Building a Machine-Actionable Pipeline for Research Data

M. Ahmed, K. Buntic, A. Smith, J. Wilson, R. Macneil, J. Hetherington

Centre for Advanced Research Computing (ARC), University College London

Abstract

At UCL, the Centre for Advanced Research Computing (ARC) is building and supporting the vital integrated digital research infrastructure that is critical in allowing researchers to work with data at scale in an efficient and innovative manner.

Alongside its research and academic activities, ARC is providing a range of inter-connected research tools, services and technological initiatives with the aim of amplifying the reusability and impact of research data by making data FAIR-compliant and machine-actionable.

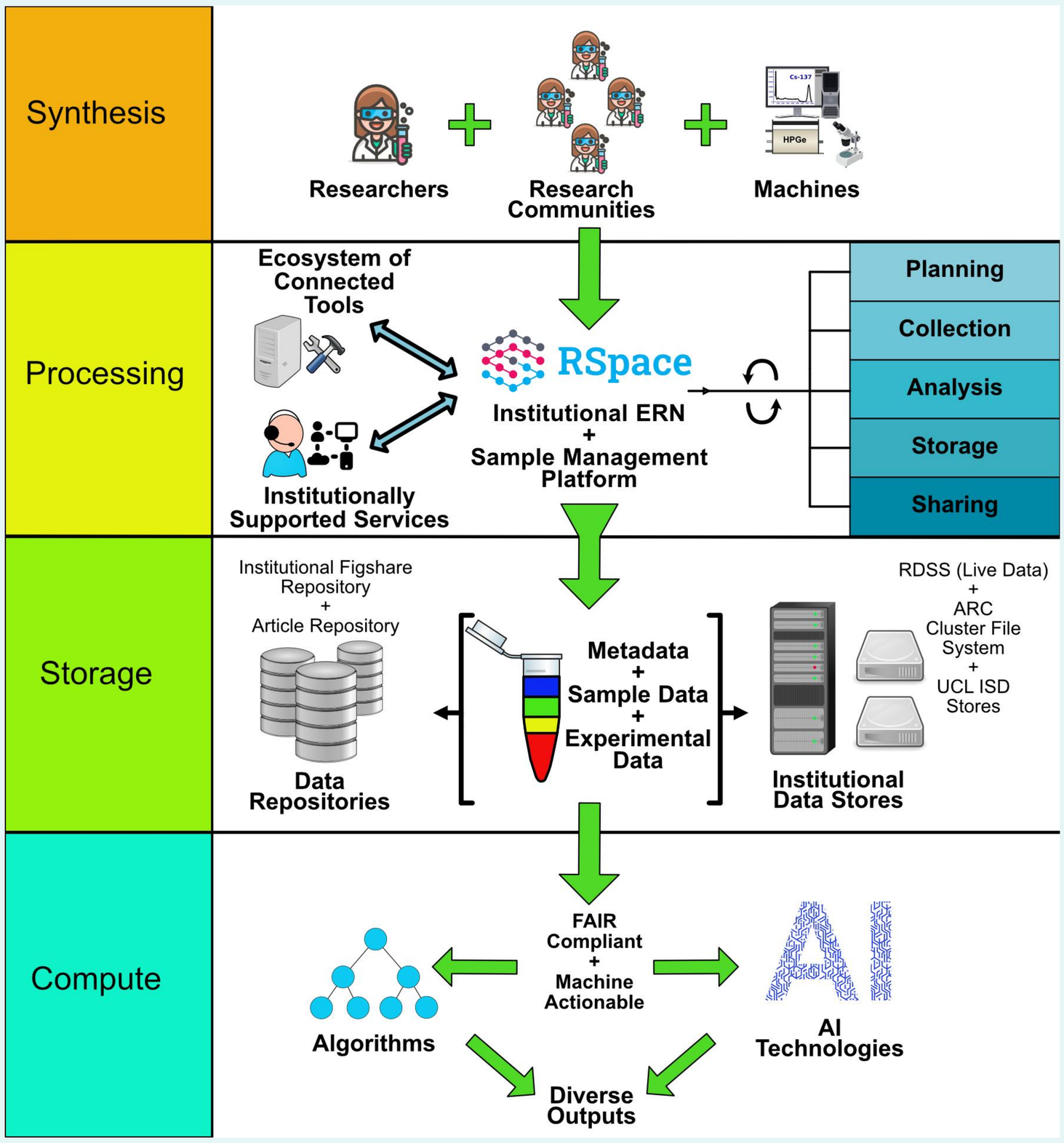
Ultimately, creating an ecosystem of institutionally supported tools and services will enable AI technologies and computing resources to be employed in extracting maximal value from research data produced at UCL across a variety of fields.

Context

Increasingly, the importance of properly managing, preserving and sharing research data is being recognized across the globe.

Currently, the true value of vast amounts of research data is unrealized, in part as a result of information silos, extensive use of paper notebooks and lack of associated metadata.

As global movements to make research data FAIR continue to evolve, it is becoming clear that dedicated support and initiatives must be put in place to enable computation at scale.



At UCL, ARC provides:

Large scale compute facilities including cloud computing, high-throughput and high-performance systems.

Data storage facilities and repositories including the institutional Figshare deployment and on-premise Research Data Storage Service for live data.

A range of tools, services, software and support to enable computationally empowered science and scholarship.

Research technology professionals- data scientists, informaticians, research software engineers, HPC systems engineers, dev-ops specialists, data engineers and data stewards - who collaborate with researchers.

The Role of Data Stewards

Data stewards at UCL operate in a cross-functional capacity to support, advise and collaborate with researchers. Their work involves translating research data needs into infrastructure and service requirements for deployment at the institutional level.

By providing support for uptake and use of the RSpace institutional ERN and contributing to development and integration with existing and future tools and systems at UCL, ARC is supporting the endeavors of researchers while encouraging the research community to take bold strides in making data both FAIR and machine-actionable.

Impact & Outcomes

ARC's hybrid mission focuses on both providing and developing state-of-the-art integrated digital research infrastructure needed to drive computationally empowered science and scholarship at UCL.

ARC aims to computationally enable the research community both at UCL and beyond through the delivery of reliable and secure interconnected tools, services and infrastructure, as well as through continual innovation in the application of advanced computational and data intensive research methods.

James Hetherington, Director of UCL ARC, j.hetherington@ucl.ac.uk

Rory Macneil, CEO of ResearchSpace, rmacneil@researchspace.com