

<b>Title</b>	COPO - Bridging the Digital Gap from Data to Publication
<b>Author</b>	<i>Presenting Author</i> , Author Two, Author Three ...
<b>Affiliation</b>	The Genome Analysis Centre
<b>Contact</b>	robert.davey@tgac.ac.uk
<b>URL</b>	<a href="https://github.com/collaborative-open-plant-omics/COPO.git">https://github.com/collaborative-open-plant-omics/COPO.git</a>
<b>License</b>	Common Public Attribution License Version 1.0 (CPAL)

The aim of open science is to make scientific research accessible. This involves mechanisms for preserving and publishing research objects such as data, methods, software and manuscripts in an efficient and timely manner thus facilitating experimental reproducibility and transparency.

Existing services for enabling open science in plant research within the “omics” fields such as the European Nucleotide Archive (ENA) ([www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena)) and GigaScience ([www.gigasciencejournal.com](http://www.gigasciencejournal.com)) are limited 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.

COPO (Collaborative Open Plant Omics) is a brokering service which enables aggregation and publishing of research outputs by Bioinformaticians, as well as providing easy access to services across disparate sources of information. COPO (based on the Django web framework) is a web front-end and a set of APIs (Application Programming Interfaces), which facilitates the creation of logical profiles containing disparate but related research objects representing a body of work. A profile can contain for example, references to sequence data, source codes, pdf files, image data or posters and presentations, all of which are fully described with associated metadata. COPO seamlessly allows deposition of research objects to services such as ENA, GigaScience and f1000 ([www.f1000research.com](http://www.f1000research.com)) as well as grid based data infrastructures such as iRODS ([www.irods.org](http://www.irods.org)) and will interface with Orcid ([www.orcid.org](http://www.orcid.org)), thus unburdening Bioinformaticians. References to such deposited objects (which are not stored in COPO) will be maintained in a COPO profile. DOIs (Digital Object Identifiers) will be minted and mapped to COPO profiles providing a permanent digital identity to these collections of research objects. COPO will take advantage of the ISATools suite and the ISATab format developed by our collaborators at the University of Oxford ([www.isa-tools.org](http://www.isa-tools.org)) to provide technical continuity between services and researchers.

COPO will contribute to state-of-the-art open science by encouraging and implementing community-accepted standards for data representation; facilitating submission to persistent archival resources for data publication and citation; enabling seamless transition from data to analysis services and improving discoverability via contextualised ontological inference. Using APIs and virtualised resources, COPO will integrate other existing services such as Galaxy ([www.galaxyproject.org](http://www.galaxyproject.org)) and iPlant ([www.iplantcollaborative.org](http://www.iplantcollaborative.org)); figshare ([www.figshare.com](http://www.figshare.com)), GitHub ([www.github.com](http://www.github.com)), the Dryad Digital Repository ([www.datadryad.org](http://www.datadryad.org)), Metabo-Lights ([www.ebi.ac.uk/metabolights](http://www.ebi.ac.uk/metabolights)) and Zenodo ([www.zenodo.org](http://www.zenodo.org)), thus providing access to disparate and distributed resources through a single point of entry. For example, by hooking into an instance of Galaxy using data and workflows referenced in COPO, experimental results may be easily reproduced and verified, and comparative studies may be conducted with alternative data.

Since research objects deposited through COPO will be accompanied by rich metadata implemented in JSON-LD, cutting edge technologies such as MongoDB ([www.mongodb.org](http://www.mongodb.org)) and neo4J ([www.neo4j.org](http://www.neo4j.org)) will be used to create a web of semantic inferences, providing 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.

So far, a prototype front-end has been developed with the ability to create user accounts and work profiles. Data transfers to the ENA are possible with associated metadata being input into intuitive forms. Metadata is stored in a document based database, with transfers between data formats (JSON-LD and ISATab) being handled by ISASuite tools.