAIR data standards; reproducible workflows; Docker-based deployment

Scientific & Quantitative Computing: High-dimensional data analysis; statistical modelling; NumPy; SciPy; MATLAB; VTK-based 3D computational visualisation

Databases: PostgreSQL, MariaDB/MySQL, Oracle, SQLite

Data Visualisation: Matplotlib, Plotly, ggplot

Web & Application Frameworks: Django, Flask, Spring, JavaScript (jQuery)

Employment History

Royal Botanical Gardens, Kew

Lead Developer - AGENT Project

Mar 2022 - Oct 2025

- Engineered FAIR-compliant, analytics-ready digital infrastructure to integrate and harmonize global plant genetic resource datasets across multiple genebanks, ensuring interoperability, semantic consistency, and reproducibility.
- Designed and implemented relational data models, ETL pipelines, and BrAPI-based RESTful APIs, enabling programmatic access to harmonized datasets for advanced analysis, modelling, and cross-institutional research.
- Orchestrated automated and manual data validation workflows, improving data quality and enabling reliable quantitative analysis of high-dimensional biological datasets.
- Collaborated with multidisciplinary teams spanning biology, bioinformatics, and data architecture to translate complex scientific requirements into robust, production-grade software systems.

Royal Botanical Gardens, Kew

Developer – Useful Plants and Fungi of Columbia

Dec 2020 - Feb 2022

- Engineered a robust data management platform for Colombian plant and fungal datasets, enabling standardized recording, integration, and analysis across multiple scientific domains.
- Collaborated with experts in taxonomy, genomics, bioinformatics, spatial analysis, and data architecture to structure and harmonize complex, high-dimensional datasets for downstream analytics.
- Designed and implemented relational data models and data ingestion pipelines, incorporating cleaning, deduplication, and validation workflows to ensure data quality, integrity, and reproducibility.

- Iteratively refined platform functionality through close collaboration with users, supporting diverse research and analytical workflows.

Medical Physics and Bio-engineering, UCLH
Volunteer Software Developer

Nov 2020

- Translated medical imaging algorithms from C++ to Python, improving accessibility and enabling further quantitative analysis.
- Developed 3D imaging visualisations using VTK, including vertex-to-vertex scan comparisons to support quantitative assessment and visual interpretation of anatomical differences.

3DAM LTD
Website Designer / Developer

Aug 2020 - Oct 2020

- Designed and delivered a client-facing business website, translating requirements into a functional, user-friendly solution.

Physics, Royal Marsden Hospital
Volunteer Back-end / Server Engineer

Jun 2020 - Aug 2020

- Configured servers, monitoring systems, and backup pipelines using Docker and ZFS/RAIDZ2, ensuring reliable infrastructure under COVID-19 constraints.

Medical Physics and Bio-engineering, UCLH
Junior Clinical Scientist: Bioinformatics (Physical Sciences)

Sep 2018 - Oct 2019

- Delivered computational and quantitative solutions across multiple clinical and scientific teams, supporting medical physics and healthcare informatics projects.
- Designed and implemented statistical and mathematical software models to assist radiation protection planning and clinical decision-making.
- Built, maintained, and enhanced ETL data pipelines for clinical datasets, including cleaning, deduplication, validation, and reporting workflows.
- Conducted exploratory data analysis and produced actionable reports for clinical and technical stakeholders, adhering to data governance and international best practices.

Publications

Contributor to publications involving large-scale biological datasets, data standardisation, and international research collaboration.

- [Diazgranados, M., Ulian, T. \[and 54 others, including Ruff, J.\] \(2020\). **ColPlantA: Colombian resources for Plants made Accessible \(2nd ed.\)**. Royal Botanic Gardens, Kew, Richmond, UK.](#)
- [Gaya E., Diazgranados, M. \[and 39 others, including Ruff, J. \] \(2021\). **ColFungi: Colombian resources for Fungi Made Accessible**. Royal Botanic Gardens, Kew, Richmond, UK.](#)
- [Selby, P., Abbeloos, R. \[and 75 others, including Ruff, J.\] \(2025\). **BrAPI v2: real-world applications for data integration and collaboration in the breeding and genetics community**. Database, Volume 2025, 2025, baaf048. <https://doi.org/10.1093/database/baaf048>](#)
- [Le Floch, E., Adam-Blondon, AF., Alaux, M. \[and 52 others, including Ruff, J.\]. **Wheat historical phenotypic data from European genebanks as an important resource for research and breeding**. Sci Data 13, 566, 2026. <https://doi.org/10.1038/s41597-026-06908-x>](#)

Education

Loughborough University
Computer Science BSc; Second Class Honours, Upper Division

Leicestershire, United Kingdom
2014 - 2017