

The Value and Impact of the European Bioinformatics Institute

Executive Summary
January 2016

Analysis and report by:

Neil Beagrie, Charles Beagrie Ltd.,
Salisbury, United Kingdom
John Houghton, Victoria University,
Melbourne, Australia

About the Authors

Neil Beagrie is director of consultancy at Charles Beagrie Ltd, an independent consultancy company specialising in the digital archive, library, science and research sectors. He has extensive experience in cost-benefit analysis, research data management and digital preservation. Neil has been the lead consultant or principal investigator for a wide range of research studies including the Keeping Research Data Safe (KRDS) projects, which investigated the costs and benefits of digital research data curated by UK universities. He is a Fellow of the RSA and was awarded the 2014 SMPTE Archival Technology Medal for his long-term contributions to digital preservation.

John Houghton is a Professorial Fellow at Victoria University's Centre for Strategic Economic Studies (CSES) and an Associate at Charles Beagrie Ltd. He is an economist specialising in applying economic assessment techniques to information technology policy, science and technology policy and in exploring the value and economic implications of open access scholarly publishing and open data. A number of John's studies have also focussed on the economics of research data services. In 1998, John was awarded a National Australia Day Council, Australia Day Medal for his contribution to IT industry policy development.

Neil and John have co-authored a series of four major studies on the value and impact of large research data services in the UK.

About this publication

© 2016 EMBL-EBI. This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>

The information in this publication may be freely reprinted and distributed for non-commercial use via print, broadcast and electronic media, provided that proper attribution to authors and designers is made.

The artwork in this executive summary was taken from the EMBL-EBI Annual Scientific Report 2014. Cover image credit: Spencer Phillips, EMBL-EBI.

This summary and the full report are available online in printable format, at www.beagrie.com/EBI-impact-summary.pdf and www.beagrie.com/EBI-impact-report.pdf

Introduction to EMBL-EBI

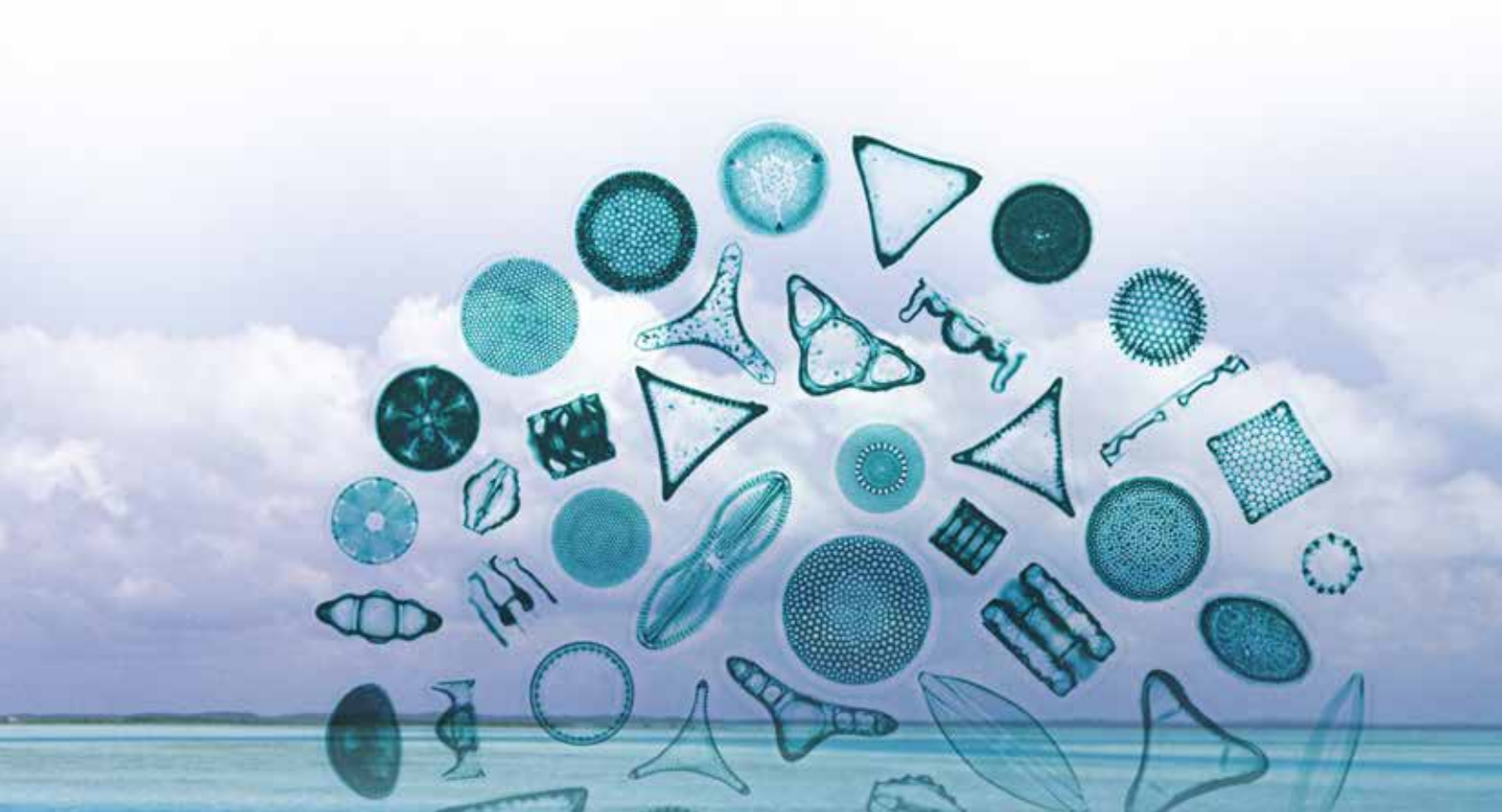
At EMBL-EBI, we manage public life-science data on a very large scale, offering a rich resource of carefully curated information. We make our data, tools and infrastructure openly available to an increasingly data-driven scientific community, adjusting to the changing needs of our users, researchers, trainees and industry partners. This proactive approach allows us to deliver relevant, up-to-date data and tools to the scientists who depend on our services.

We are a central partner in global efforts to exchange information, set standards, develop new methods and curate complex information. Our core databases are produced in collaboration with other world leaders including the National Center for Biotechnology Information in the US, the National Institute of Genetics in Japan, the SIB Swiss Institute of Bioinformatics and the Wellcome Trust Sanger Institute in the UK.

We are also a world leader in computational biology research, and are well integrated with experimental and computational groups on all EMBL sites. Our research programme is highly collaborative and interdisciplinary, regularly producing high-impact works on sequence and structural alignment, genome analysis, basic biological breakthroughs, algorithms and methods of widespread importance.

EMBL-EBI received substantial funding as part of the UK Government's Strategy for Life Sciences via its Large Facilities Capital Fund (LFCF), to support both the computational and building infrastructure for sustaining biological big-data capacity. The building houses the Technical Hub for ELIXIR, which will in future coordinate the delivery of biological data infrastructures throughout Europe.

The UK's Biotechnology and Biological Sciences Research Council (BBSRC), through the LFCF programme, has actively encouraged this analysis and on-going studies of the wider value and impact of EMBL-EBI as an infrastructure supporting the global life-science community.



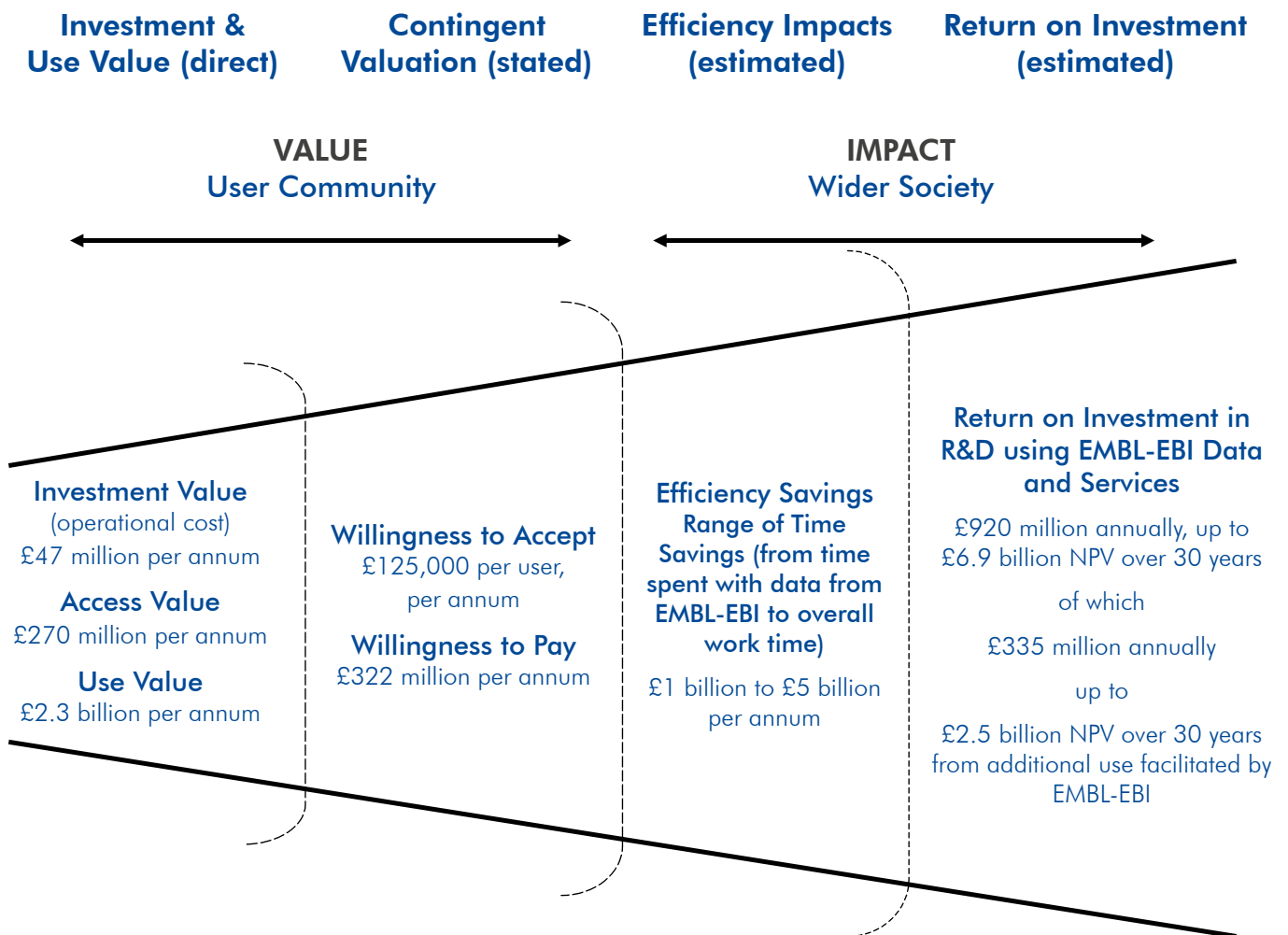
Executive Summary

The European Bioinformatics Institute (EMBL- EBI), located on the Wellcome Genome Campus in Hinxton, UK, manages public life-science data on a very large scale, making a rich resource of information freely available to the global life science community. EMBL-EBI is one of a handful of organisations in the world involved in global efforts to exchange information, set standards, develop new methods, and curate complex genome information.

We present here the results of a quantitative and qualitative study of the Institute, examining the value and impact of its work. Our focus is the economic impact and can be seen as complementary to traditional academic measures, such as citation counts.

This study was conducted as part of an ongoing programme, led by EMBL-EBI, to develop a framework and evidence base for demonstrating how the Institute captures and curates the increasingly vast output of genome research and allows it to be easily located, understood, and applied.

The value and impact of EMBL-EBI data and services



This summary and the full report are available online:

<http://www.beagrie.com/EBI-impact-report.pdf> and <http://www.beagrie.com/EBI-impact-summary.pdf>

Key Findings

The qualitative and quantitative analyses reveal that EMBL-EBI services are utilised widely and valued highly by their user community. The quantitative analysis, summarised in the figure, explores the value and impact of EMBL-EBI data and services, and shows:

- **Access (use) value:** The most direct measure of the value is the time, and therefore costs, spent by users accessing EMBL-EBI data and services – an estimated £270 million during the year to May 2015.
- **Contingent valuation:** is an alternative approach to estimating what something is worth, measuring the value users place on a freely provided service, and is an estimated £322 million during the year to May 2015.
- These estimates give a sense of the minimum direct value of EMBL-EBI's data and services to its user community, and compare very favourably with the approximately £47 million annual operational expenditure, with a minimum direct value to users that is equivalent to around 6 times the direct operational cost.
- **Efficiency impacts:** Users reported that EMBL-EBI data and services made their research significantly more efficient. This benefit to users and their funders is estimated, at a minimum, to be worth £1 billion per annum worldwide - equivalent to more than 20 times the direct operational cost.
- **Return on Investment in R&D:** during the last year the use of EMBL-EBI services contributed to the wider realisation of future research impacts conservatively estimated to be worth some £920 million annually, or £6.9 billion over 30 years in net present value.
- A large number 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 data and services underpinned future research impacts worth £335 million annually, or £2.5 billion over 30 years in net present value, that could not otherwise have been realised.

**Efficiencies gained:
Worth ~£1 billion per
annum worldwide,
or 20 times the direct
operational cost**

**Underpinning future
research impacts:
~£335 million
annually that could
not otherwise have
been realised.**

**EMBL-EBI data: 45% of survey
respondants stated that they could
neither have created/collected the
last data they used themselves, nor
obtained it elsewhere.**

This summary and the full report are available online:

<http://www.beagrie.com/EBI-impact-report.pdf> and <http://www.beagrie.com/EBI-impact-summary.pdf>

About the impact survey

The qualitative analysis reveals a similar picture of the value and impact of EMBL-EBI data and services.

The user survey, launched in May 2015, received 4509 responses, providing an excellent foundation for analysis.

The survey found that users spend considerable time on research. The mean number of hours spent in research was 34 hours per week. More than half of this time (mean of 56 per cent) was spent working with data. One-fifth (20 per cent) of this was with data from EMBL-EBI. These numbers reflect the vital importance of molecular data (and related services) and the intensity of its use among life-science researchers.

More than half of all respondents (55 per cent) said that not having access to EMBL-EBI services and resources would have a “major” or “severe” impact on their work or study.

The quantitative assessment and the qualitative one independently revealed similar findings about the economic value and wider impact of EMBL-EBI data and services. This gives added credence to the findings.

A selection of the comments received as part of the survey are summarised as an appendix in the full report, which is available online in printable format at <http://www.beagrie.com/EBI-impact-report.pdf>

“EMBL-EBI resources are essential to all modern life sciences research. It would be impossible to work with data of any kind relating to genomes, genes, proteins, small molecules, 3D structures or other related data without reference to EMBL-EBI resources. We simply could not function without the core, reliably maintained data collections and the world-leading expertise that is centred on EBI for the organisation and analysis of biological data.”

“Services from EMBL-EBI are a well-established and trusted resource for bioinformatic and proteomic research, as well as for structural biology and drug discovery/design. Even more, it gets better every year, making life easier for users, and ultimately speeding up the research process”.

“I use EMBL-EBI services on a daily basis. Most of these cannot be sourced elsewhere on the internet and my institution lacks the resources to provide them locally. EMBL-EBI training sessions have been by far the best NGS and RNASeq courses I have attended.”

This summary and the full report are available online:

<http://www.beagrie.com/EBI-impact-report.pdf> and <http://www.beagrie.com/EBI-impact-summary.pdf>

Methodology

Quantifying the value and impact of EMBL-EBI is complicated due to many factors: its open and free provision of services, the collaborative nature of the work researchers undertake, the wide range of services it provides, and the diversity of communities it serves. Our approaches took account of the practical limitations inherent in collecting data through interview and survey techniques.

The quantitative economic approaches used included: estimates of access and use value, contingent valuation using stated preference techniques, an activity-costing approach to estimating the efficiency impacts of EMBL-EBI data and services, and a macro-economic approach that seeks to explore the impacts of EMBL-EBI use on returns to investment in research.

These approaches allowed 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.

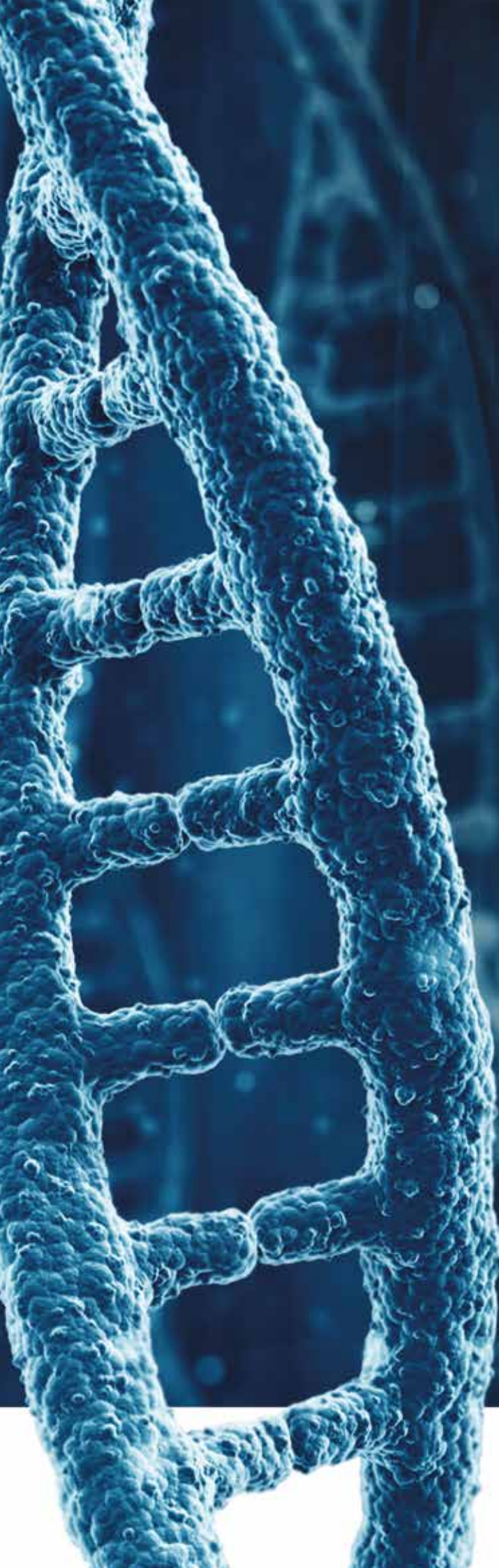


Knowledge of the EMBL-EBI user population numbers and levels of use underpin estimates of the value and impact of EMBL-EBI as a whole. EMBL-EBI data and services are almost entirely open resources. Users are not required to register and are not directly identified or recorded.

The various proxy measures of users and levels of use, such as Unique IP addresses or Web downloads, have significant limitations, which are amplified by the large number of individual EMBL-EBI services surveyed. In the absence of user registration, we estimated EMBL-EBI's user population by combining information from log data, a user survey, and external studies. The result was an estimate of 198 000 direct active users during the year to May 2015, accessing data 88 million times. This figure is felt to be conservative as it excludes what is known to be extensive secondary use. There are a number of observations in the conclusions which may help to validate and improve future knowledge of the user population and related research data metrics.

This summary and the full report are available online:

<http://www.beagrie.com/EBI-impact-report.pdf> and <http://www.beagrie.com/EBI-impact-summary.pdf>



The LFCF Programme was facilitated and managed by the Biotechnology and Biological Sciences Research Council (BBSRC) on behalf of Research Councils UK (RCUK).