Relational network graph for the Biodiversity Knowledge Hub (BKH)

Deliverable D2.1

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 Pensoft Publishers
Deliverable title:	Relational network graph for the Biodiversity Knowledge Hub (BKH)
Deliverable n°:	D2.1
Nature of the deliverable:	Other
Dissemination level:	Public
WP responsible:	WP2
Lead beneficiary:	Global Biodiversity Information Facility
Citation:	Copas, K. & Miller, J. (2022). *Relational network graph for the Biodiversity Knowledge Hub (BKH)*. Deliverable D2.1 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	25 July 2022	Kyle Copas & Joe Miller, GBIF
0.2	Draft	26 July 2022	Jerry Lanfear, ELIXIR, & Lyubomir Penev, Pensoft
0.3	Draft	27 July 2022	Christos Arvanitidis, LifeWatch Eric
1.0	Final	28 July 2022	Kyle Copas & Joe Miller, GBIF

Table of contents

Preface

The relational network graph for the Biodiversity Knowledge Hub (BKH) provides a lightweight, flexible framework for guiding the development of the BiCIKL starting community's content, its users, its partners and its stakeholders. Using an open-source web interface prepared as BiCIKL Milestone 3, consortium members can now conduct more detailed analyses aimed at engaging and enlisting new partners as well as expanding functional connections across a wider network of data infrastructures from the biodiversity and life sciences.

Summary

The network graph leverages an existing open data resource, FAIRsharing.org, to display an interactive catalogue that displays first- and second-order connections between BiCIKL project members and other organizations through a browser-based web application (currently located at https://gbif.shinyapps.io/bicikl-network-v2/). These connections provide the basis for BiCIKL community members to identify additional opportunities for intervention and deployment of innovation in data curation and enrichment. By evaluating, engaging and expanding on the possible interest of this wider network and its relevance to BiCIKL, the network graph can help guide the development of content for the BKH as well as the consortium's educational and outreach activities.

List of abbreviations

BGBM	Botanic Garden and Botanical Museum Berlin-Dahlem
BiCIKL	Biodiversity Community Integrated Knowledge Library
BKH	Biodiversity Knowledge Hub
CERN	European Organization for Nuclear Research
CETAF	Consortium of European Taxonomic Facilities
EMBL-EBI	European Molecular Biology Laboratory-European Bioinformatics Institute
EU	European Union
Europe PMC	European PubMed Central
FAIR	Findable, Accessible, Interoperable and Reusable
GBIF	Global Biodiversity Information Facility
LifeWatch	LifeWatch ERIC
MBG	Meise Botanic Garden
Naturalis	Naturalis Biodiversity Center
Pensoft	Pensoft Publishers
Plazi	Plazi
SIB	Swiss Institute of Bioinformatics
Sp2000	Species 2000
Tartu	University of Tartu
TDWG	Biodiversity Information Standards *formerly* the Taxonomic Databases Working Group

1. Goals and objectives

The relational network graph for the Biodiversity Knowledge Hub (BKH) provides a framework for guiding the development of the BiCIKL starting community's content, its users, its partners and its stakeholders.

Like the consortium, the BiCIKL network graph is rooted in biodiversity and life sciences, so what the tool seeks to reveal are the starting community's existing connections to wider areas of scientific and technical research. Leveraging an existing open data resource, FAIRsharing.org, the tool extends its data analysis and visualization capabilities and enables exploration of first- and second-order organizational connections behind the data standards, infrastructures and policies in FAIRsharing.

The network graph supports various tasks and work programmes (Tasks 1.2, 2.1, 2.2, WP3, JRA/WP10) by helping consortium members identify potential users, organizational partners and other stakeholders. The graph also supports targeted engagement of key prospects and communities in this wider network, whether as an input for developing BiCIKL content, a guide for partnerships needed to implement shared solutions, or a framework for prioritizing educational and outreach activities.

2. The network graph database

2.1. Open data on organizations from FAIRsharing

The BiCIKL network graph database 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 visualize the complex, interlocking web of connections between the data standards, infrastructures and policies—hereafter referred to collectively as "FAIRsharing objects." (Figures 1-3)

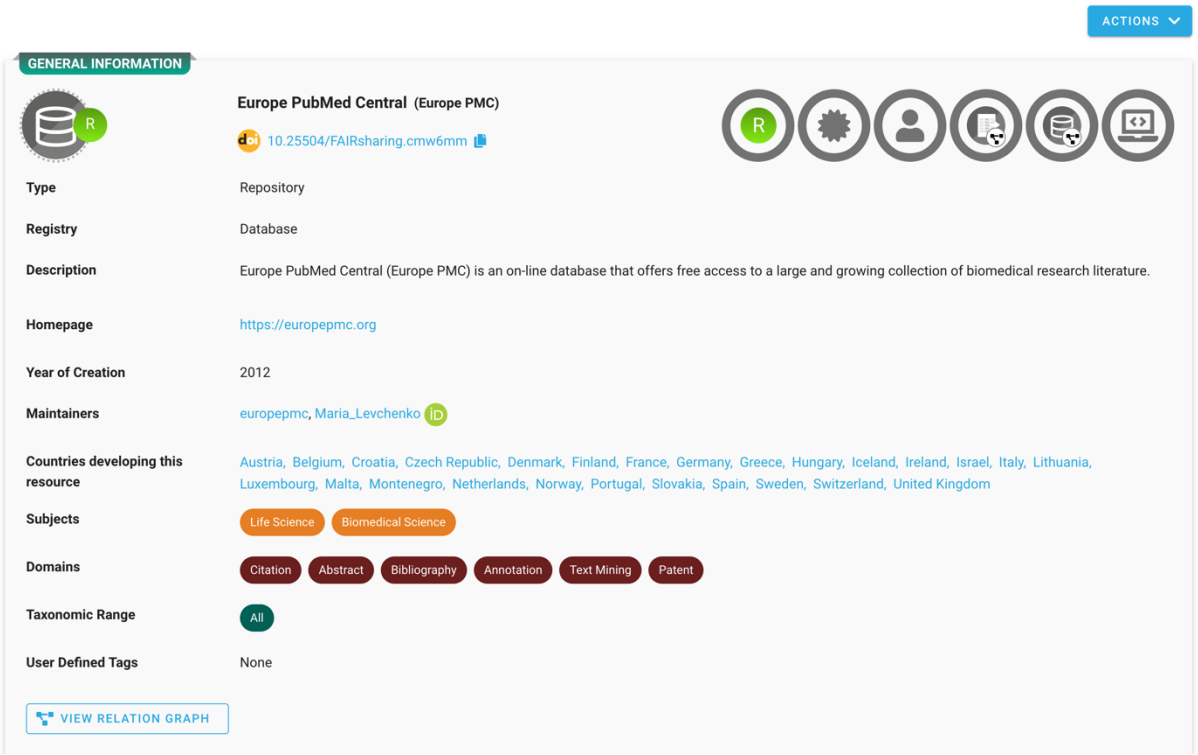

Fig. 1: *Sample record for Europe PMC shows the top-level information displayed on FAIRsharing.org. The record is maintained in part by EMBL-EBI staff.* https://fairsharing.org/FAIRsharing.cmw6mm

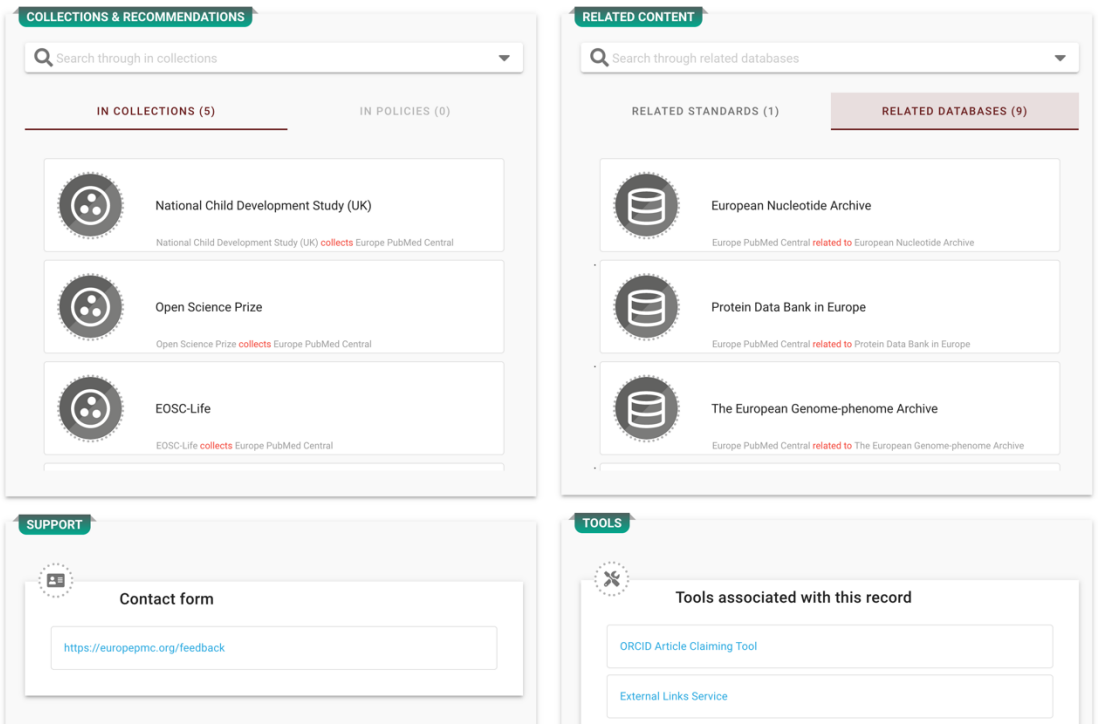

Fig. 2: *Tabular views of links of EuropePMC with other FAIRsharing objects.*

The declared connections with other FAIRsharing resources provide the basis for understanding Europe PMC's place in the FAIR-enabling landscape. https://fairsharing.org/FAIRsharing.cmw6mm

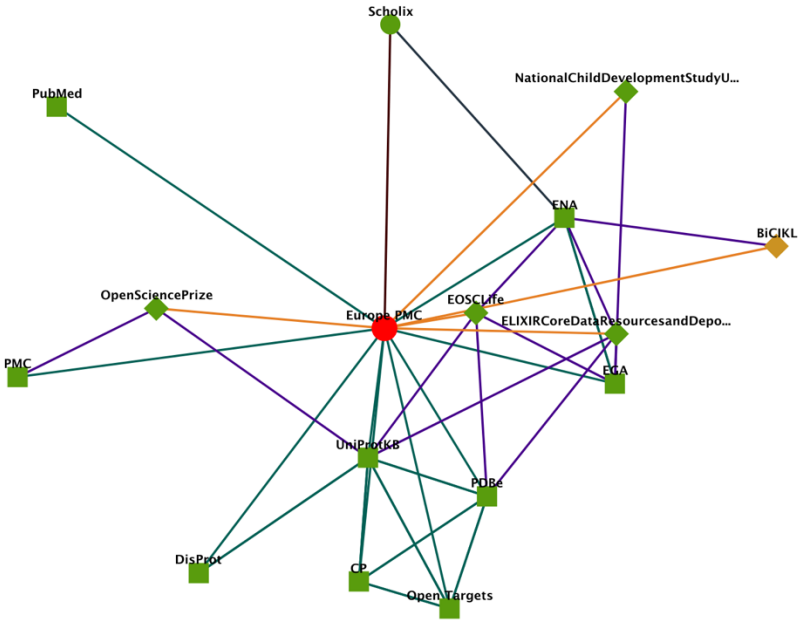

Fig. 3: *Sample force-directed graph detailing the different connections of EuropePMC to other databases, standards and collections on FAIRsharing.org. Complete legend and configuration details available at https://fairsharing.org/graph/1848.*

In addition to the declared connections to other FAIRsharing objects, each FAIRsharing record also links to an organization or set of organizations. BiCIKL consortium members agreed that this information had greater value for developing the network graph, because, by shifting the focus from connections between the individual FAIRsharing objects to the organizations that lay behind them, we could draw out links to a wider network of institutions. Tapping into this portion of FAIRsharing data thus promises to support in the ongoing work of:

- identifying and cataloguing relevant information about stakeholders
- documenting their interest in the BiCIKL's outputs
- engaging them as contributors to, or users of, the consortium's infrastructures

Use of this resource made sense on several levels. Applying existing data from a stable infrastructure is efficient and non-duplicative—especially considering that FAIRsharing is a flagship resource of BiCIKL consortium member ELIXIR whose development was supported in part through earlier EU funding from the Horizon 2020 programme (H2020-EU.3.1, 634107, H2020-EU.1.4.1.3, 654241, H2020-EU.1.4.1.1, 676559). In addition, several BiCIKL partners had previous experience claiming and maintaining FAIRsharing records associated with their infrastructures, making it easy to help the others (where necessary) to create, adopt and maintain records for their own resources. Finally, by curating these records, BiCIKL consortium members could contribute directly toward refining and improving the quality and accuracy of FAIRsharing data.

Some technical hurdles did intervene, however. While organizational data is documented as a first-order data object in the FAIRsharing schema, it was (and is still) not displayed at the same level of detail on FAIRsharing.org as the other objects. We also discovered that the permissions model for the API limited our access to this data. But through discussions and correspondence with FAIRsharing staff, who were keen to support an external use of the data, they quickly implemented the necessary improvements to the FAIRsharing API to resolve these issues.

With full access to organizational data from FAIRsharing enabled, we pursued one final preparatory step. After extracting the relevant FAIRsharing objects and organizations into a local database, we transformed the connections between the former into inferred ones for the latter and loaded FAIRsharing data for more than 3,300 organizations into an interactive web application.

2.2. Web application for the relational network graph

The browser-based web app for accessing the relational network graph (https://gbif.shinyapps.io/bicikl-network-v2) is an R-based Shiny app, technologies that are widely adopted across the biodiversity informatics community. Its source code will be hosted by GBIF in a publicly accessible GitHub repository (available shortly from https://github.com/gbif), where consortium partners and others can submit issues, comments and requests for enhancements. The code will also be released under an open licence, enabling others to access and repurpose it for use in similar analyses.

The web app extends the functionality of FAIRsharing.org in several ways. Firstly, it infers, expresses and visualizes the links between the organizations that maintain, collaborate or are otherwise involved with the resources described in FAIRsharing (Figure 4). Secondly, it shifts the focus from connections between the individual data standards, databases and data policies to the organizations responsible for these resources, revealing a wider network of institutional relationships. Finally, it extends the visualizations of FAIR-enabling organizations to a second order of connections, displaying the web of transitive relations between the institutions responsible for or associated with FAIRsharing objects (Figure 5).

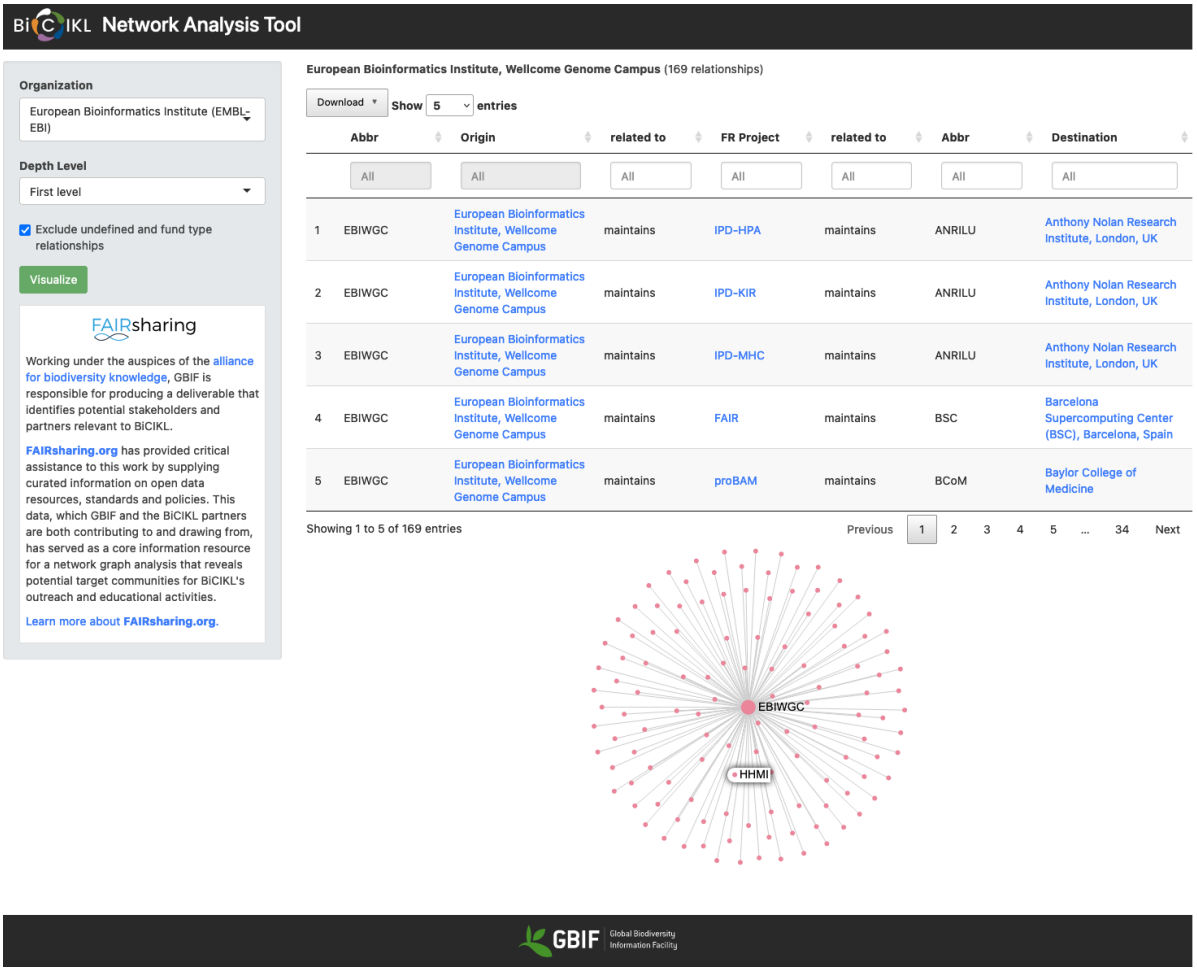

Fig. 4: *Sample first-level organizational record from the network graph web app, displaying first-level connections between EMBL-EBI and other organizations that maintain, collaborate on or are associated with other FAIRsharing objects.*
https://gbif.shinyapps.io/bicikl-network-v2/

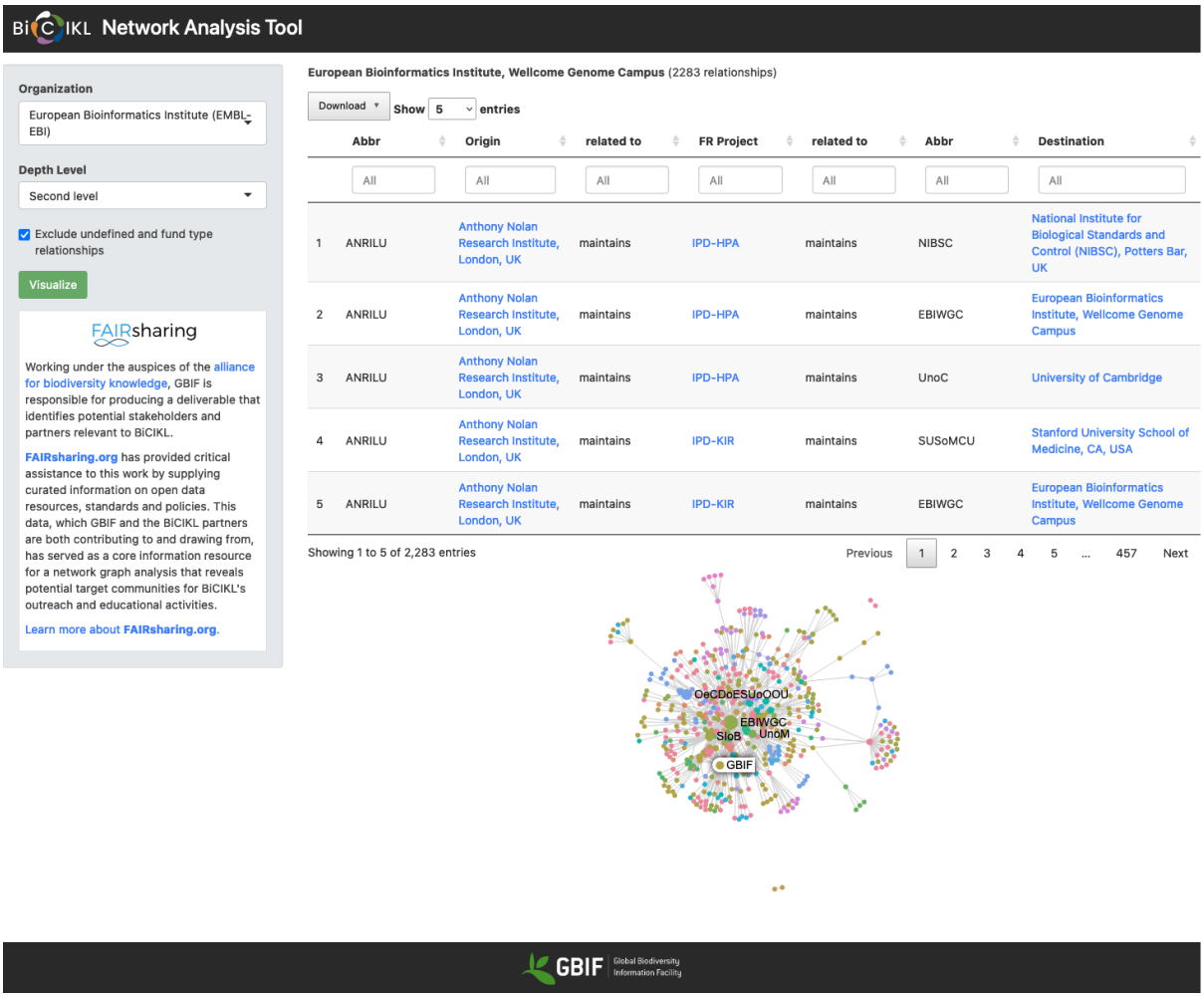

Fig. 5: *The web app extends the relational network graph displayed on FAIRsharing.org to show second-level connections between BiCIKL partners (e.g. EMBL-EBI) and other organizations. https://gbif.shinyapps.io/bicikl-network-v2/.*

Users generate network graphs on the web app by

1. Selecting the BiCIKL consortium partner
2. Selecting whether to view the first- or second-order relationships
3. Clicking “Visualize”

By default, the web app displays three types of relationships between the organization and associated FAIRsharing records: ones that “maintain”, “collaborate on” or are “associated with” FAIRsharing objects. Users who wish to see a network graph that includes organizations that “fund” or have an “undefined” relationship can uncheck the box above the visualize button.

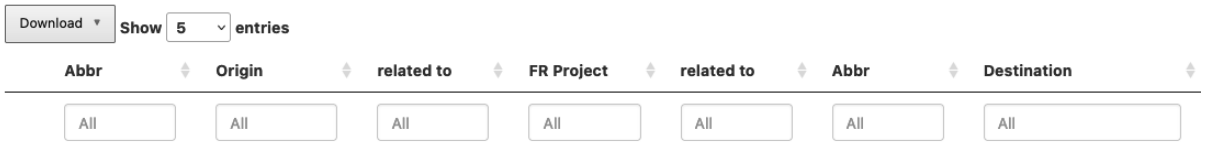

Fig. 6: *A customizable table view allows users to filter search results across all columns for any network graph displayed on the web app. https://gbif.shinyapps.io/bicikl-network-v2/*

Users can customize the view at the top of the table to show more rows and narrow their searches across all of the tables' columns to further filter results (Figure 6). Connections to the original information in FAIRsharing are also readily available through hyperlinks to either the related organizations or the FAIRsharing object that serves as its source (Figure 7a-7c).

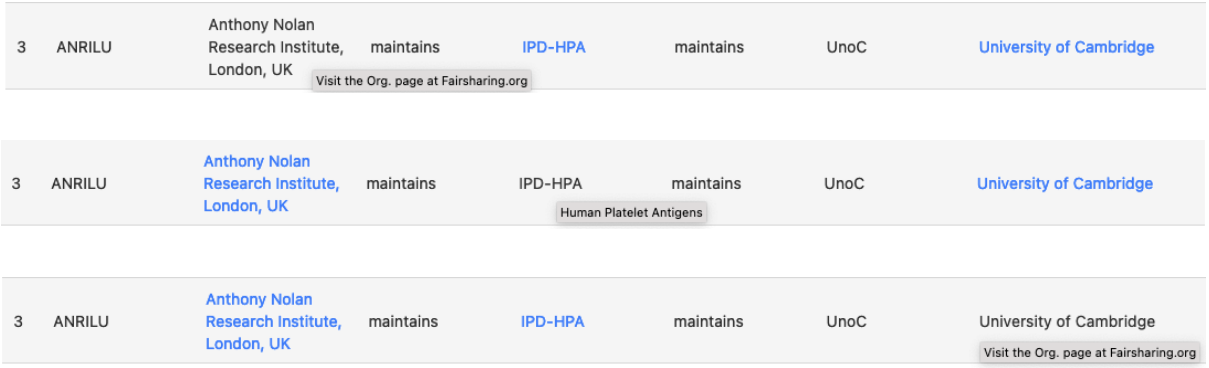

Fig. 7a, 7b, 7c: *The web app preserves FAIRsharing provenance for any result in the table by providing users with access to the original FAIRsharing records for (**7a**: top) the first-order organization and (**7b**: middle) the FAIRsharing object that links to (**7c**: bottom) the second-order organization. https://gbif.shinyapps.io/bicikl-network-v2/*

These links to FAIRsharing sources in the network graph will enable quick identification of possible partners across adjacent research subdomains along with easy access to basic information about them. In this way, the graph will help the BiCIKL consortium
target and tailor development and outreach activities to fit potentially underserved but relevant areas of research.

In addition to supporting prioritization of other activities, the patterns of linkages reflected in the network graph provide a preliminary (if coarse) baseline set of metrics that signal the current range and breadth of the BiCIKL starting community (Table 1). The change in the number of organizational connections among the consortium partners, both individually and collectively, will provide a simple but indicative metric for charting the community's growth over time.

Table 1: *Baseline counts of organizational connections between BiCIKL and other FAIRsharing resources.*

BiCIKL organization	First-order connections	Second-order connections
BGBM	30	675
CERN	23	597
CETAF	25	651
EMBL-EBI	169	2,283
GBIF	59	1,225
LifeWatch	13	530
MBG	25	651
Naturalis	26	656
Pensoft	20	557
Plazi	15	530
SIB	73	1,142
Species 2000	24	612
Tartu	14	544
TDWG	14	531

With their individual and collective commitments to support research and policy addressing the twin crises around biodiversity and climate change, the web app's design dramatically reduces its operating footprint to only the minimum required resources. Data from FAIRsharing is updated and processed once a day, storing data locally within the web app to reduce API calls and associated bandwidth usage. Use of hosted services from shinyapps.io also provides web-app users with secure, scalable and efficient access to the tool.

3. Next steps

By leveraging community-curated data from an external registry with stable infrastructure, BiCIKL partners have both a stake in and method for maintaining the web app. Each consortium member has staff designated as maintainers of their records and in several cases have identified additional resources to add to FAIRsharing. Working as a team of approved curators, the BiCIKL consortium can improve the quality and accuracy of FAIRsharing data and ensure its sustainability as a resource for their efforts and those of other researchers.

For the remainder of the project, members of the BiCIKL consortium Iterative can 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, it supports collaborative dialogue with potential users and stakeholders, whether they support needs of targeted users through training resources (WP3) or develop technical specifications to mobilize and interlink additional data resources or building entirely new services.

While designed as a tool to help prioritize the project team's wider outreach and development activities around the BKH, the use of FAIRsharing data in the web app also suggests other possible extensions. For example, among other opportunities for applying the information that the tool provides, BiCIKL members may wish to use the visualizations to present and explain the consortium. Displaying the networks that link the partners as part of the BKH may help explain existing or possible research paths to potential users.

The web app should also prompt consortium members to consider whether FAIRsharing data can serve as part of the BKH infrastructure. In such an event, the team could recommend that new collaborators create and/or update their own records on FAIRsharing as a first step toward any partnership. Wider deployment of FAIRsharing data could also provide the basis for capturing more precise technical information such as:

- 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

4. 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 of Royal Botanic Gardens Kew and David Shorthouse, both of whom attended the May 2022 BiCIKL meeting in Seville, provided constructive feedback on the first iteration of this work.

We are also grateful for the invaluable assistance provided by the FAIRsharing team—particularly Allyson Lister, Delphine Dauga, Milo Thurston, and Susanna-Asunta Sansone—as well as our application development team from Extendo, especially Carlos Pravia, Paul Fervoy and Gloria Edith Rodríguez Ramírez.

5. 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

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