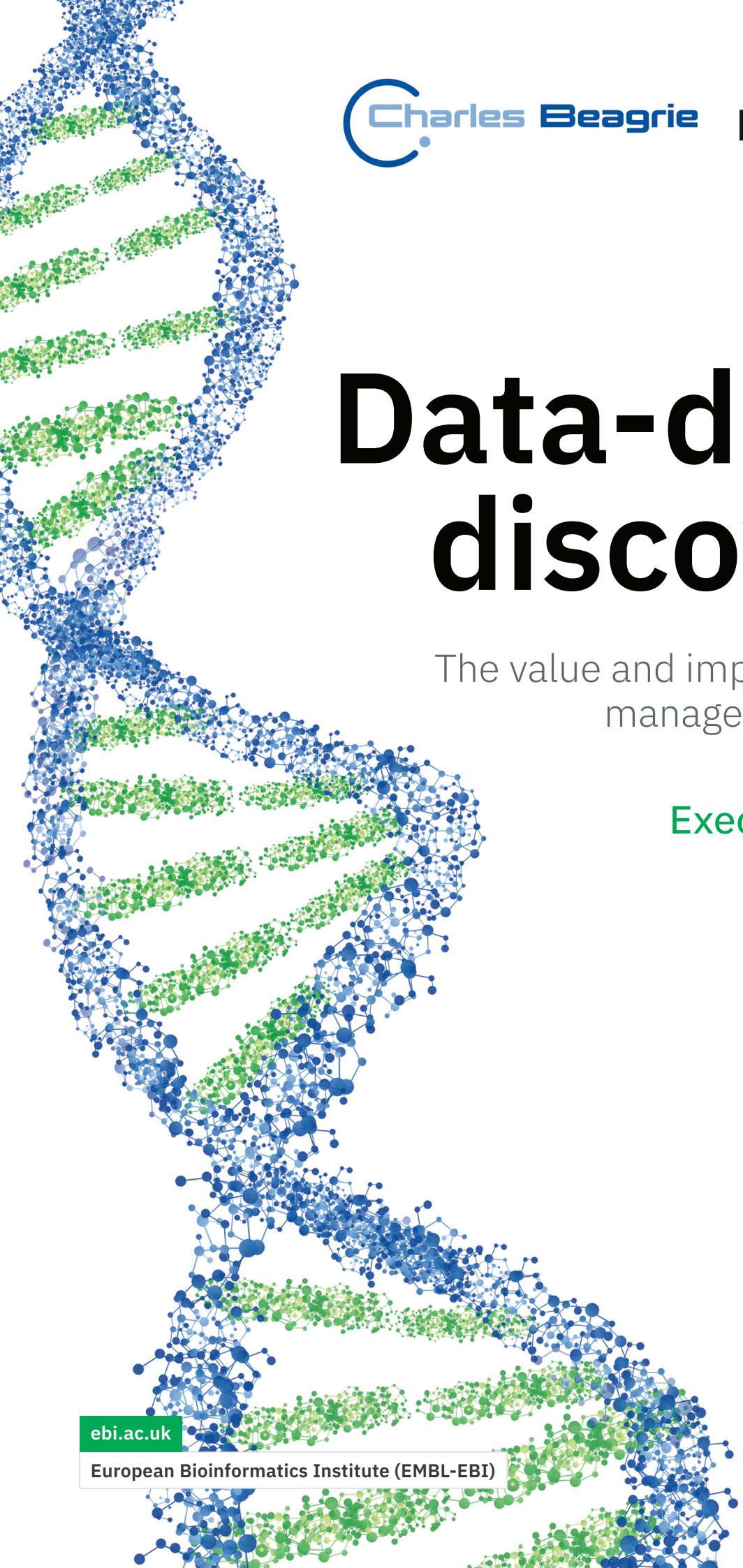


Data-driven discovery:

The value and impact of EMBL-EBI managed data resources

Executive summary

Neil Beagrie & John Houghton
Charles Beagrie Ltd
www.beagrie.com

ebi.ac.uk
European Bioinformatics Institute (EMBL-EBI)

About this publication

Cover image credit: AdobeStock image edited by Karen Arnott, EMBL-EBI.

This summary and the full report are available online in printable format, at www.embl.org/documents/document/embl-ebi-impact-report-2021 and www.embl.org/documents/document/embl-ebi-impact-report-summary-2021 respectively.

Executive summary

The findings in this report are the outcome of an independent evaluation undertaken by Charles Beagrie Ltd in 2020/2021.

Introduction

The European Molecular Biology Laboratory (EMBL) is Europe's only intergovernmental laboratory for life science research. Established to advance the study and understanding of molecular biology, nurture young talent, new ideas and technologies, it now performs its activities across six sites in five host nations.

EMBL's European Bioinformatics Institute (EMBL-EBI), located on the Wellcome Genome Campus near Cambridge, is Europe's hub for biomolecular data and an acknowledged world leader in the management and analysis of big data in biology. It is a highly collaborative organisation, with data resources run in partnership with organisations throughout the world.

Investment by EMBL member states and other funders alongside the contributions of collaborators, enable EMBL-EBI to host the world's most comprehensive and integrated collection of data resources in the life sciences.

Charles Beagrie Ltd undertook the first EMBL-EBI Value and Impact study in 2015-2016 and the current study after a gap of five years. Both studies have been conducted as part of an on-going programme, led by EMBL-EBI, to develop a framework and evidence base for demonstrating the economic value and impact of the open data resources.

The user survey for the current study launched in March 2021 and received 4 920 usable responses, providing an excellent foundation for analysis. The response to the user survey is one of the world's largest from surveys on open research and open data in recent years.

The current study used multiple approaches to assess the economic value and impact of 44 open data resources managed by EMBL-EBI. The quantitative economic approaches used in the study include: estimates of access and use value; contingent valuation; estimating the efficiency impacts of EMBL-EBI data resources; and a macro-economic approach that seeks to explore the wider impacts of EMBL-EBI data resources on returns to investment in research. These approaches allow us to develop a picture, beginning with estimates of minimum direct values for the EMBL-EBI's user community and moving progressively toward approaches that measure wider social and economic value. Methods rely on triangulated conservative estimates of EMBL-EBI user populations, levels of use, and investment value (expenditure) required by EMBL-EBI and its many collaborators in developing and delivering the data resources. Both these methods and their application, and so the estimates used by us in this study, are conservative.

In order to isolate attributable impacts, we also use a counter-factual approach, focussing on what users could and would have done if the EMBL-EBI data resources did not exist. For both research efficiency and returns to R&D, this enables us to distinguish between the impacts arising from activities *facilitated* by EMBL-EBI data resources and the impacts that *depend* on EMBL-EBI data resources.

Economic impact studies often focus on a single or very limited approach(es) and metrics, and counter-factuals are rarely used. Although this is simpler to present and easier to undertake, all individual approaches have weaknesses. A strength of this study is the breadth and depth of the multiple approaches employed and the confidence this can give in the overall picture of its findings.

COLLABORATIVE BY DESIGN

EMBL-EBI develops many of its data resources through collaborations with other organisations around the world. This is done in many ways: through joint grants, consortium agreements (e.g. Protein Data Bank in Europe), direct partnerships (e.g. Reactome), and federated models (e.g. European Genome-phenome Archive). Some data resources are jointly managed (e.g. UniProt).

The scale and depth of these collaborations vary, creating a complex ecosystem.

EMBL-EBI does not just store the data, it brings value by curating, standardising, and making the data FAIR (Findable, Accessible, Interoperable and Reproducible). For the sake of brevity, this report refers to “EMBL-EBI managed data resources”, regardless of the type of collaboration agreement.

EMBL-EBI are grateful to its partners and collaborators for their continued support in developing and maintaining robust data infrastructure for the scientific community.

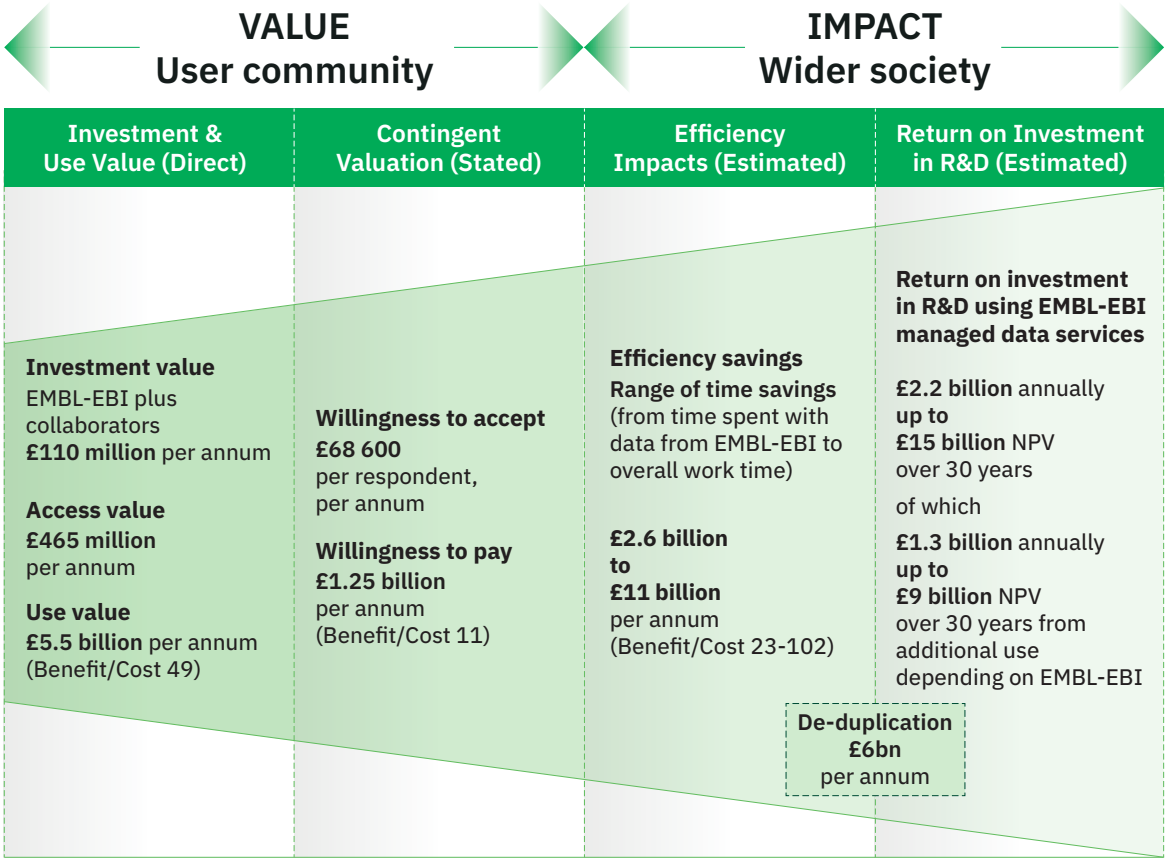
Key findings

The qualitative and quantitative analyses reveal that EMBL-EBI managed data resources are utilised widely and valued highly by their users.

We find that EMBL-EBI managed data resources present exceptional value for money in terms of the value returned and impact compared to the costs of running them.

FIGURE S1

The value and impact of EMBL-EBI managed data resources



Note: NPV is Net Present Value. All estimates rounded.

Source: Authors' analysis

The quantitative analysis summarised in Figure S1 (above) explores the value and impact of EMBL-EBI managed data resources and shows:

- **Use value:** The most direct measure of the value is the time users spend using EMBL-EBI managed data resources - **an estimated £5.5 billion per annum** (more than £3 billion higher than in our 2015-16 impact study, reflecting increased use). **This compares very favourably with the estimated £110 million total annual expenditure with the use value being 49 times the estimated total costs.**[1]
- **Contingent valuation:** Measures the value users place on a freely provided service by asking what they would be willing to pay for that service in a hypothetical market situation, which for EMBL-EBI managed data resources is **an estimated £1.25 billion per annum** (some £925 million higher than 2015-16). This contingent valuation estimate gives a sense of the minimum value of EMBL-EBI's managed data resources to users, **equivalent to 11 times the estimated total costs**.
- **Efficiency impacts:** Researchers reported that EMBL-EBI managed data resources made their research significantly more efficient. This benefit to users and their funders is estimated, at a minimum, to be worth £2.6 billion per annum worldwide, and at a possible maximum worth £11 billion per annum worldwide (almost 120 per cent higher than 2015-16). The estimated efficiency impacts give a sense of the **possible wider value of EMBL-EBI's managed data resources, equivalent to 23 to 102 times the estimated total costs**.
- **Return on Investment in R&D facilitated by EMBL-EBI managed data resources:** During the last year the use of EMBL-EBI managed data resources contributed to the wider realisation of research impacts conservatively estimated to be **worth some £2.2 billion annually** (almost 140 per cent higher than 2015-16), or up to £15 billion over 30 years in net present value.
- **Return on Investment in R&D depending on EMBL-EBI managed data resources:** Some 58 per cent of survey respondents stated that they could neither have created/collected the last data they used themselves nor obtained it elsewhere. As a result, it is estimated that during the last year EMBL-EBI managed data resources underpinned **research impacts worth £1.3 billion annually** (£930 million higher than 2015-16), or up to £9 billion over 30 years in net present value, **that could not otherwise have been realised**.
- **De-duplication of effort:** A key value of open data lies in the de-duplication of research effort. A separate new calculation combining approaches in the study (therefore shown in Figure S1 as an insert) provides a supplementary way of estimating a major part of the value of research efficiencies. ***If*** the time saved by users from not having to (re)create the data enabled more research to be done

[1] Note that during the last year EMBL-EBI managed data resources underpinned more than 140 million hours of research.

(i.e., adding the value of the time saved (£4.3 bn) to the potential returns from the additional research facilitated (£1.5 bn)), it could be **worth almost £6 billion per annum** (more than 200 per cent higher than we would calculate for 2015-16).

The qualitative analysis reveals a similar picture of the value and impact of EMBL-EBI managed data resources:

- **More than two-thirds of all respondents (69 per cent) said that not having access to EMBL-EBI managed data resources would have a "major" or "severe" impact on their work or study**, and 90 per cent in total said not having EMBL-EBI would have a major, severe or moderate impact on their work. This represents a substantial increase compared to our 2015-16 study when the equivalent responses were 55 per cent and 84 per cent.
- The study shows that **EMBL-EBI's managed data resources have been making a major contribution to science throughout the COVID-19 Pandemic**. Some 32 per cent of respondents said they valued EMBL-EBI data resources more as a result of the pandemic and 25 per cent said they had used them more.

Value and impact statement have been based on estimated values of EMBL-EBI user populations and levels of use. This has focused on an estimated 450 000 to 500 000 unique primary users, those who access the data resources managed by EMBL-EBI directly and who could be invited to respond to our user survey, possibly 20 per cent of worldwide life science researchers. The study did find extensive indirect secondary use and a considerable number of secondary users of open data managed by EMBL-EBI, which are not included in our impact assessment.

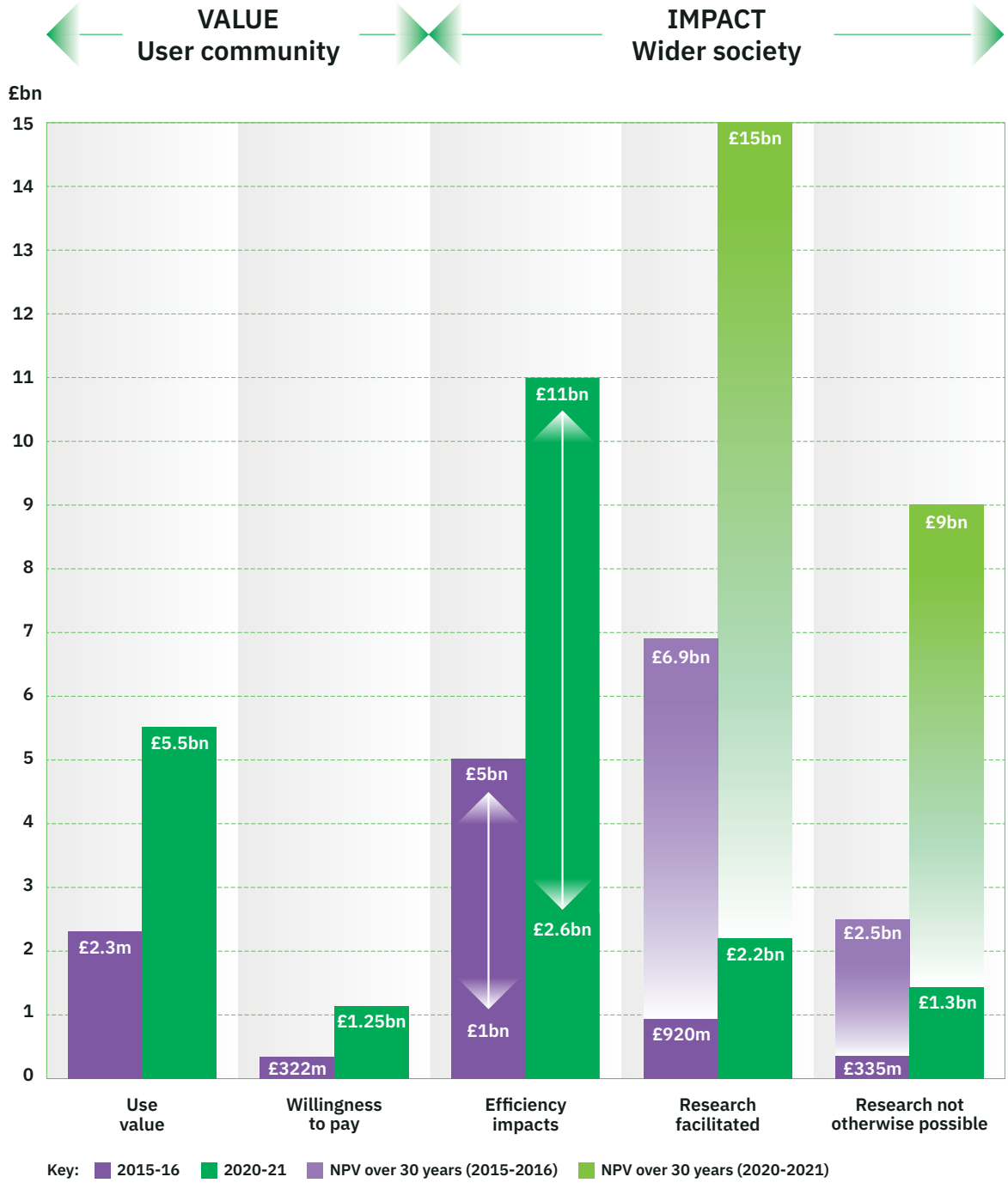
Putting these findings into context

The current study and the original value and impact study conducted in 2015-16 are snapshots in time and there are limits to how they can be compared. Nevertheless, the economic estimates can provide a sense of the scale of economic impacts at the time and changes over the last five years.

The current study shows higher levels of value and impact across the board compared to 2015-16, reflecting the increased level and intensity of use of EMBL-EBI managed data resources, the increasing maturity of the data resources, and evolving research practices. Figure S2 (below) shows the 2015-16 results alongside those from 2020-21.

FIGURE S2

The increasing value and impacts of EMBL-EBI managed data resources (2015-16 & 2020-21)



Note: All values are per annum apart from the Net Present Values (NPV) for returns on research which are over 30 years.

Source: Authors' analysis

It is difficult to make direct comparisons with other studies. To be truly comparable studies must use the same methods and processes of application (e.g., in activity costing, estimation of user population, etc.), they must focus on similar types of resources and be done in a similar timeframe and locale. However, the current study compares favourably against a range of studies of library and information services, national statistics and other information and data services (See Box S1 below).

BOX S1

Putting the value and impact of EMBL-EBI managed data resources into context

While individual studies focus on different information services and content, and use a variety of methods and measures, it is possible to explore their findings to give a sense of how the value and impact of EMBL-EBI data resources compare:

- Houghton (2011) estimated the benefit/cost ratio of the Australian Bureau of Statistics making data and publications freely available online and using Creative Commons licensing at 5.3 to 1.
- Tennison (2015) reported that a report by Nesta and the ODI adds to the evidence of the impact of open data. The report's analysis, undertaken by PwC, examines the effects of the Open Data Challenge Series (ODCS) and predicts the programme will result in a potential 10x return (£10 for every £1 invested over three years), generating up to £10.8m for the UK economy.
- Like other meta reviews, *Measuring the Value of California's Public Libraries, 2017-2020* found that investment in public libraries is a sound use of public funds: for every dollar invested in libraries, about $2-$10 are returned, with an average of between $3 and $6.
- King (2010) summarized findings relating to library services and concluded that: special libraries exhibit a return of 2.9 to 1, academic libraries 3.4 to 1 (for staff), and public libraries 5.8 to 1.

Source: Authors' analysis

European Bioinformatics Institute (EMBL-EBI)
Wellcome Genome Campus
Hinxton, Cambridge, CB10 1SD
United Kingdom

www.ebi.ac.uk
+44 (0)1223 494 444
comms@ebi.ac.uk

@emblebi
/EMBLEBI
/EBImedia
/company/ebi/

EMBL-EBI is a part of the European Molecular Biology Laboratory.

EMBL member states and associate member states: Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom, Argentina, Australia

Prospect member states: Estonia, Latvia