# Stakeholder input into the network graph provided by Third Global Biodiversity Informatics Conference (GBIC3)

## Deliverable D2.2

29 July 2022

Kyle Copas, Joe Miller

*Global Biodiversity Information Facility*

**BiCIKL**

**BIODIVERSITY COMMUNITY INTEGRATED KNOWLEDGE LIBRARY**

This project receives funding from the European Union's Horizon 2020 Research and Innovation Action under grant agreement No 101007492.

| | |
|---|---|
| Start of the project: | May 2021 |
| Duration: | 36 months |
| Project coordinator: | Prof. Lyubomir Penev<br>Pensoft Publishers |
| Deliverable title: | Stakeholder input into the network graph provided by Third Global Biodiversity Informatics Conference (GBIC3) |
| Deliverable n°: | D2.2 |
| Nature of the deliverable: | Report |
| Dissemination level: | Public |
| WP responsible: | WP2 |
| Lead beneficiary: | Global Biodiversity Information Facility |
| Citation: | Copas, K., & Miller, J. (2022). *Stakeholder input into the network graph provided by Third Global Biodiversity Informatics Conference (GBIC3)*. Deliverable D2.2 EU Horizon 2020 BiCIKL Project, Grant Agreement No 101007492. |
| Due date of deliverable: | Month 15 |
| Actual submission date: | 29 July 2022 |

Deliverable status:

| Version | Status | Date | Author(s) |
|---|---|---|---|
| 0.1 | Draft | 26 July 2022 | Kyle Copas & Joe Miller<br>GBIF Secretariat |
| 0.2 | Draft | 27 July 2022 | Anton Gunsch, BGBM |
| 0.3 | Draft | 27 July 2022 | Christos Arvanitidis, LifeWatch ERIC |
| 0.4 | Draft | 28 July 2022 | Ana Casino, CETAF |
| 1.0 | Final | 28 July 2022 | Kyle Copas & Joe Miller<br>GBIF Secretariat |

# Table of contents

# Preface

Members of the BiCIKL consortium have provided input on the relational network graph (D2.1) both during and after the project meeting hosted by LifeWatch ERIC in Seville on 1-5 May 2022. This deliverable documents that feedback and how it guided the development of the next version of the network graph and its associated web application.

# Summary

This deliverable documents feedback from members of the BiCIKL consortium and its allies on the relational network graph (D2.1). Project team members provided input during the project meeting hosted by LifeWatch ERIC in Seville on 1-5 May 2022 and in subsequent correspondence. Screenshots from the first version of the web application included in this deliverable provide the basis for comparing changes made in response to stakeholder input.

Opportunities for more in-depth discussions were unfortunately limited by Covid infections to key staff (including at least one case contracted at the Seville meeting). However, the lightweight and open-source approach to designing the network graph will enable the BiCIKL consortium's ongoing use and further refinement of the web application, currently located at https://gbif.shinyapps.io/bicikl-network-v2.

As described in D2.1, the BiCIKL network graph reuses and repurposes data from FAIRsharing (Sansone et al. 2019). This resource supplies a community-curated set of registries describing and interlinking data standards, databases and data policies that support or aspire to fulfil the FAIR Principles. Its website, FAIRsharing.org, allows users to search, browse and even visualise the complex, interlocking web of connections between the data standards, infrastructures and policies. The web app for the relational network graph extends the functionality of FAIRsharing.org, shifting the focus from FAIR-enabling resources to the organisations responsible for such resources and extending visualisations of FAIR-enabling organisations to a second order of connections, revealing a wider network of potential influence for BiCIKL.

Note: while initially proposed as a "3rd Global Biodiversity Informatics Conference" (GBIC3) jointly organised by the *alliance for biodiversity knowledge* and LifeWatch, the May 2022 meeting in Seville shifted to a much more operational focus on the project. Partly due to pandemic-related delays and, more importantly, the need for GBIF to establish appropriate governance structures for the *alliance*, pursuit of the wider agenda pursued in the previous GBIC events in 2011 and 2018 was deferred and GBIC3 postponed.

# List of abbreviations

| | |
|---|---|
| AAFC | Agriculture and Agri-Food Canada |
| BGBM | Botanic Garden and Botanical Museum Berlin |
| BiCIKL | Biodiversity Community Integrated Knowledge Library |
| BKH | Biodiversity Knowledge Hub |
| CERN | European Organization for Nuclear Research |
| CETAF | Consortium of European Taxonomic Facilities |
| DRAO | |
| EMBL-EBI | European Molecular Biology Laboratory-European Bioinformatics Institute |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable and Reusable |
| GBIC3 | Third Global Biodiversity Informatics Conference |
| GBIF | Global Biodiversity Information Facility |
| LifeWatch | LifeWatch ERIC |
| MBG | Meise Botanic Garden |
| Naturalis | Naturalis Biodiversity Center |
| Pensoft | Pensoft Publishers |
| Plazi | Plazi |
| RGB Kew | Royal Botanic Gardens, Kew |
| SIB | Swiss Institute of Bioinformatics |
| Sp2000 | Species 2000 |
| SRAO | FAIRsharing Subject Ontology (SRAO) |
| Tartu | University of Tartu |
| TDWG | Biodiversity Information Standards<br>*formerly* the Taxonomic Databases Working Group |

# 1. Stakeholder input

The initial version of the relational network graph (BiCIKL D2.1) was shown as a work-in-progress at the BiCIKL meeting in Seville on 3 May 2022.

## 1.1. Version 1 of the network graph

The first draft of the web application shown in Seville reflected the significant challenges of giving BiCIKL partners efficient and effective access for exploring and analysing the rich vein of institutional information stored in FAIRsharing.org.

The initial premise was to allow the user to start by selecting a BiCIKL consortium member and choosing to view the first- or second-level organizational connections. Once the user clicked "Visualize," the web app followed the precedent of FAIRsharing.org to reveal the tabular data first (Figure 1).

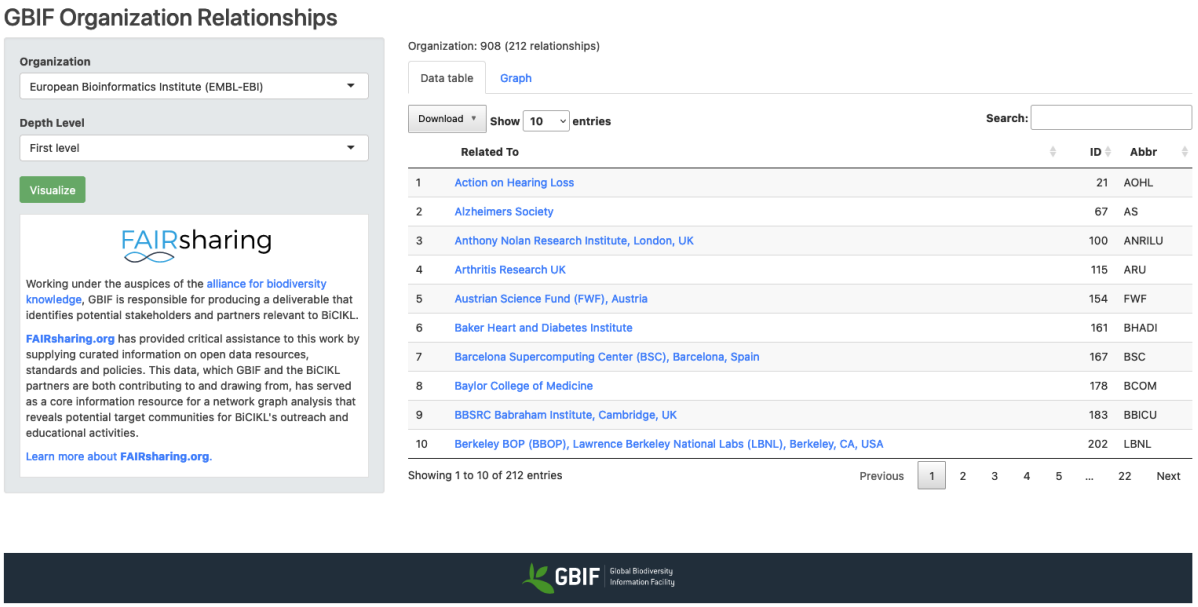


**Fig. 1:** *Example of table view showing the first-order relationships of EMBL-EBI with other organizations in FAIRsharing.org.*

This approach allowed the application to load a graphical view of the relationship in the "Graph" tab behind the table (Figure 2), which the user clicked to switch.

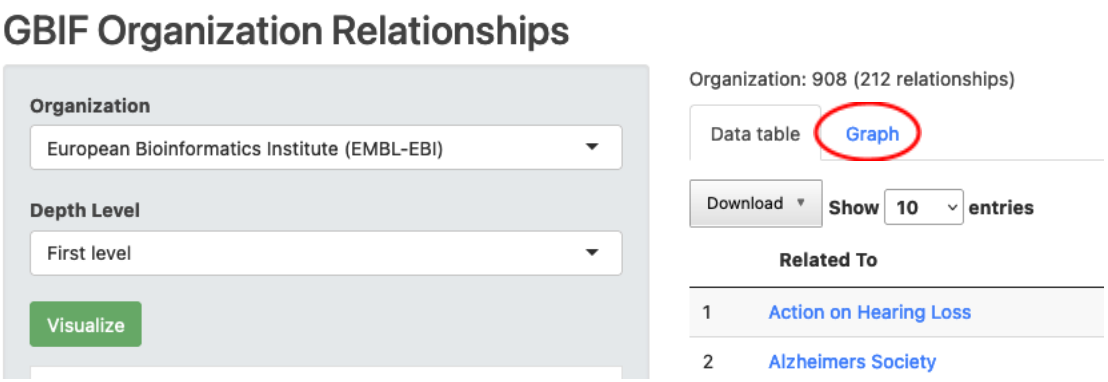


**Fig. 2:** *Detail of table view, showing the toggle to switch to a graph view of the first-order relationships of EMBL-EBI with other organizations.*

Upon switching to the graphic visualisation, the user saw an interactive arc graph that they could click on to highlight connections between individual nodes (Figure 3). This approach came after failures with initial tests of the "force-directed" graphs often used to depict social networks.

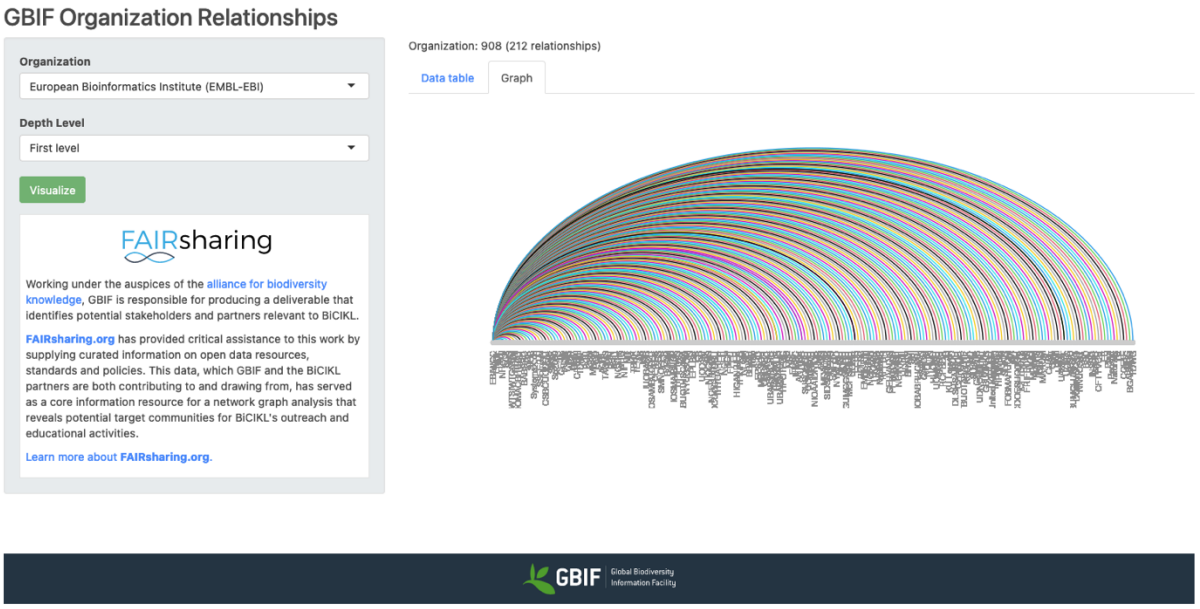


**Fig. 3:** *Example of arc-graph view showing the first-order relationships of EMBL-EBI with other organisations in FAIRsharing.org.*

One of the proposed extensions beyond the options available at FAIRsharing.org was to reveal the second-order connections between FAIR-enabling organisations, which users could access by choosing the "second level" depth option (Figure 4).

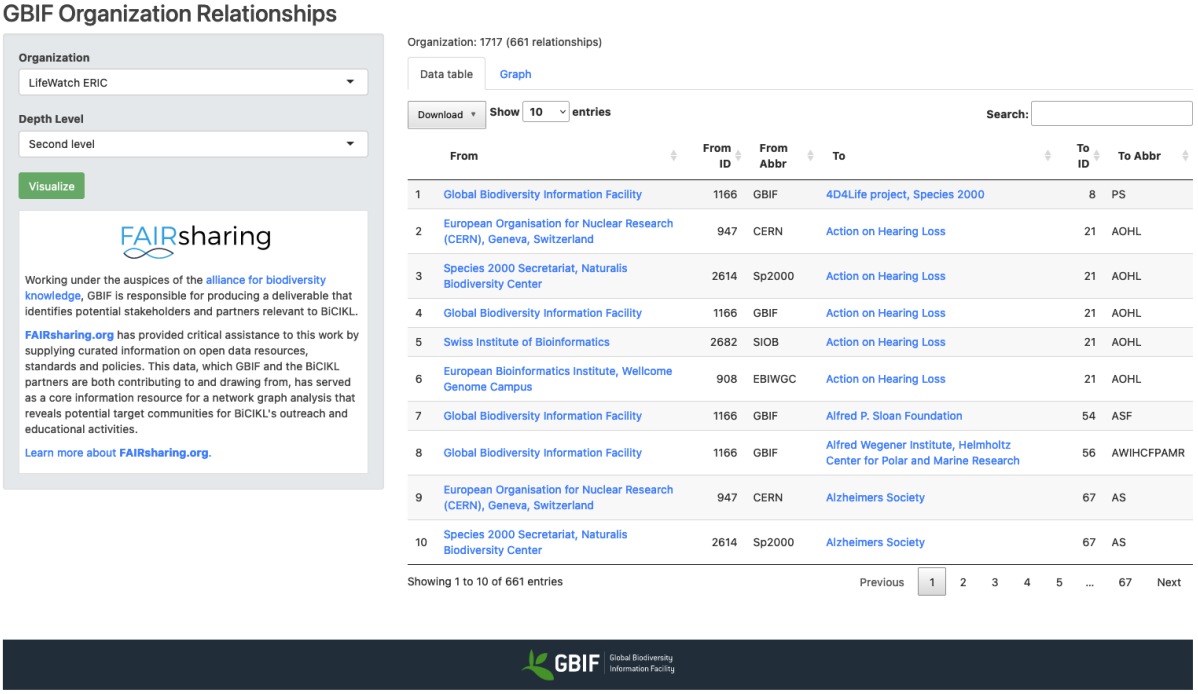


**Fig. 4:** *Example of table view showing the second-order relationships of LifeWatch ERIC with other organizations in FAIRsharing.org.*

As with the first-order view, this second-order arc graph would also be interactive, allowing users to highlight and reveal these connections (Figure 5).

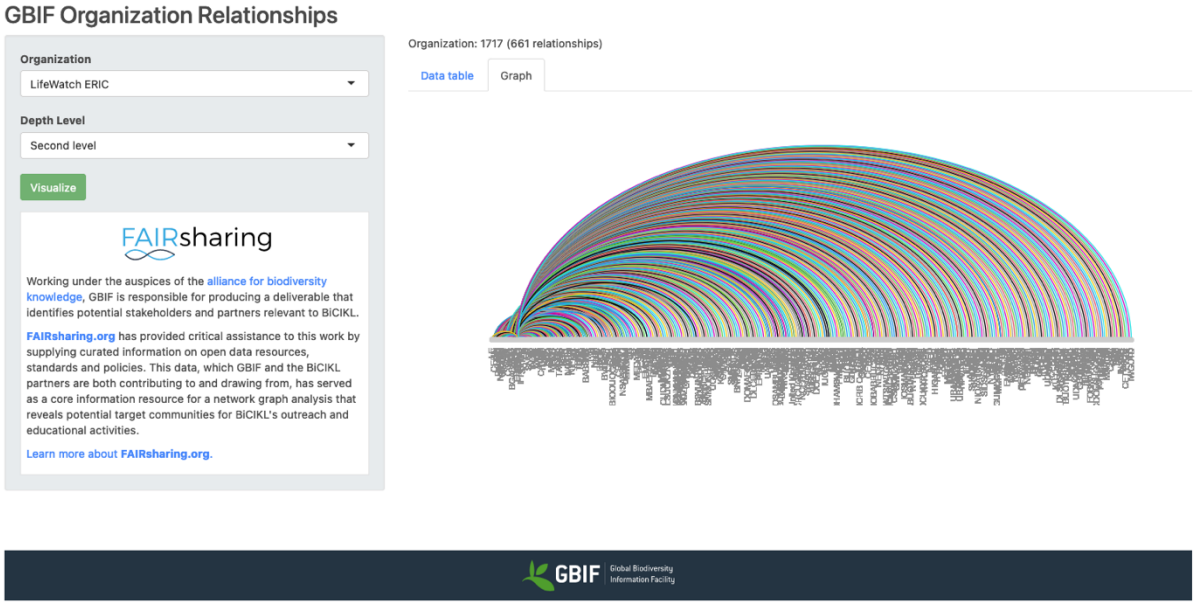


**Fig. 5:** *Example of arc-graph view showing the second-order relationships of LifeWatch ERIC with other organisations in FAIRsharing.org.*

## 1.2. Stakeholder input in Seville

During the discussion of the web app demo in plenary, comments focused almost exclusively on the graph views. Quentin Groom of MBG began by pointing out several issues with the arc graph, particularly:

- Relationships between the individual nodes are undifferentiated
- Displaying hundreds of relationships in an arc graph was a non-scalable approach

Groom strongly encouraged us to return to earlier experiments with the force-directed format.

One of the meeting's invited observers, Nicky Nicolson of RGB Kew, echoed Groom's recommendation. Nicolson offered another suggestion of applying a Sankey chart, a format often used to depict flows in, for example, energy analyses, using the specific example of a coal trade map: https://www.sankey-diagrams.com/coal-trade-flows-map/. However, in following up after the Seville meeting, Nicolson clarified her thought that such diagrams could be useful in laying out the life cycle of a specimen. In this framework, a Sankey chart could depict a series of events—specimen collection, accession, sequencing, identification, citation as type, support of hypothesis—each record of which could be used to visualise individual connections between the research infrastructures. Such relationships could also analyse the flow of information between the infrastructures' countries and the country of the researchers whose published work cites those specimens. However, dependence on record-level metrics likely places this well beyond the scope of BiCIKL's initial start-up.

Patricia Mergen of MBG called attention to a recent preprint that offered a precedent for various visualisations describing "contact zones" between several of the infrastructure involved in BiCIKL (Figure 6: Smith VS et al. 2022).

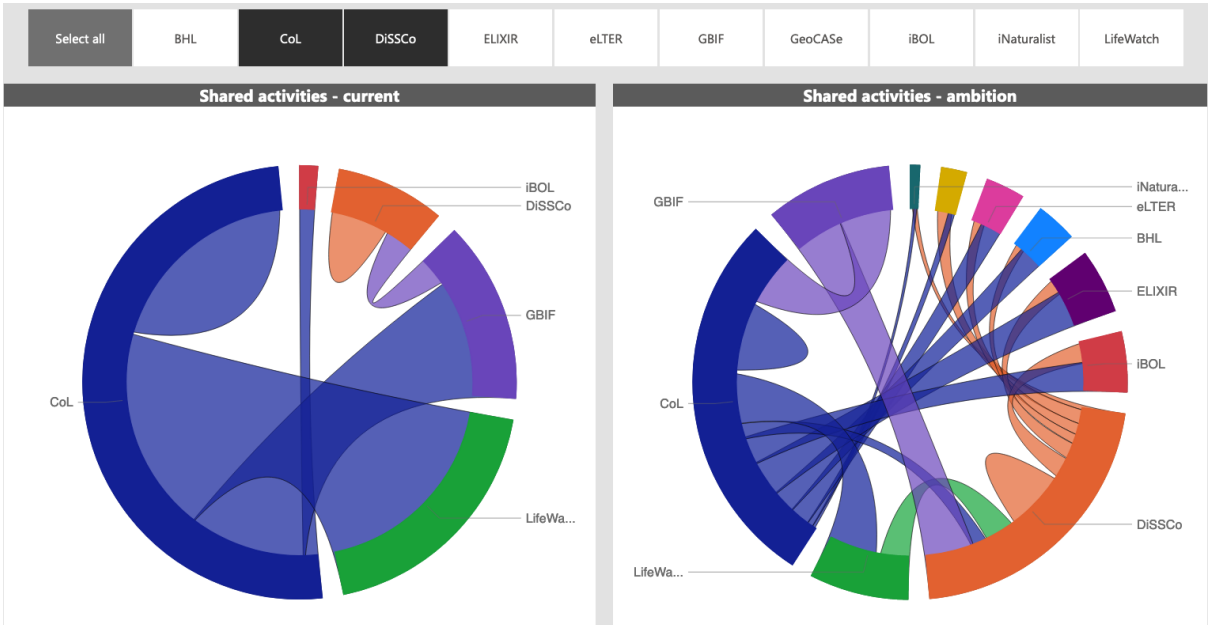


**Fig. 6:** *Visualisation precedent highlighting the contact zones connecting current and proposed activities between several BiCIKL and other research infrastructures.*

Discussions continued throughout the days following the plenary session. Another invited meeting participant, David Shorthouse of AAFC, emphasised the importance of distinguishing between the types of relationships by prioritising the most important ones and by adding

edge weighting to the connections. This first suggestion is quite visibly incorporated in the second version of the web app in the tool dialogue, where “undefined” and “funds” relationships are excluded by default.

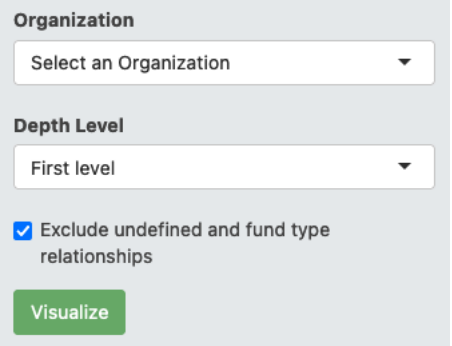


**Fig. 7:** *The default exclusion of “undefined” and “funds” relationships from FAIRsharing reduces the noise and amplifies the signal for connections of importance to BiCIKL.*

Other suggestions raised in discussion (whose provenance was sadly not captured) included:

- Don’t privilege the table view of the data over the graphic view
- Provide direct access to the original FAIRsharing source records
- Explore the use of FAIRsharing’s ontologies to refine and filter results

The first two suggestions are both embodied in the current version of the web application. Now, once users click “Visualise”, the web app reveals a small portion of the table while loading the graph below it in the same browser window. In addition, each row in the resulting tables offer hyperlinks to the first- and second-order organisations as well as the FAIR-enabling resources that provide the connection.

After exploring options for supporting the last suggestion of working with FAIRsharing, further work would be required. The subject ontology (SRAO) seems more promising, as it provides a broader, flatter hierarchy of subjects, when compared to the domain ontology (DRAO). Future development of the web app may be able to target and extract useful information from the SRAO.

# 2. Next Steps

For the remainder of the project, members of the BiCIKL consortium Iterative will use the graph to reveal latent connections outside the initial starting community and, where appropriate, act upon them by adding new services and cultivating new partnerships. As such, the work supports ongoing collaboration and dialogue with potential users and stakeholders, whether focused on addressing additional targeted user needs through training resources (WP3) or developing technical specifications to mobilise and interlink additional data resources or building entirely new services.

Additional stakeholder input of value to the project could include:

- refinement of the web app to optimise its use in visualising and explaining the consortium
- display of networks on the BKH to help explain existing or possible research paths to potential users

- recommendations for new collaborators create and/or update their own records on FAIRsharing as a first step toward any partnership
- enrichment of data available from FAIRsharing with other more technical data, including
    - access requirements for different types of data
    - constraints on data mining, linkage access and usage
    - benchmarks for ensuring data quality and applying accepted standards
    - compliance with FAIR practices

# 3. Acknowledgements

We wish to extend our thanks to all members of the BiCIKL consortium who have contributed their ideas and suggestions to the development of the relational network graph. Nicky Nicolson (RBG Kew) and David Shorthouse (AAFC), both of whom attended the May 2022 BiCIKL meeting in Seville, provided constructive feedback on the first iteration of this work.

# 4. References

Sansone SA, McQuilton P, Rocca-Serra P et al. (2019) FAIRsharing as a community approach to standards, repositories and policies. Nature Biotechnology 37: 358–367. https://doi.org/10.1038/s41587-019-0080-8

Smith VS, French L, Vincent S et al. (2022) Research Infrastructure Contact Zones: a framework and dataset to characterise the activities of major biodiversity informatics initiatives. ARPHA Preprints. https://doi.org/10.3897/arphapreprints.e82955

Wilkinson M, Dumontier M, Aalbersberg I et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. https://doi.org/10.1038/sdata.2016.18