Title	COPO - Bridging the Gap from Data to Publication in Plant Science
Authors	*A Etuk[1]*, F Shaw[1], A Gonzalez-Beltran[2], P Rocca-Serra[2], A Abdul-Rahman[2], P Kersey[3], R Bastow[4], S Sansone[2], R Davey[1]
Affiliations	The Genome Analysis Centre (1), Oxford e-Research Centre, University of Oxford (2), European Bioinformatics Institute (3), University of Warwick (4)
Contact	robert.davey@tgac.ac.uk
URL	https://documentation.tgac.ac.uk/display/COPO/
License	Common Public Attribution License Version 1.0 (CPAL)

The aim of open science is to make scientific research accessible, facilitating experimental reproducibility and transparency. Mechanisms exist for preserving and publishing research objects in plant science within the "omics" fields. However, researchers are often hindered by: (i) complicated and time-consuming procedures for repository deposition; (ii) a lack of interoperability between disparate information sources and mechanisms; (iii) sub-optimal search and retrieval facilities across data repositories; (iv) a lack of public awareness of existing services.

To address these issues, we are developing COPO (Collaborative Open Plant Omics), a brokering service which enables aggregation and publishing of research outputs by plant scientists, and provides access to services across disparate sources of information via web interfaces, and Application Programming Interfaces (APIs) for bioinformaticians. COPO comprises a web front-end (based on the Django framework), data grid infrastructure using iRODS (www.irods.org), and a number of APIs which facilitate the creation and management of logical profiles containing heterogeneous but related research objects representing a body of work. A profile can contain, for example, references to omics and image data, source code, PDF files, or posters and presentations, which are fully described with associated ISA-based metadata. COPO leverages the Investigation/Study/Assay (ISA) formats and ISA software suite tools (http://www.isa-tools.org) to enable experimental metadata attribution and conversion between metadata formats. Based on ISA research objects, this metadata comprises information about the investigators, objectives/hypotheses, related publications, subjects, experimental design and assays. Deposition to public repositories is enabled, providing a technical continuity between services. COPO will provide access to distributed resources through a single point of entry, integrating existing data services such as: European Nucleotide Archive (ENA) (www.ebi.ac.uk/ena) and MetaboLights (www.ebi.ac.uk/metabolights) for raw data; GigaScience for data and workflow citation (www.gigasciencejournal.com); f1000 (www.f1000research.com); Nature Publishing Group Scientific Data (www.nature.com/sdata/) for data publication; figshare (www.figshare.com) for a variety of supplementary data types; Galaxy (www.galaxyproject.org) and iPlant (www.iplantcollaborative.org) for data analysis; GitHub (www.github.com) for source code and ORCID (www.orcid.org) to provide user information. COPO does not store the data itself, but rather references deposited objects, thus keeping the system lightweight and decentralised from third party repositories. DOIs (Digital Object Identifiers) will be minted and mapped to COPO profiles providing permanent digital identities for these collections of research objects.

As research objects deposited through COPO will be accompanied by rich metadata implemented in JSON-LD (www.json-ld.org), cutting edge technologies such as MongoDB (www.mongodb.org) and neo4J (www.neo4j.org) are being used to create a web of linked semantic knowledge, allowing for user customised suggestions for future avenues of investigation. Such inferences are domain agnostic, with the potential for bioinformaticians to discover useful results and techniques from other fields.

We present a prototype front-end with the ability to create user accounts and work profiles. Data transfers to the ENA are possible with associated metadata being input into intuitive online forms. Metadata is stored in a document based database, with transfers between data formats (ISA metadata in RDF and JSON-LD) being handled by ISA software suite tools.