

Building Excellence in Genomics and Computational Bioscience

AIDING THE JOURNEY FROM DATA TO PUBLICATION IN THE LIFE SCIENCES

Robert Davey

robert.davey@tgac.ac.uk

 @froggleston

<http://www.tgac.ac.uk>



Collaborative Open Plant Omics

- (Good) Science is founded on reproducibility
- Reproducibility depends on:
 - reducing reinvention (*“friction”*)*
 - describing methods and data
 - maximising benefit to the researcher
- Methods pretty well covered through traditional publishing
- Data description sorely under-represented and used
- Benefits are often opaque
 - Fear of being scooped, loss of control, reputation, etc

* <http://cameronneylon.net/blog/network-enabled-research/>

Collaborative Open Plant Omics

- What prevents plant scientists from openly depositing their data and metadata?
 - **Lack of interoperability** between:
 - metadata annotation services
 - data repository services
 - data analysis services
 - data publishing services
 - Researchers might not:
 - **be aware** that the services exist
 - **have the expertise** to use them
 - **see the value** in properly describing their data

Collaborative Open Plant Omics

- Data:
 - Sample, Sequence, Genome, Proteome, Metabolome
- Code:
 - GitHub, BitBucket, Zenodo
- Analysis:
 - Galaxy, iPlant, Bioconductor, Taverna, local code/services
- Publication:
 - figshare, Scientific Data, Dryad, F1000, PeerJ, PLOSone
- Beyond the PDF:
 - Utopia, GitHub
- Training:
 - Materials, examples, workshops, bootcamps

Collaborative Open Plant Omics

- It's not because these services don't exist!
- Clearly, barriers exist between the scientist and the service
- Infrastructure can help by:
 - wiring existing services together
 - improving access to services
 - facilitating collaboration
 - raising profile of the benefits of open science
- How do we collaborate successfully to make this happen?

Mapping services with Application Programming Interfaces

Collaborative Open Plant Omics

APIs!

Collaborative Open Plant Omics

COPO

- Recently awarded BBSRC BBR grant
- TGAC, Univ. Oxford, Univ. Warwick, EMBL-EBI
 - Supported by GARNET community, iPlant, Eagle Genomics
- Empower bioscience plant researchers to:

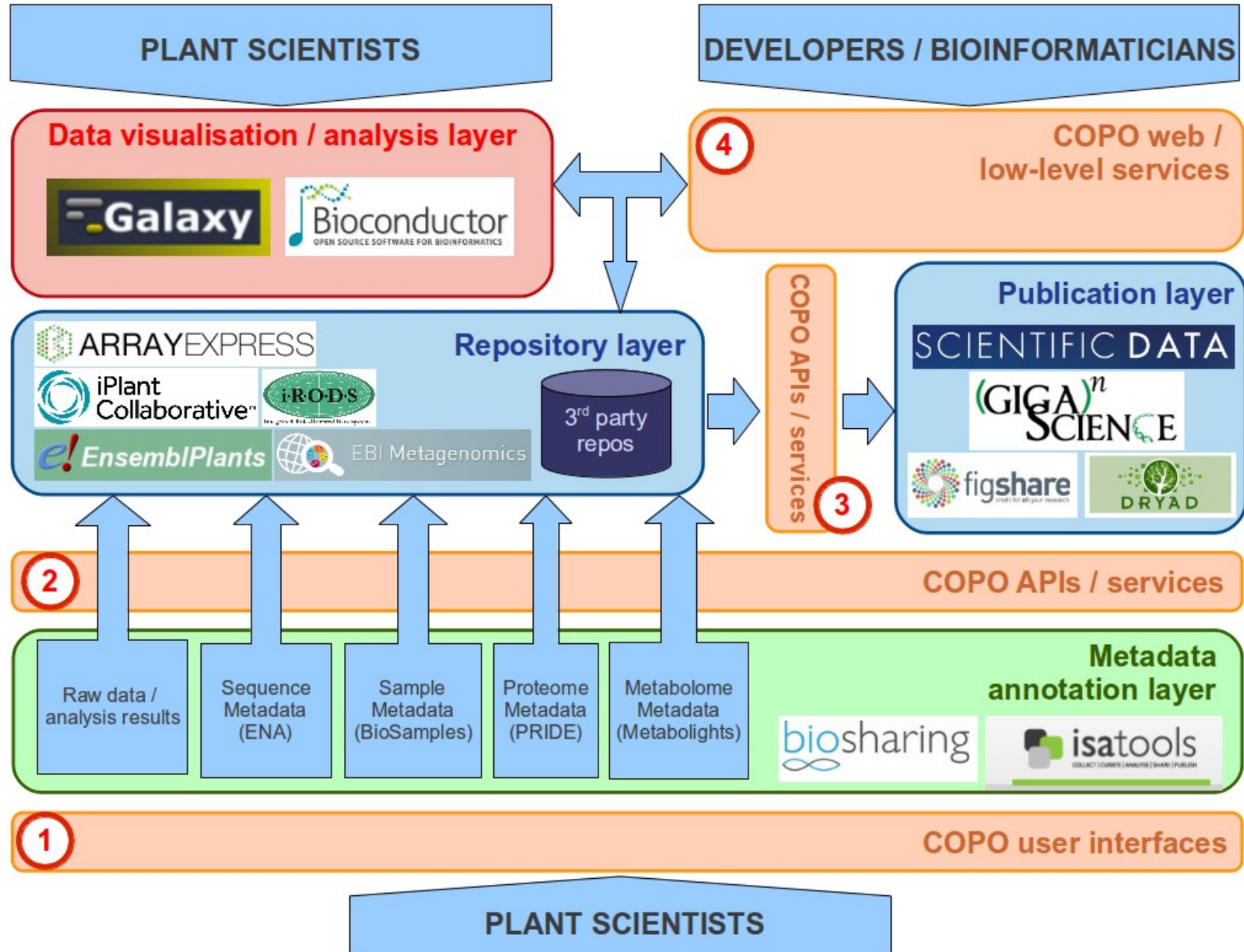
1. Enable standard-compliant data collection, curation and integration

2. Enhance access to data analysis and visualisation pipelines

3. Facilitate data sharing and publication to promote reuse

- Train plant researchers in best practice for data sharing and producing Research Objects

Collaborative Open Plant Omics



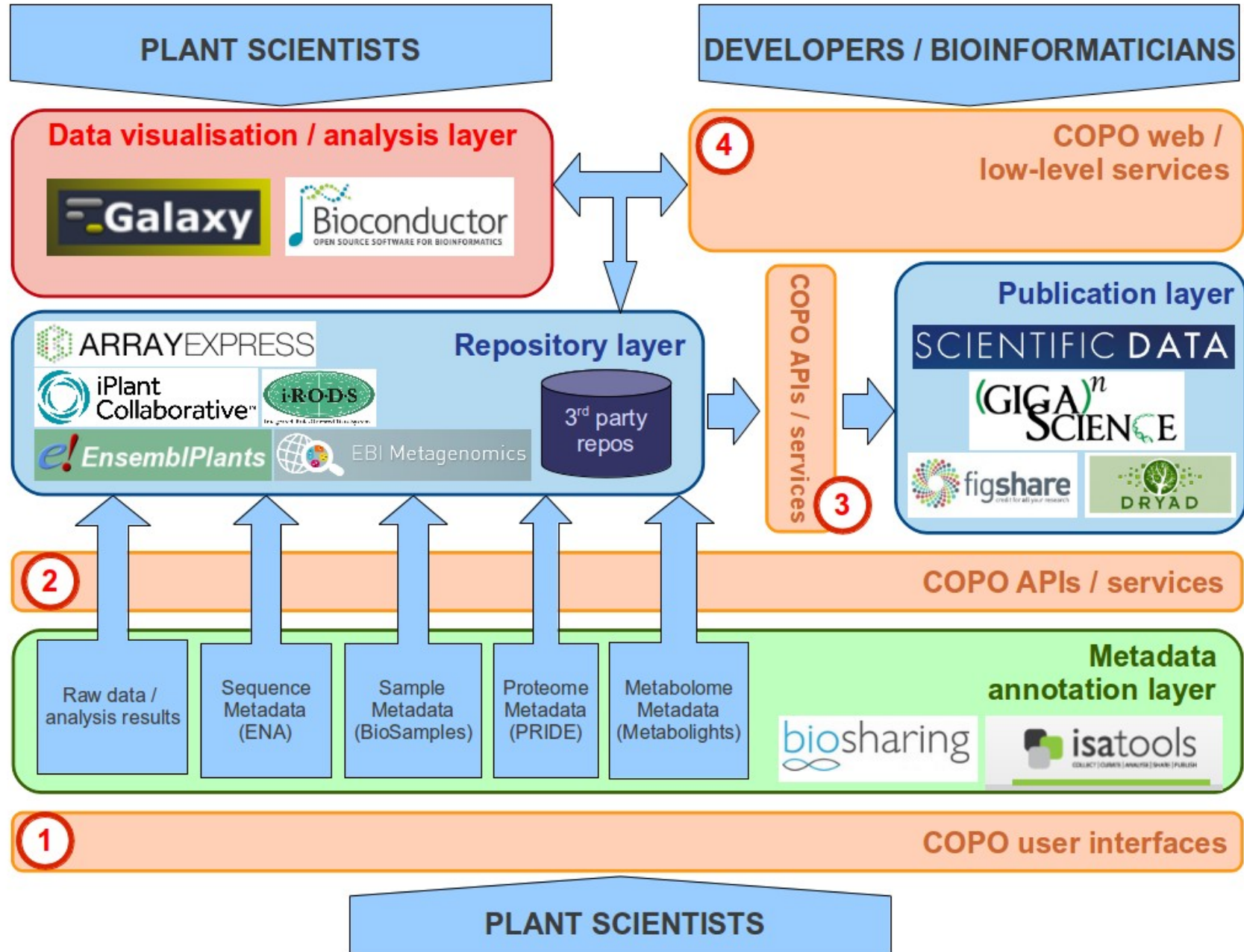
Collaborative Open Plant Omics

COPO user interfaces 1

- Implementation of user interfaces to
 - Access system as a whole (*brokerage*)
 - Existing metadata tools
- ISA suite
 - Metadata attribution suite developed at Oxford
 - Helps you to provide rich description of experimental metadata
 - Utilises widely accepted standards and supports submission to public repositories
 - Ontological underpinnings support validation and interoperability
- COPO will facilitate interaction between the user and the ISA software suite
 - metadata annotation
 - raw data preparation



Collaborative Open Plant Omics

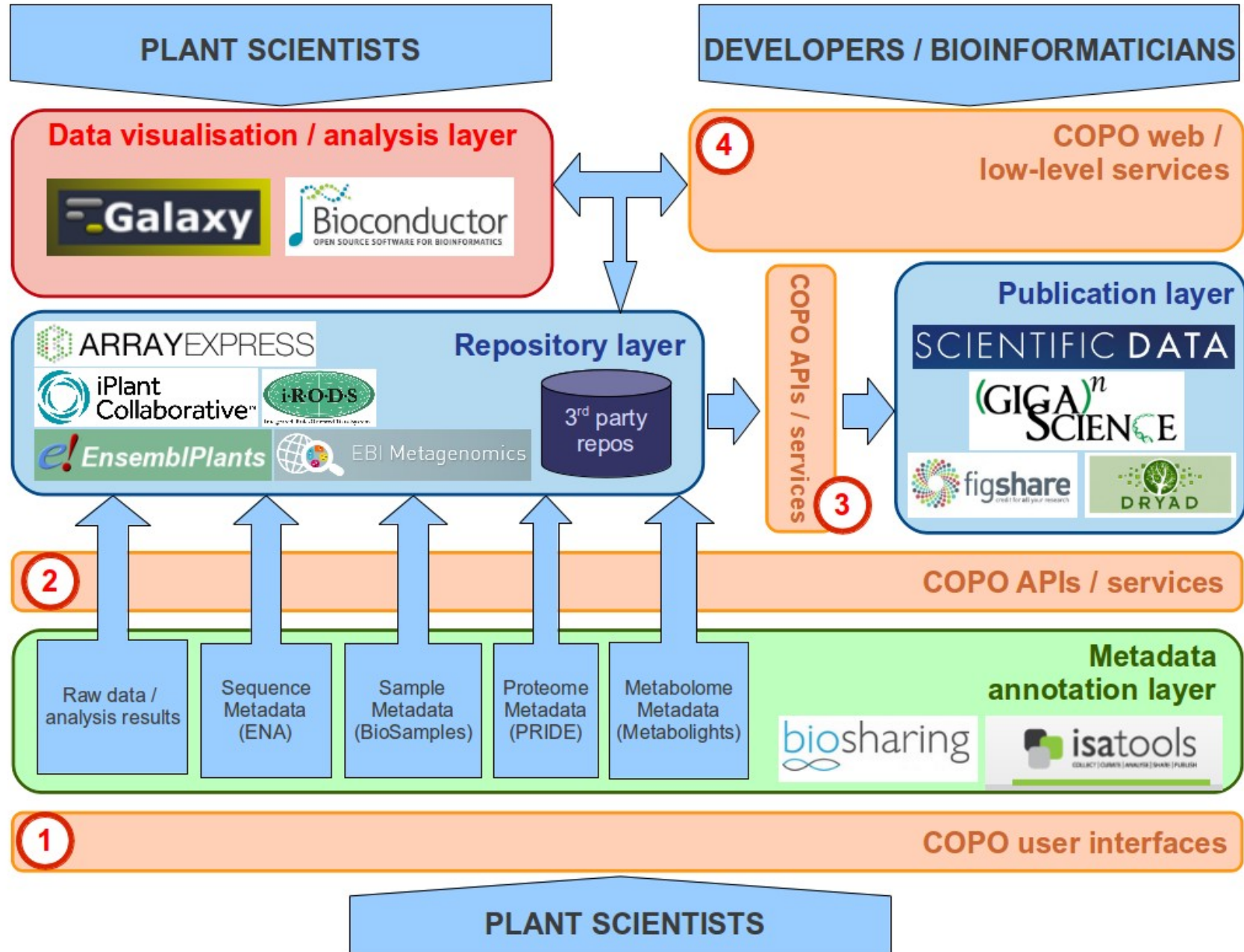


Collaborative Open Plant Omics

COPO data / metadata deposition APIs and services 2

- Allow programmatic access to deposited data and metadata
- Interact with public repository APIs
- Act as a staging post/broker for data and metadata
 - Tracked in the COPO system via its own accession paths
 - Minting DOIs eventually?
- Push packages comprising your data and descriptors to public repositories
- Return and store unique accessions to form digitally traceable objects

Collaborative Open Plant Omics

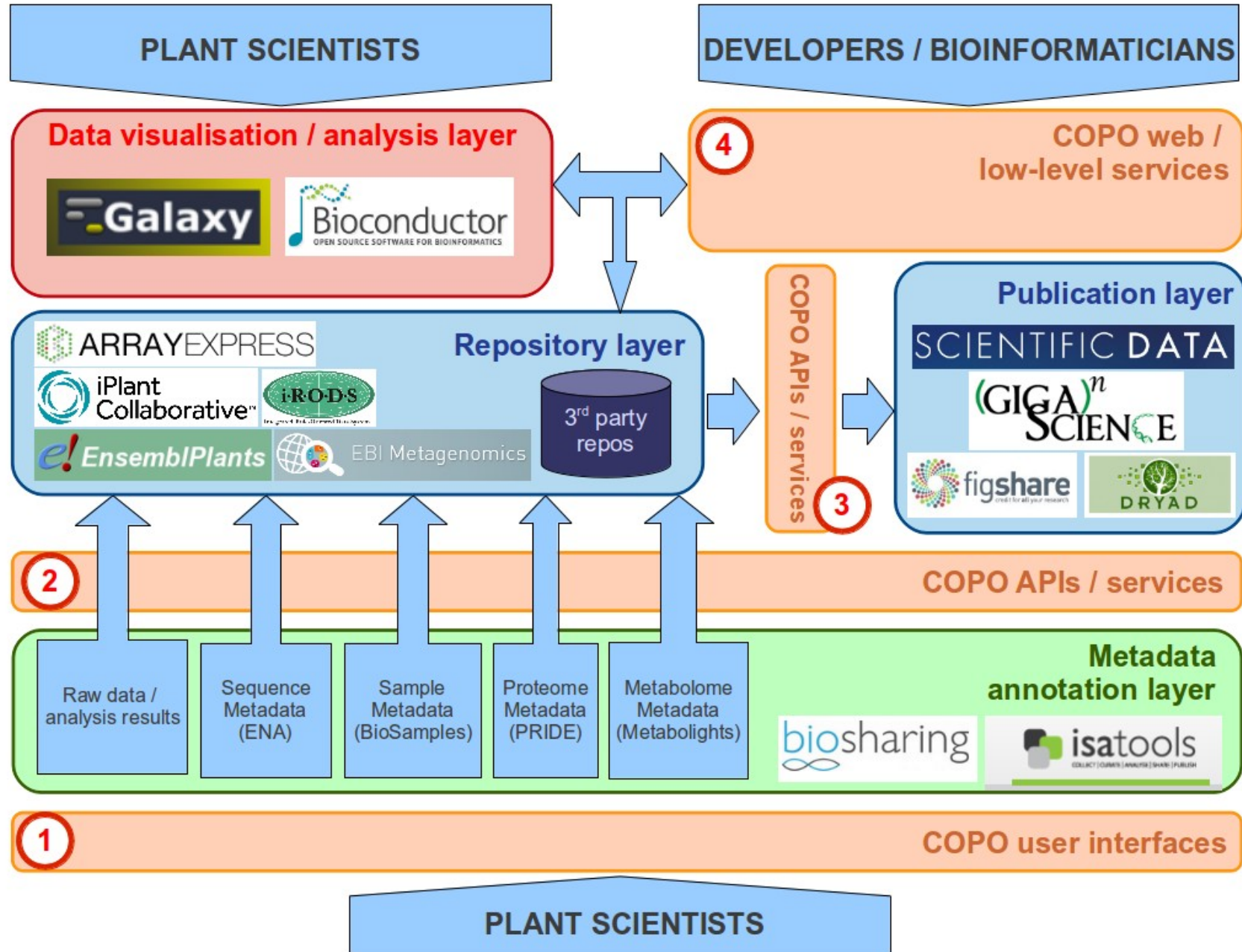


Collaborative Open Plant Omics

COPO data publication APIs and services 3

- Facilitate submission of data and metadata to publication platforms
- Allow programmatic access to published knowledge via
 - COPO accessions
 - Public repository accessions
 - Indexed search terms
 - Ontology-based searches over integrated metadata
- Store any new identifiers
 - URLs
 - DOIs
 - Secondary accessions

Collaborative Open Plant Omics



Collaborative Open Plant Omics

COPO developer / bioinformatician APIs and services 4

- Developed and provided across the COPO framework
- Allow bioinformaticians to access whole technology platform from their own programs
- Promote openness, reusability and interoperability
- Feed COPO any accession and users will be able to:
 - See all other accessions related to the supplied object
 - Original Data
 - Related Data
 - Publications and supplemental data
 - Get immediate access to any linked dataset
 - Forward these data and metadata to analysis platforms
 - Galaxy, iPlant, etc

Collaborative Open Plant Omics

We're hiring!

- COPO needs developers
- 2x software developers hosted at TGAC
- PhD not essential, but experience in bioinformatics domain very desirable

- Please contact me using:
 - Email: **robert.davey@tgac.ac.uk**
 - Twitter: **@froggleston**

Collaborative Open Plant Omics

Acknowledgements

- **TGAC**
 - Vicky Schneider
 - Mario Caccamo
- **University of Warwick**
 - Jim Beynon
 - Ruth Bastow
- **University of Oxford**
 - Susanna Sansone
- **EMBL-EBI**
 - Paul Kersey

