Session 2: Socio-economic and environmental impact of RI's:

ELIXIR's approach to enabling socio-economic impact in the context of digitalisation and AI

Tim Hubbard
ELIXIR Director

Tuesday 04 June 2024

www.elixir-europe.org

@ELIXIREurope

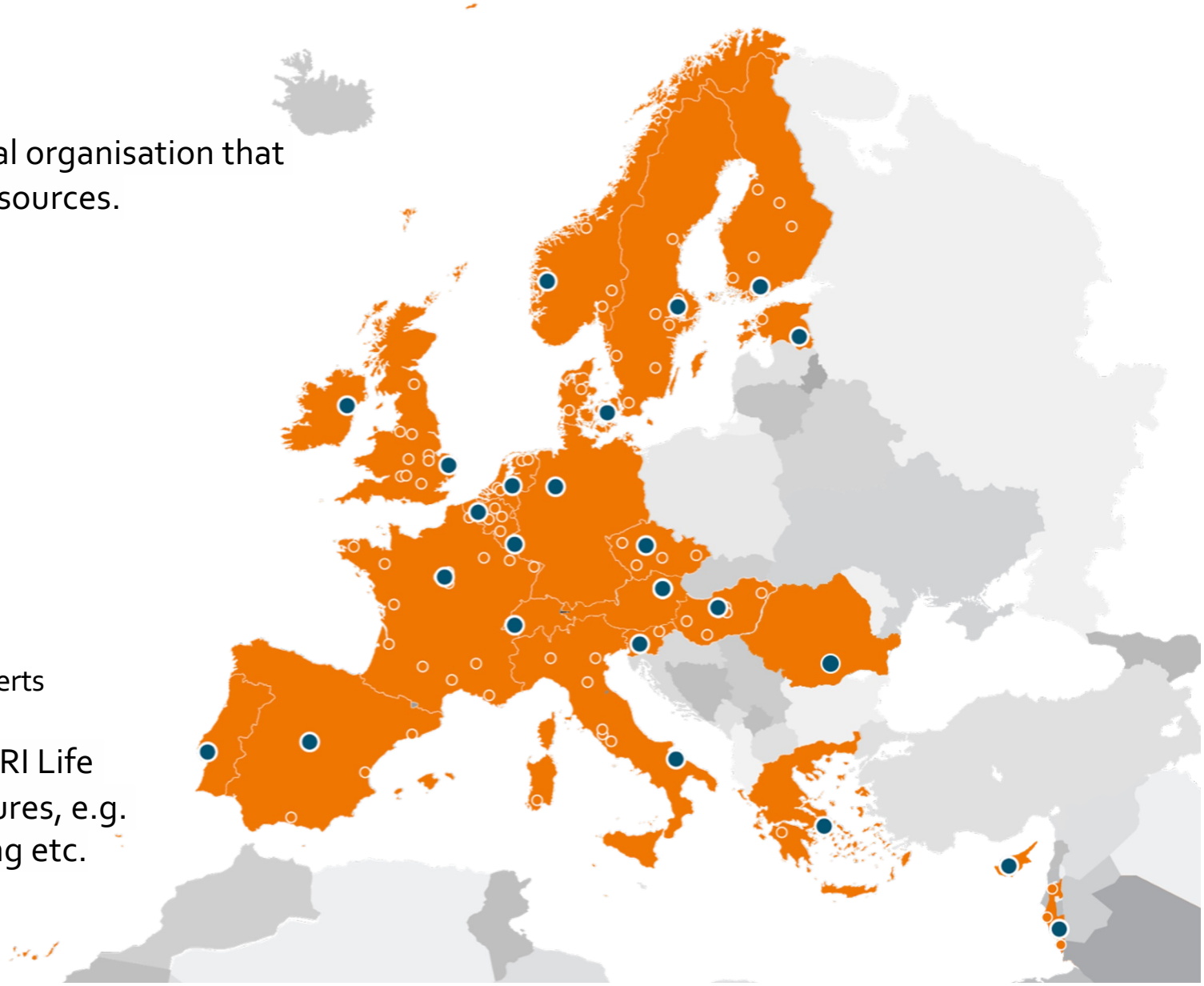
/company/elixir-europe

ELIXIR Europe

ELIXIR is an intergovernmental organisation that brings together life science resources.

- Databases
- Training
- Software tools
- Data standards
- Compute resources
- Scientific & technical experts

Works closely with other ESFRI Life Science Research Infrastructures, e.g. BBMRI-ERIC, Euro-BioImaging etc.



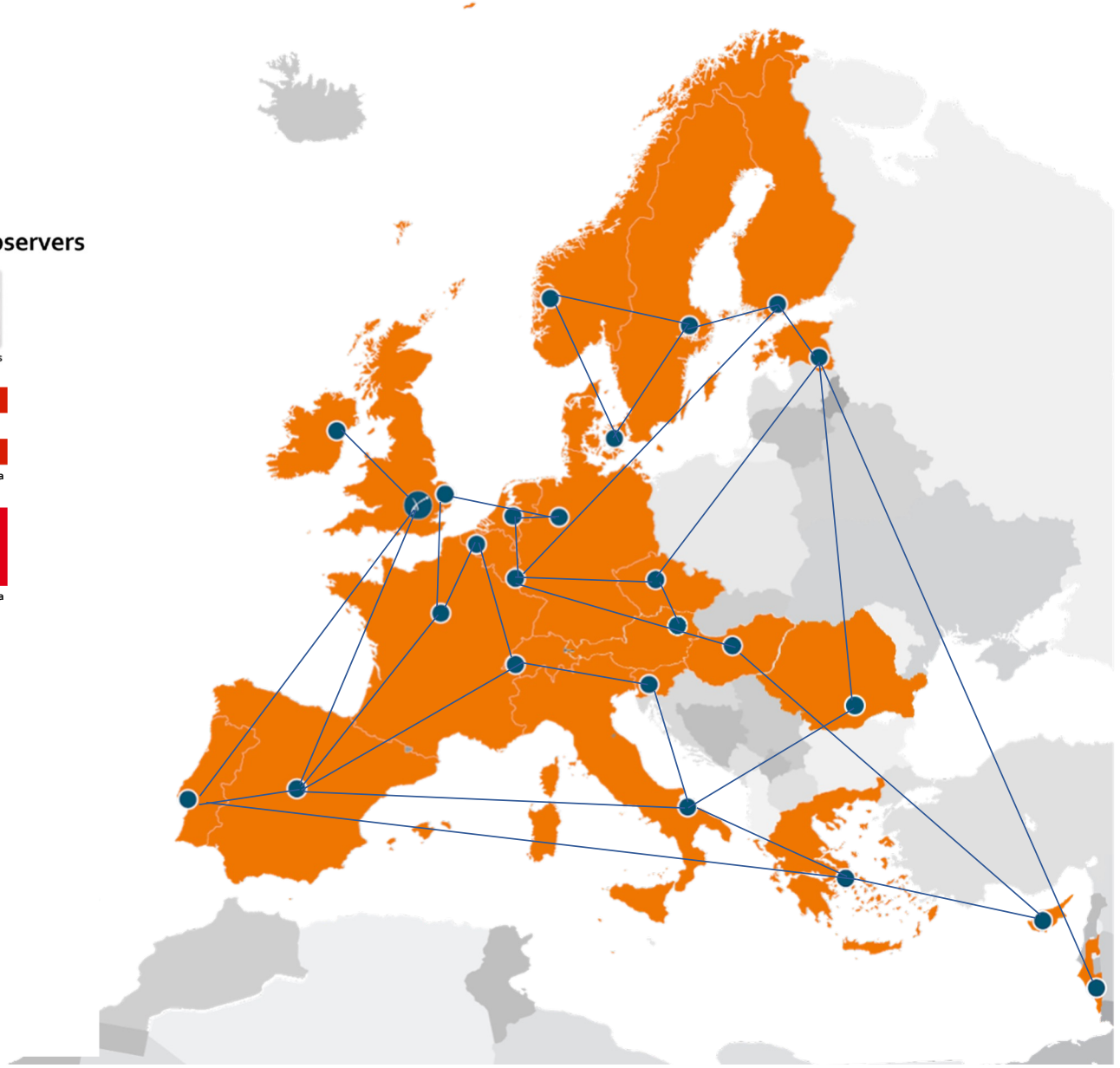
ELIXIR Europe

ELIXIR Members

Belgium
Czech Republic
Denmark
EMBL
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Luxembourg
Israel
Netherlands
Norway
Slovenia
Portugal
Spain
Sweden
Switzerland
United Kingdom

ELIXIR Observers

Cyprus
Austria
Romania



Societal challenges….

Disease prevention

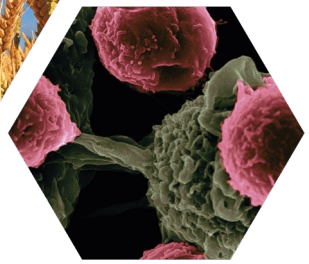
Climate change

Diminishing biodiversity

Disease diagnosis

Sustainable food production

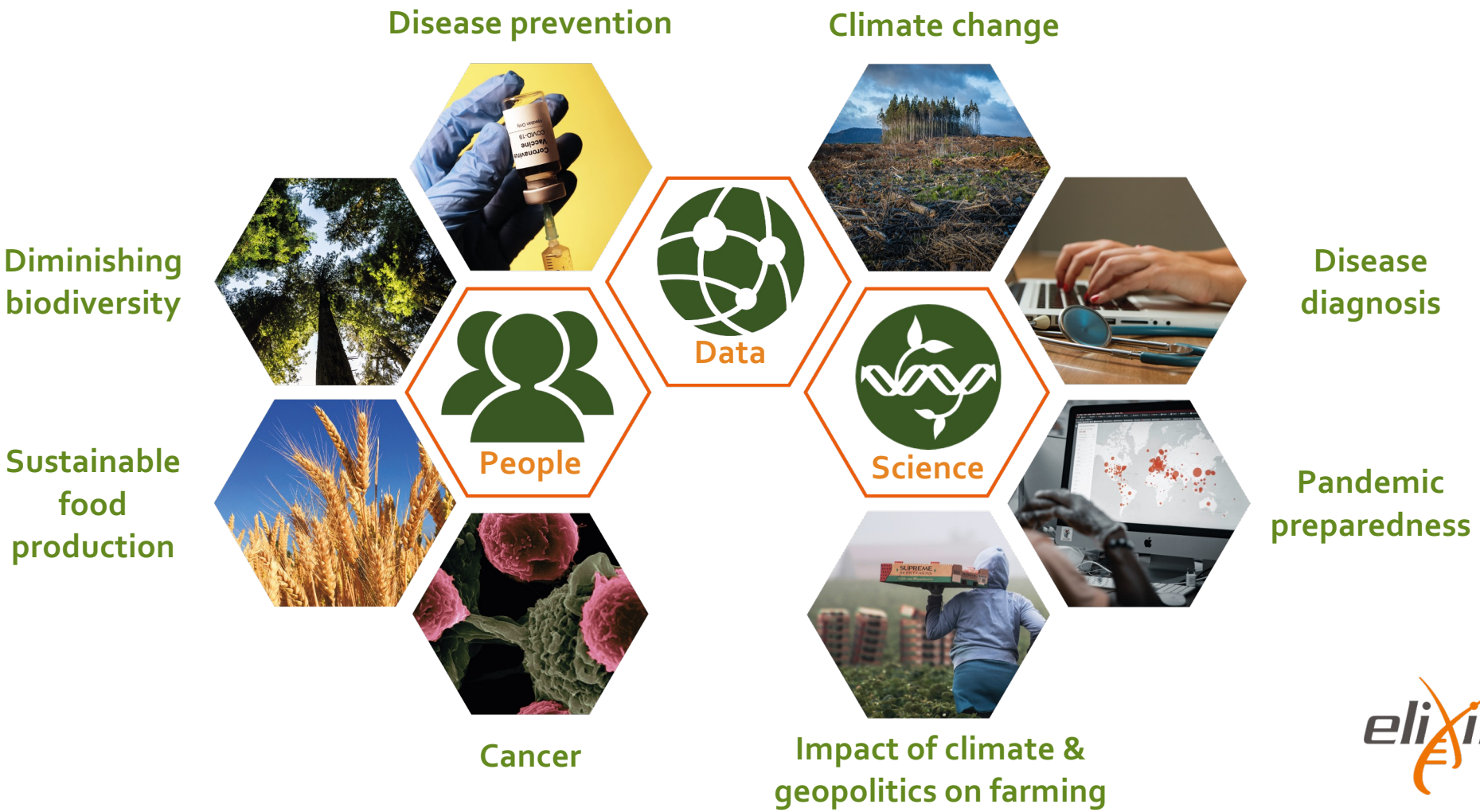
Pandemic preparedness



Cancer

Impact of climate & geopolitics on farming

...and data-driven solutions



Data management challenges in life sciences

Data growth

Data diversity

Data location

Data sensitivity

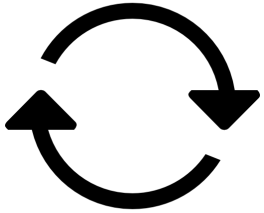
Artificial Intelligence

SCIENCE

Human data and translational research

Cellular & molecular research

Biodiversity, food security & pathogens



Common & Shared Technology

TECHNOLOGY

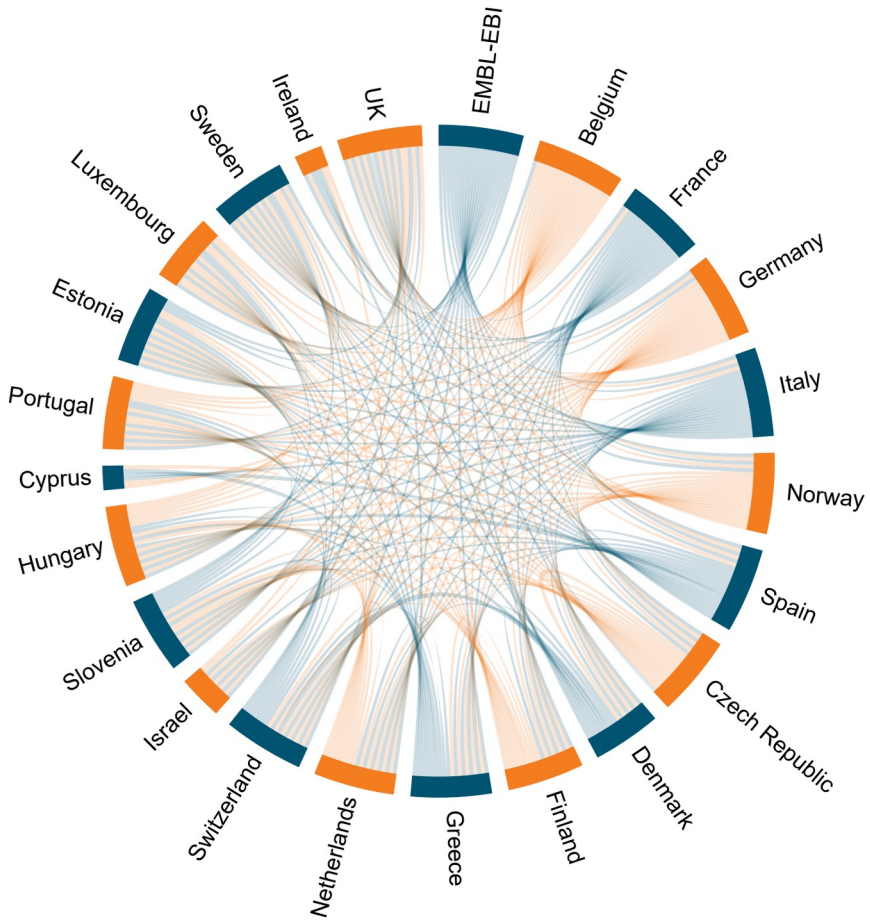
Research data management & knowledge sharing

Reproducible analytics & infrastructure

Federated service delivery

elixir

ELIXIR Commissioned Services



https://elixir-europe.org/about-us/commissioned-services

Commissioned Services are technical projects that guide future service development, drive standards adoption and connect ELIXIR Nodes

Building a single infrastructure by connecting Europe

- Over **120 projects** supported by **EUR 26 million**
- Improving **collaboration between ELIXIR Nodes**
- **Connecting, integrating and sustaining** life science data resources
- **Developing and improving services** used by life scientists globally

ELIXIR Communities – connecting infrastructure & life science experts

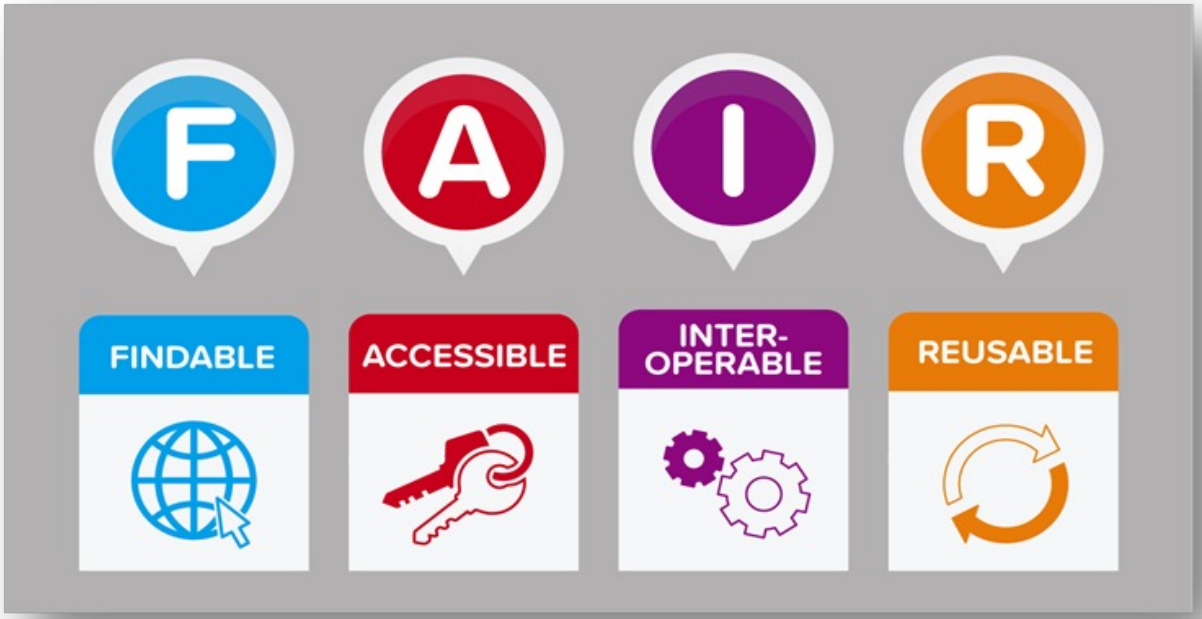
Formed around domain experts in ELIXIR Nodes (including non-ELIXIR partners)

Provide a mechanism for long-term collaborations with other ESFRIs and large-scale initiatives

Drive service developments in the ELIXIR Platforms

Provide a framework to develop and maintain community standards

The ELIXIR Communities Handbook tells you what a Community is, who can join, what the benefits are, and how Communities are structured.



Reproducible

ELIXIR's support for FAIR data and software

FAIR data services & resources

Open registries, ontologies, identifiers, data management platforms, stewardship tools, data FAIRification methodology, standards

Trusted data resources

Open deposition databases and portals, scalable curation, sustainability

Data analytics & platforms

Workflows, reproducible and portable processing, software and AI best practice, FAIR assessment, federated analytics

Open & FAIR policy/advocacy

FAIR principles, FAIR leadership & partnering at the global, European and national level

Specific communities

Human Data, Structural Bioinformatics, Rare Diseases, Plant Sciences, Microbial Biotechnology, Proteomics, Metagenomics, Systems Biology…

Stewardship and training

Capability frameworks, skills, data managers network, training portal

Open Data drives innovation

76% of respondents stated that without data shared on open repositories, they would not be able to offer their product or service.

89% of respondents stated that a product or service has more features because of access to data shared on open repositories

63% 63% of respondents stated that without access to registries, ontologies, and dictionaries published on open repositories, they would not be able to offer their product or service.

92% of respondents stated that a product or service has more features because of access to registries, ontologies, and dictionaries shared on open repositories.

https://elixir-europe.org/sites/default/files/documents/sme-report-2021.pdf

ELIXIR's innovation and industry programme

External outreach

Since 2014	17 SME forums in...	12 European countries with over...	~1,000 participants from...	308 companies

Engagement

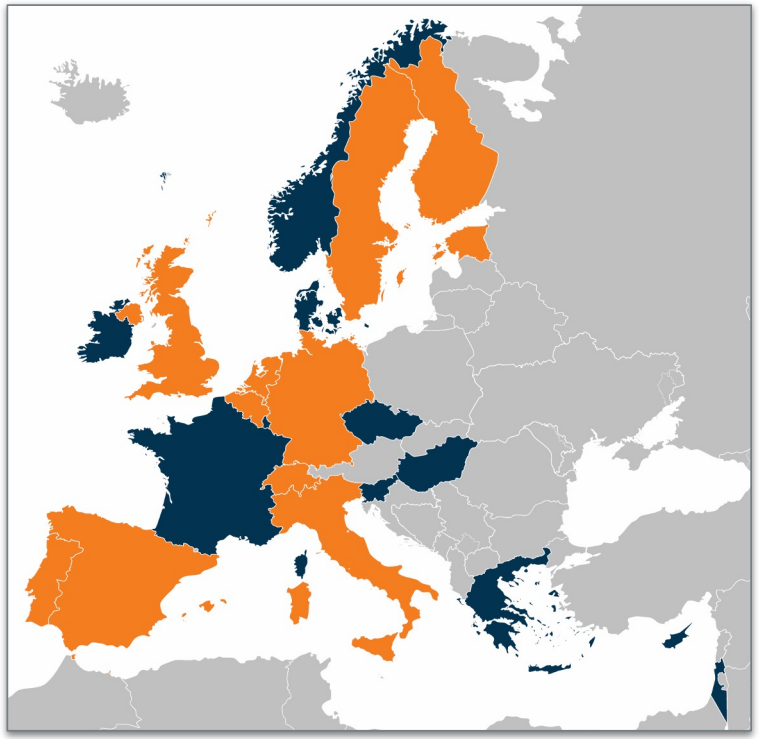
41% Nodes with industry engagement, more coming online

38% Communities with industry engagement, more coming online

Patents

>11,000 patent applications filed mentioning ELIXIR's resources

22 countries in which these patent applications have been filed



Resources to access knowledge and curated digital objects

bio.tools

bio.tools helps you find and select bionformatics software and connect it in workflows.

BioContainers

Search a repository of containerised software that you can build into workflows.

WorkflowHub

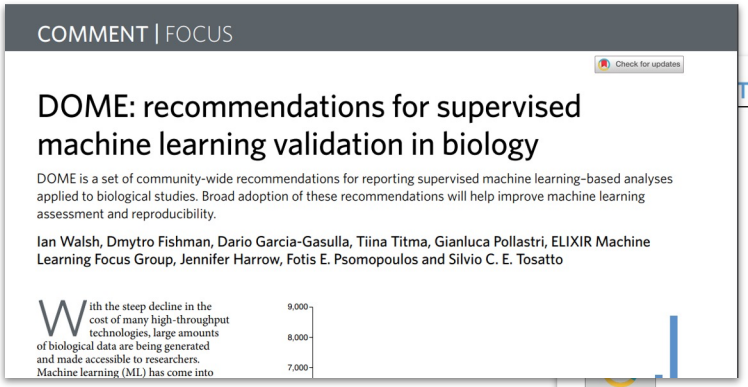
A registry for sharing and publishing scientific computational workflows.

FAIRsharing.org

FAIRsharing.org allows you to search for databases and data policies by aspects such as domain, species and country.

TeSS

Search for training courses, webinars, training materials and workflows in TeSS, ELIXIR's training portal.



COMMENT | FOCUS

Check for updates

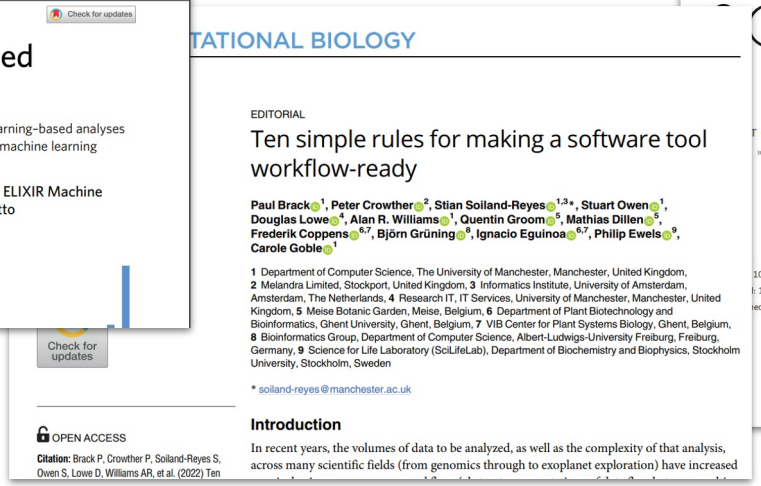
DOME: recommendations for supervised machine learning validation in biology

DOME is a set of community-wide recommendations for reporting supervised machine learning–based analyses applied to biological studies. Broad adoption of these recommendations will help improve machine learning assessment and reproducibility.

Ian Walsh, Dmytro Fishman, Dario Garcia-Gasulla, Tiina Titma, Gianluca Pollastri, ELIXIR Machine Learning Focus Group, Jennifer Harrow, Fotis E. Psomopoulos and Silvio C. E. Tosatto

With the steep decline in the cost of many high-throughput technologies, large amounts of biological data are being generated and made accessible to researchers. Machine learning (ML) has come into

9,000
8,000
7,000



TATIONAL BIOLOGY

EDITORIAL

Ten simple rules for making a software tool workflow-ready

Paul Brack[1], Peter Crowther[2], Stian Soiland-Reyes[1,3]*, Stuart Owen[1], Douglas Lowe[4], Alan R. Williams[1], Quentin Groom[5], Mathias Dillen[5], Frederik Coppens[6,7], Björn Grüning[8], Ignacio Eguinoa[6,7], Philip Ewels[9], Carole Goble[1]

1 Department of Computer Science, The University of Manchester, Manchester, United Kingdom, 2 Melandra Limited, Stockport, United Kingdom, 3 Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands, 4 Research IT, IT Services, University of Manchester, Manchester, United Kingdom, 5 Meise Botanic Garden, Meise, Belgium, 6 Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, 7 VIB Center for Plant Systems Biology, Ghent, Belgium, 8 Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Freiburg, Germany, 9 Science for Life Laboratory (SciLifeLab), Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

* soiland-reyes@manchester.ac.uk

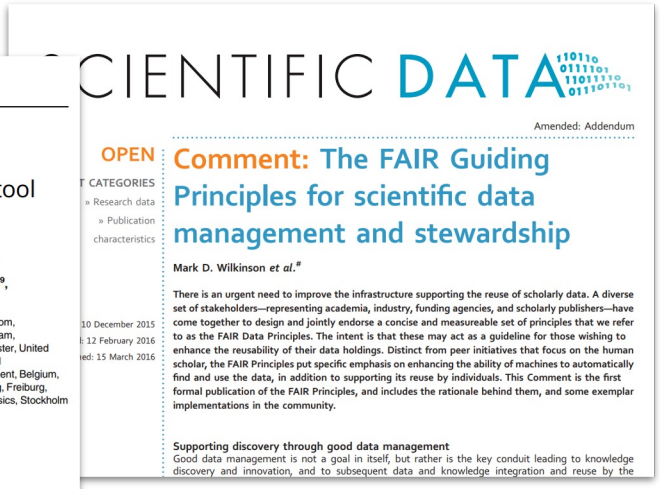
Check for updates

OPEN ACCESS

Citation: Brack P, Crowther P, Soiland-Reyes S, Owen S, Lowe D, Williams AR, et al. (2022) Ten

Introduction

In recent years, the volumes of data to be analyzed, as well as the complexity of that analysis, across many scientific fields (from genomics through to exoplanet exploration) have increased



CIENTIFIC DATA

Amended: Addendum

OPEN

T CATEGORIES

» Research data

» Publication characteristics

10 December 2015
: 12 February 2016
ed: 15 March 2016

Comment: The FAIR Guiding Principles for scientific data management and stewardship

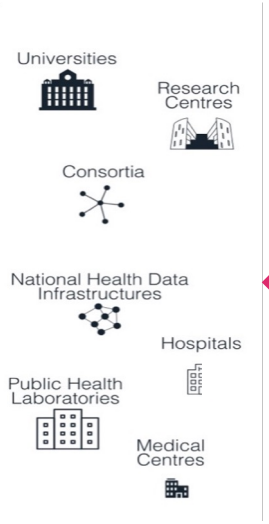
Mark D. Wilkinson *et al.*#

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

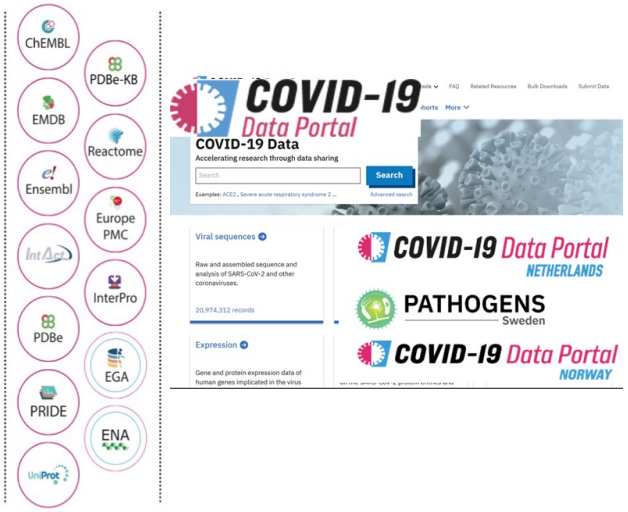
Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the

Infectious Disease data mobilisation mission



243 SARS-CoV-2 Data Hubs mobilised.



connected datasets containing data from 3m+ samples.

EMBL-EBI

Standardised metadata framework index across infectious disease resources & collect provenance.

samples in COVID-19 data portal **analysed, genomes annotated, automated monitoring** for allelic-variant surveillance, feeding dashboards, **distributed and federated analytics**

Guidance for pathogen characterisation, socioeconomic data, human biomolecular data, human clinical and health data

BY-COVID

elixir

Changing the source of data inputs for research

- Old sources of data:
  - 'cohorts' of data collected for specific research purpose;
  - mostly open, possible to centralise in databases (e.g. at EBI)
- New sources of data:
  - routinely collected data across society, industry: cheap enough to collect *everything*
  - maybe restricted, may be limited to national location and control

- Opportunity:
  - Routinely collected data noisy, but complementary and datasets much larger, good for AI
  - If data has already been collected and paid for, marginal cost in enabling research access
  - "Data visiting" rather than "data distribution" has a lower climate footprint
- Challenges:
  - Privacy, commercial ownership, culture, federation

EC understands the idea

English

Search

Shaping Europe's digital future

Home | Policies | Activities | News | Library | Funding | Calendar | Consultations | AI Office

Home > Policies > Common European Data Spaces

Common European Data Spaces

Common European Data Spaces will make more data available for access and reuse. This will be done in a trustworthy and secure environment for the benefit of European businesses and citizens.

https://digital-strategy.ec.europa.eu/en/policies/data-spaces

Realising a practice of personalised medicine & health

Long-term strategy: cross-border access to genomic data, implementation of genomics-based health

1+MG Group, National Mirror Groups and Thematic Working Groups, Use Cases Working Groups: cancer, infectious diseases, rare diseases, common complex diseases, industry

Design and testing

1+MG Trust Framework

- Data governance
- Standards and Quality
- Infrastructure
- Maturity model

Scale up and sustainability

EHDS

Healthdata@EU Pilot

1+MG Cancer Use Case

European Genomic Data Infrastructure

elixir

Convening a Data Ecosystem for Sensitive Data

Federated Service Delivery

Secure cross-border access to national genomic data and corresponding clinical data
data visited, metadata travels

distributed & federated analytics
1+MG Framework toolkit

Global Alliance for Genomics & Health

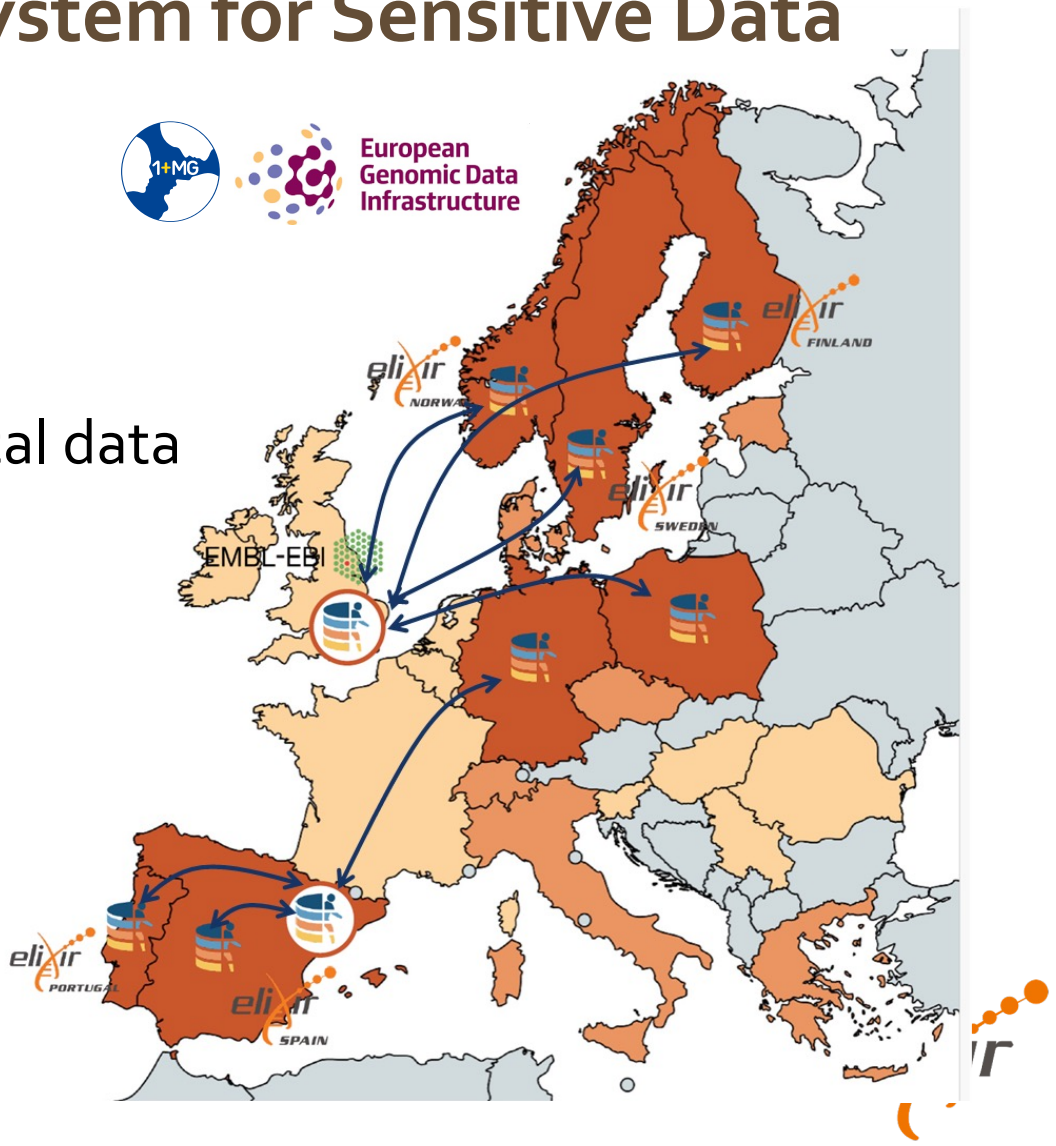
Beacon

Data Use Ontology

EUROPEAN GENOME-PHENOME ARCHIVE

EMBL-EBI

https://framework.onemilliongenomes.eu/



Changing the health outputs that Infrastructure enables

- Old translational mechanism:
  - New drug in 20+ years
- New translational mechanism:
  - Algorithms that can be applied immediately in direct care for individuals

- Opportunity:
  - EU countries spend about 10% GDP on healthcare
  - Public spending on biomedical and health research is only around 0.2% GDP*
- Challenges:
  - Privacy, commercial ownership, culture, federation

*Lancet, 2015

Biological Information +Artificial Intelligence
- *modelling in reality*

Ecosystem, system of interacting organisms

↑ predict

Organism, system of interacting organs

↑ predict

Organ, system of interacting cells

↑ predict

Cell, living system of DNA, RNA, Proteins

↑ predict

Protein (translations of genes, 3D structure)

↑ Predict (alphafold)

RNA (transcripts of active genes)

↑ predict

DNA (copy of genetic material in every cell)

← Environmental, ecosystem data

← whole organism data ← Clinical Medical Records

← organoid data

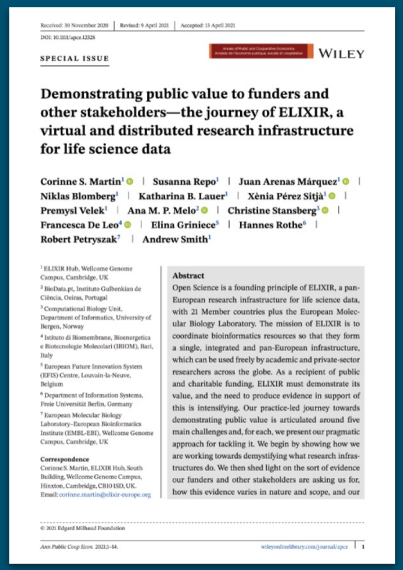
← single cell data

← Xray, NMR, cryoEM structures; MassSpec

← Expression, other omics

← Genomes ← Clinical Genome Sequences

Main categories of direct impact for work funded by and through ELIXIR



SPECIAL ISSUE

WILEY

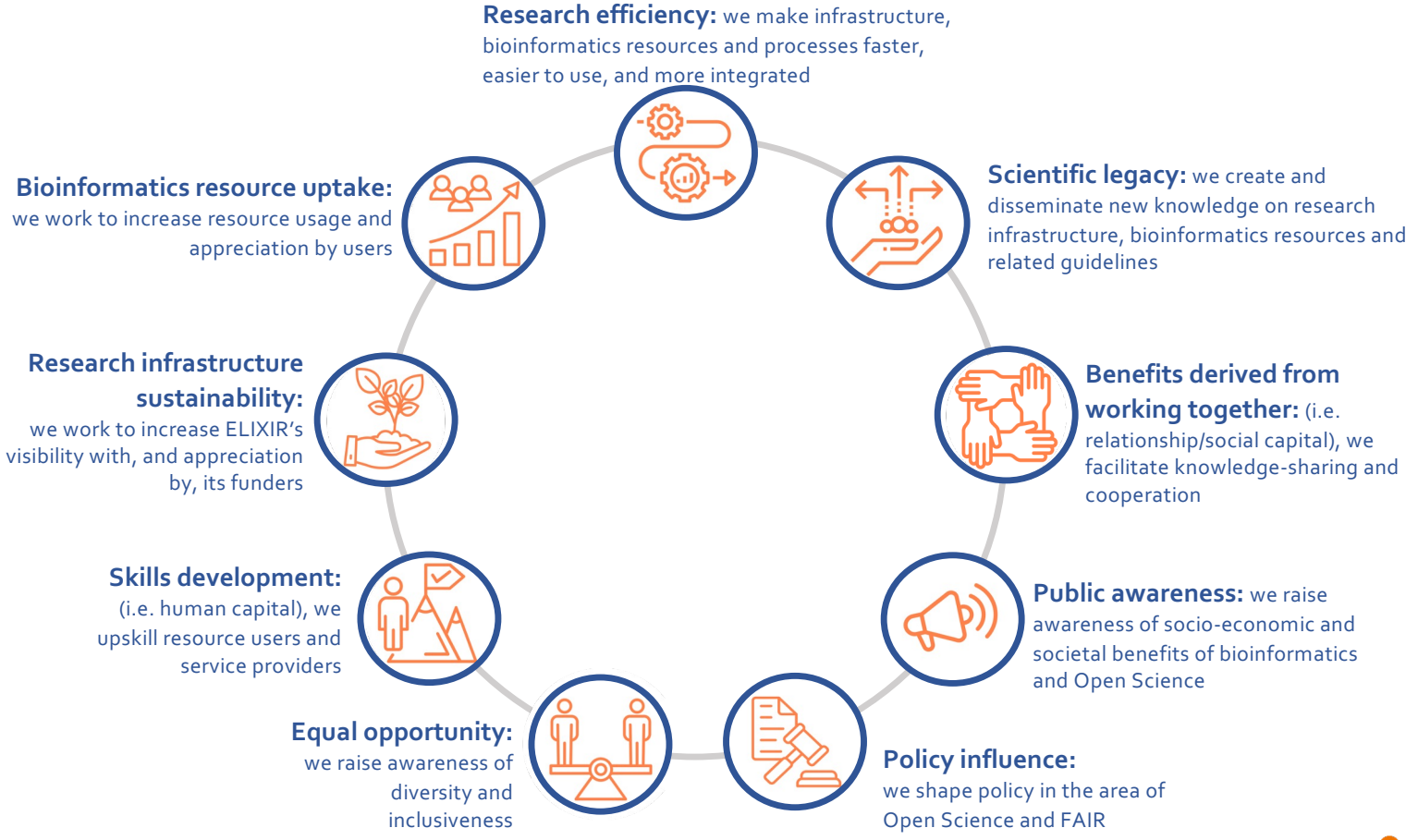
Demonstrating public value to funders and other stakeholders—the journey of ELIXIR, a virtual and distributed research infrastructure for life science data

Corinne S. Martin | Susanna Repo | Juan Arenas Márquez | Niklas Blomberg | Katharina B. Lauer | Xènia Pérez Sitjà | Premysl Velek | Ana M. P. Melo | Christine Stansberg | Francesca De Leo | Elina Griniece | Hannes Rothe | Robert Petryszak | Andrew Smith

Abstract

Open Science is a founding principle of ELIXIR, a pan-European research infrastructure for life science data, with 21 Member countries plus the European Molecular Biology Laboratory. The mission of ELIXIR is to coordinate bioinformatics resources so that they form a single, integrated and pan-European infrastructure, which can be used freely by academic and private-sector researchers across the globe. As a recipient of public and charitable funding, ELIXIR must demonstrate its value, and the need to produce evidence in support of this is intensifying. Our practice-led journey towards demonstrating public value is articulated around five main challenges and, for each, we present our pragmatic approach for tackling it. We begin by showing how we are working towards demystifying what research infrastructures do. We then shed light on the sort of evidence our funders and other stakeholders are asking us for, how this evidence varies in nature and scope, and our

Martin et al. (2021) Demonstrating public value to funders and other stakeholders — the journey of ELIXIR, a virtual and distributed research infrastructure for life science data. *Ann Public Coop Econ*, **00**: , 1-14. https://doi.org/10.1111/apce.12328



Adapted from Martin et al. (2021)
https://doi.org/10.1111/apce.12328

Acknowledgements

- ELIXIR Nodes and Hub
- EMBL-EBI
- Carole Goble

www.elixir-europe.org

@ELIXIREurope

/company/elixir-europe