technopolis |group|

Measuring the value and impact of the Europe PMC repository

Final Report

technopolis |group|

Measuring the value and impact of the Europe PMC repository

Final Report

technopolis |group| April 2019

Anoushka Davé
Vivek Seth
Emma Pottinger
Cristina Rosemberg
Xavier Potau
Peter Varnai

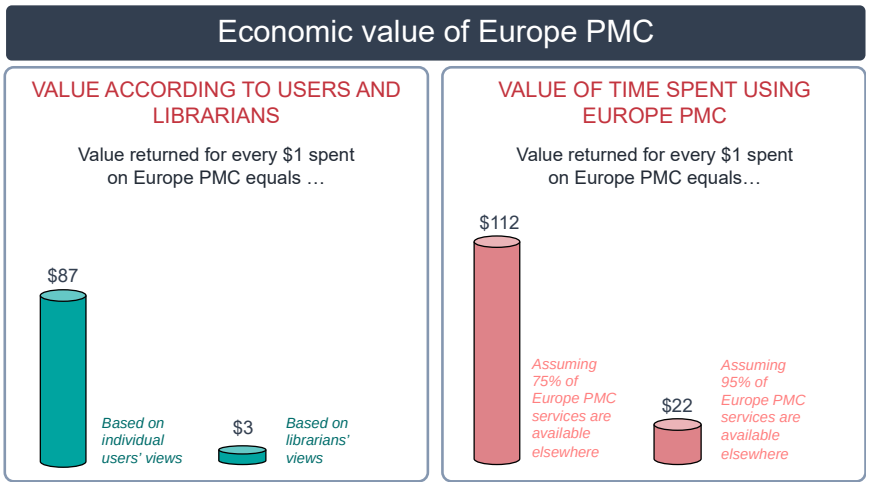
Executive Summary

Europe PMC is an open access non-commercial online repository, providing free access to life sciences and biomedical information including articles, books, patents and clinical guidelines. In addition, it offers a range of tools for users, such as ORCID linking, programmatic and text-mining access via publicly available APIs and a grant finder tool to search grants awarded by Europe PMC funders. The repository is operated by the European Bioinformatics Institute (EBI) based at the European Molecular Biology Laboratory (EMBL) in Cambridge and is funded by a grant, which is managed by the Wellcome Trust.

Europe PMC includes the full body of content within PubMed Central (PMC) and allows access to both PubMed and PMC content via a single search interface. Europe PMC has a global reach in terms of its funders, content, contributors and user community.

Key findings

The economic value of Europe PMC



- Our surveys show that users put a mean value of about $28 per user per year on Europe PMC while librarians value the service at around $1 per user per year. On multiplying these mean values with the mean number of users (unique IP addresses as proxy) in a year (about 10 million), we obtain a contingent valuation per year of $286 million and $11.3 million, respectively for Europe PMC.

  Based on Europe PMC funders' contributions and estimated in-kind contributions from EMBL-EBI, Europe PMC costs about $3.3 million to run per year. This is a conservative estimate as the actual value of in-kind contributions made by EMBL-EBI and others is not known. Moreover, Europe PMC benefits from obtaining much of its publication content from PMC for no cost.

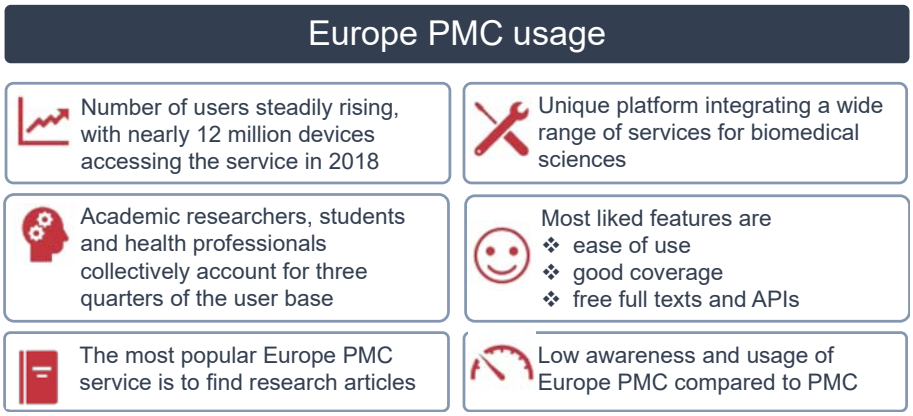
  On comparing the contingent valuation figures to the known costs of running Europe PMC per year, we find that *for every dollar contributed to Europe PMC, the value returned is $87 or $3 based on the views of users and librarians, respectively*.

- Additionally, we calculated the monetary value of the time spent by users on Europe PMC (usage valuation) which is akin to assessing the related opportunity cost. Here we assume that time spent on the platform represents value to users in terms of accessing and using knowledge rather than any potential inefficiency of the system. We found that this equated to $1.5 billion for all users in 2018. However, since many of Europe PMC's services can be replicated using other alternatives, the usage value needs adjustment based on the extent to which Europe PMC is unique. Crucially, there is no direct comparator for Europe PMC as a whole and the extent of duplication between services depends on the specific type of use and availability of alternatives to users. Moreover, although certainly positive, it is difficult to put a quantitative value on the

benefit of having several services in one place. Nevertheless, we estimated that the extent of duplication ranges from high (75%) to very high (95%). Applying this adjustment to the usage valuation, we found that *for every dollar spent on Europe PMC, the usage value returned is between $112 and $22 respectively*.

- Both these valuations show that Europe PMC presents a very high value to users relative to the moderate investment made by its funders to run the services. It is however important to stress that the contingent valuation (based on willingness to pay) and usage valuation (based on self-reported usage times) of Europe PMC as a knowledge platform cannot be understood as a genuine market price. The approaches used may suffer from cultural-strategic biases in stated preferences and thus may lead to overestimation of the value Europe PMC provides.
- To compare, a valuation of EMBL-EBI data and services using the same two methods demonstrated a value to users equivalent to around 6 times the direct operational cost[1]. That is, for every dollar spent, the value returned was $6. While the contingent valuation of EMBL-EBI data and services in that study was about twice that of Europe PMC (as calculated based on users' willingness to pay data), the annual operational cost of EMBL-EBI was 24-fold higher than that of Europe PMC.[2] These numbers indicate that the high multiples of Europe PMC valuation are the result of a combination of the high value attributed to Europe PMC services by the global user community with the relatively low cost of running services that are embedded in an established research infrastructure.

Europe PMC usage



- The *number of Europe PMC users has been steadily rising over the past five years*. In addition, the service has a truly global reach, with the US, China, India and the UK collectively accounting for almost half of the unique IPs. This finding also indicates that the research funded by Europe PMC funders is being accessed, read and potentially built upon by the wider global research community, in the true spirit of 'open science'.
- Survey findings show that the majority of Europe PMC users are academic researchers, students and health professionals. Users are most likely to use Europe PMC to find research articles, but they also use the platform to upload manuscripts, check open access compliance, link to citations and ORCIDs (Open Researcher and Contributor IDs). Third-party developers and industry users are the main users of the application programming interfaces (APIs), relying on them for text and data mining.
- One of the main comments regarding Europe PMC is that the service is not well publicised, and hence many potential users are not yet aware of Europe PMC and the value it could add.

[1] Beagrie, N. and Houghton, J. (2016) *The value and impact of the European Bioinformatics Institute*

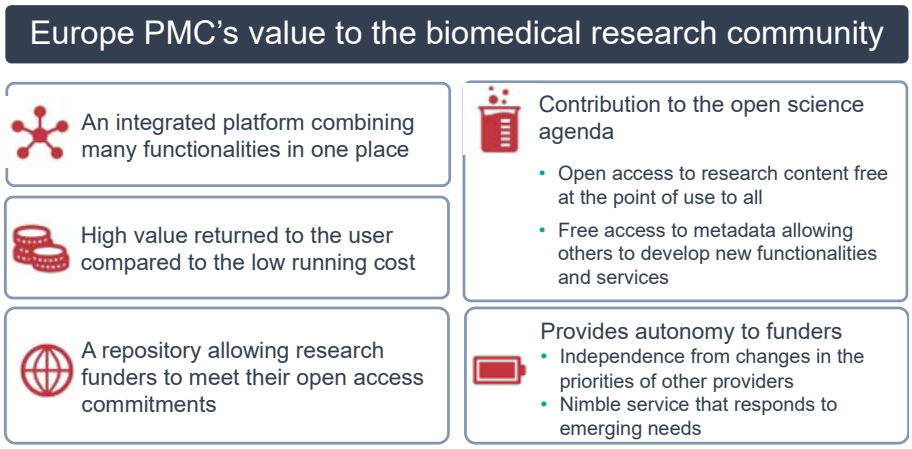
[2] Contingent valuation based on users' data gave a value of $492 million for EMBL-EBI in 2015, while the annual operational cost was $72 million, using OECD's PPP exchange rate of 0.655 in 2015.

- *While no other platforms provide exactly the same range of services as Europe PMC, collectively they cover most of Europe PMC's services and functionalities.* For many users, Europe PMC is one of several different tools to find biomedical content. Pubmed, PMC, search engines and Google Scholar are the most commonly used comparator services. However, Europe PMC is seen as providing better open access provision, a more user-friendly interface, and better functionalities such as faster search speeds, broader range of services and easier access to DOIs than other similar services. Nevertheless, usage of Europe PMC remains especially low in comparison to PMC. For example, usage statistics for articles published in Wellcome Open Research, Proceedings of the National Academy of Sciences (PNAS) and PLOS Biology show that Europe PMC use is one-tenth or lower that of PMC.

What would happen if Europe PMC no longer existed?

- If Europe PMC ceased to exist, it would cause noticeable problems to several user groups. Poorly resourced researchers (e.g. those from developing countries, citizen scientists, and students) would have reduced access to research publications. Some third-party developers would have to completely rebuild their applications, entailing additional time and financial costs.
- It was also noted that Europe PMC has made a vital contribution to the Open Access movement and could play an important role in the context of Plan S[3]. Thus, loss of Europe PMC would be a loss to the Open Science agenda. In addition, funders who currently use Europe PMC as their open access repository would have to make alternative arrangements, incurring additional costs.
- Finally, Europe PMC offers a degree of autonomy compared to PMC, which is subject to changes in the US political climate. This autonomy also allows Europe PMC to be nimbler in responding to emerging needs, thus enabling it to rapidly introduce new functionalities and services such as indexed preprints.

Conclusions



Overall, Europe PMC presents exceptional value for money in terms of the value returned compared to the relatively low costs of running it. While Europe PMC is not completely unique, it offers an integrated, free platform combining several of its competitors' services that potentially saves users' time and money. For instance, in the absence of Europe PMC, its funders would have to make alternative arrangements for an open access repository. Notably, Europe PMC contributes greatly to the Open Science agenda and has a potentially important role in the context of Plan S. Finally, its autonomy offers protection against

[3] Plan S is an initiative for open-access science publishing that was launched by Science Europe on 4 September 2018. It is supported by the European Research Council and major national research agencies and funders from 12 European countries. The plan requires researchers funded by state-funded research organisations and institutions to publish their work in open repositories or open access journals by 2020.

changes in the priorities of commercial providers and the US government, and an ability to be nimble in response to emerging needs in the community.

Table of Contents

Executive Summary iii
Key findings iii
Conclusions v
1 Introduction 5
2 Overview of Europe PMC 6
2.1 History of Europe PMC 6
2.2 Current purpose and remit 7
2.3 Organisation and operation 8
3 Usage of Europe PMC 9
3.1 Number and location of Europe PMC users 9
3.2 Types of users 10
4 Europe PMC services and functionality 11
4.1 Types of usage 12
4.2 What users like about Europe PMC 14
4.3 What users think can be improved 15
5 Comparison of Europe PMC to other services 16
5.1 Comparator services 16
5.2 Pros and cons of Europe PMC compared to other services 18
5.3 Europe PMC versus PMC usage statistics 20
6 Economic Valuation 21
6.1 Cost of running Europe PMC 21
6.2 Economic valuation of Europe PMC 21
6.3 Other considerations of impact 27
7 What would happen if Europe PMC no longer existed? 29
8 Conclusions 31
Appendix A Approach and Methodology 33
Appendix B Email-based researcher survey questionnaire 41
Appendix C Website user survey 44
Appendix D Librarian survey questionnaire 47
Appendix E Survey analysis 50
Appendix F Usage valuation 64

Tables

Table 1 Europe PMCs main services and functionality 11
Table 2 The main comparator services to Europe PMC 16
Table 3 Comparison of Europe PMC and PMC usage statistics (June 2017 to November 2018) for Wellcome Open Research, Proceedings of the National Academy of Sciences (PNAS) and PLOS Biology 20
Table 4 Summary of financial support requested for Europe PMC (2016 to 2021) 21
Table 5 Cost per user per year for running Europe PMC (2016 to 2021) 21
Table 6 Total net value of Europe PMC usage per year 23
Table 7 Alternatives that replicate Europe PMC services 24
Table 8 Usage valuation of Europe PMC under different deadweight scenarios (based on Scottish Enterprise ready reckoner in Table 14) 24
Table 9 Finalised Methodological Approach 33
Table 10 Geographical distribution of email addresses contacted for user survey 34
Table 11 List of interviewees 35
Table 12 User and stakeholder interview questionnaire 36
Table 13 Calculating the gross effect of Europe PMC usage 38
Table 14 Deadweight ready reckoner 40
Table 15 Segmentation of Europe PMC users by geography 64
Table 16 Breakdown of segments by user type 65
Table 17 Median duration of each Europe PMC visit (minutes) 65
Table 18 Mean number of visits to Europe PMC per month 66
Table 19 Time valuation per user type (median wage per hour worked- PPP $) 67
Table 20 Value of Europe PMC usage for all users in developed economies 67
Table 21 Value of Europe PMC usage for all users in transition economies 68
Table 22 Value of Europe PMC usage for all users in developing economies 69
Table 23 Total economic value of Europe PMC usage (PPP $) 69

Figures

Figure 1 Intervention logic model for Europe PMC 8
Figure 2 Unique IPs accessing Europe PMC per year 9
Figure 3 Geographical Distribution of IPs (2018) 9
Figure 4 Distribution of website survey respondents by user type (n=958) 10
Figure 5 Distribution of respondents by usage type (n=1086) 13
Figure 6 Use of other services by user survey respondents (n=1256, includes Europe PMC users and non-users) 18
Figure 7 Services promoted by libraries or information services (n=137) 18
Figure 8 Summary of usage valuation modelling approach 22

Figure 9 Analysis of willingness-to-pay responses by amount (a; n=856) and common reasons cited for choice (b; n=381) 25

Figure 10 Librarians' responses on their institution's willingness-to-pay amounts (a) and number of life sciences users (b) [n=118] 26

Figure 11 Distribution of respondents by geography (a) and user type (b) 50

Figure 12 Distribution of respondents by geography 51

Figure 13 Distribution of survey respondents by user type (n=398) 52

Figure 14 Distribution of respondents by number of visits to the Europe PMC website per month (a) and time spent per visit (b) (merged data from both the website user survey and the email-based researcher survey) 53

Figure 15 Distribution of respondents by type of use (a) and helpfulness (b) of Europe PMC services (merged data from both the website user survey and the email-based researcher survey. 53

Figure 16 Use of other services (merged data from both the website user survey and the email-based researcher survey) [n=1256] 54

Figure 17 Responses to the librarians' survey by country (a) and organisation type (b) 54

Figure 18 Number of life science users that the respondents are providing services to 56

Figure 19 Services promoted by respondents' library or information services portal [n=140] 56

Figure 20 Distribution of respondents by helpfulness (n=43) 57

Figure 21 Distribution of UK respondents by user type (n=83, across both user surveys) 57

Figure 22 Distribution of UK respondents by number of visits to the Europe PMC website per month (a) and time spent per visit (b) (across both user surveys) 58

Figure 23 Distribution of UK respondents by type of use (a) and helpfulness (b) of Europe PMC services (across both user surveys) 58

Figure 24 Use of other services by UK respondents (n=83, across both user surveys) 59

Figure 25 Distribution of willingness to pay amounts for UK respondents (n=54, across both user surveys) 60

Figure 26 UK responses to the librarians' survey by organisation type [n=35] 60

Figure 27 Number of life science users (a) and willingness to pay amounts (b) for UK respondents to the librarian survey 60

Figure 28 Services promoted by UK respondents' library or information services portal [n=35] 61

Figure 29 Distribution of UK respondents by helpfulness (n=14) 62

Figure 30 Geographical distribution of Europe PMC survey respondents (left panel) and manner of use (right panel) 63

Figure 31 Use of comparator services by Europe PMC users 63

1 Introduction

This document is the final report of the project titled 'Measuring the value and impact of the Europe PMC repository'.

The Wellcome Trust commissioned Technopolis on behalf of the funders of Europe PMC to conduct a review of the value and impact of the Europe PMC repository service – both to the consortium of funders that support the service and the global scientific community. The broader aim of this commission is to inform funders in advance of an anticipated new funding request to support the ongoing development and maintenance of the repository from 2021.

We used a mixed methods approach including document and data review, surveys and a programme of interviews to develop an economic valuation of Europe PMC and impact case studies to demonstrate the value and impact of Europe PMC (a detailed methodology is available in Appendix A).

The document is further structured into the following sections

- Overview of Europe PMC
- Usage of Europe PMC
- Europe PMC services and functionality
- Comparison of Europe PMC to other services
- Economic Valuation
- What would happen if Europe PMC no longer existed?
- Conclusions

The survey questionnaires, survey analysis and usage valuation calculations are presented as annexes to the main report.

2 Overview of Europe PMC

2.1 History of Europe PMC

In 2006, the Wellcome Trust (hereafter described as Wellcome) conducted an exercise examining articles where they were attributed as funder, finding that only 6% of these articles were freely available on the internet. As a result, Wellcome adjusted its grant conditions to improve open access, ensuring that research papers partly or wholly funded by Wellcome were made freely accessible no later than 6 months after publication[4]. This new policy led to the genesis of UK PubMed Central (UKPMC). UKPMC was devised by Wellcome and developed in partnership with the British Library, the University of Manchester and the European Bioinformatics Institute. It mirrored PubMed Central (PMC), a free archive and repository for biomedical and life sciences journal articles funded by the US National Institutes of Health (NIH). UKPMC hosted the same content, search and browse mechanisms as PMC. Later, other major funders of UK biomedical research including but not limited to the Medical Research Council (MRC), British Heart Foundation (BHF), and Cancer Research UK came onboard as funders of UKPMC. Since all UKPMC funders had open access policies, they introduced a new requirement that publications funded by UKPMC funders had to be made freely and publicly available on UKPMC[5].

In January 2010, UKPMC introduced a new interface to improve navigation and content searches, including a function to search abstracts and full-text from the same search box. Further functions were added in the years that followed, including integrated text-mining, manuscript submissions, and grant reporting tools.[6]

Following this, in November 2012, UKPMC was re-named Europe PMC, reflecting that it had now gained funders from across Europe. At present, Europe PMC has 29 research funders (listed below), including some from outside Europe.

- The Academy of Medical Sciences
- Action on Hearing Loss
- Alzheimer's Society
- Arthritis Research UK
- Austrian Science Fund (FWF)
- Biotechnology and Biological Sciences Research Council (BBSRC)
- Bloodwise
- Breast Cancer Now
- British Heart Foundation
- Cancer Research UK
- Chief Scientist Office
- Diabetes UK
- The Dunhill Medical Trust
- European Research Council (ERC)
- Marie Curie
- Medical Research Council (MRC)
- Motor Neurone Disease Association
- Multiple Sclerosis Society
- Myrovlytis Trust
- National Centre for the Replacement, Refinement and Reduction of Animals in Research
- Department for Health and Social Care
- Parkinson's UK
- Prostate Cancer UK
- Swiss National Science Foundation
- Telethon
- Wellcome Trust
- Wellcome Trust/DBT India Alliance
- World Health Organization
- Worldwide Cancer Research

[4] Walport M & Kiley R. Open access, UK PubMed Central and the Wellcome Trust. *J R Soc Med.* 2006; 99: 438-439

[5] McEntyre JR, Ananiadou S, Andrews S, et al. UKPMC: a full text article resource for the life sciences. *Nucleic Acids Research.* 2011; 39(Database issue):D58-D65. doi:10.1093/nar/gkq1063.

[6] Ibid.

2.2 Current purpose and remit

Europe PMC's current mission is "to build open, full text scientific literature resources and support innovation by engaging users, enabling contributors, and integrating related research data"[7]. It continues to share PMC content, but also contains additional material which Europe PMC funders have themselves supported. As such, Europe PMC consists of two main components: Europe PMC (the main website) and Europe PMC+ (for manuscript submission and grant data management). The platform has a wide remit covering a range of different facets:

- Acting as a free and non-commercial database of mainly life sciences and biomedical research information. It provides easy and quick access to a wide variety of sources including articles, books, patents and NHS clinical guidelines. The full body of content within PubMed and PMC is available from Europe PMC via a single search interface. Europe PMC also provides links to records in databases such as Uniprot, the European Nucleotide Archive (ENA), Protein Data Bank in Europe (PDBe) and BioStudies[8]
- Allowing bulk download via FTP (File Transfer Protocol) of full text open access content and other datasets
- Providing the infrastructure to help users derive maximum benefit from literature held on Europe PMC. It does this by supporting text and data mining services as well as ORCID (Open Research and Contributor ID) linking and citation tracking
- Providing a data resource for Application Programming Interfaces (APIs)[9], which help to disseminate information to potential wider users and stakeholders
- Providing research funders (mainly European) with a repository and infrastructure that is independent of PMC and the associated US federal funding it receives

Although Europe PMC focuses on the life sciences, it also contains research in the humanities and social sciences related to life sciences (e.g. Europe PMC funder supported papers on the history of medicine) and NASA research through its sharing of PMC content. Europe PMC also sees itself as having a global reach with its content, authorship and users, all being spread worldwide.

The logical sequence and causal relationships among the objectives, resources (inputs), results (outputs) and expected impacts of Europe PMC are laid out in the intervention logic model below.

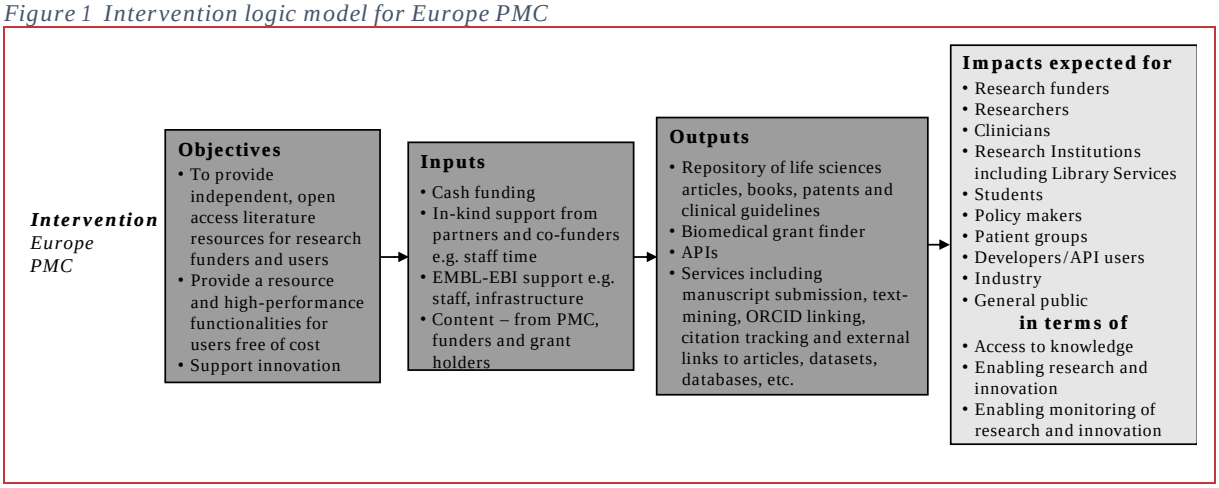
The model lays out Europe PMC's objectives as envisaged by the funders and the inputs they and EMBL-EBI (described in further detail below) provide to achieve these objectives. Next, it shows the outputs emerging from the inputs and the impacts that these are expected to lead to.

[7] Europe PMC Roadmap. Available at: https://europepmc.org/docs/EuropePMCRoadmapJul-Oct2018.pdf

[8] Sarkans U et al. (2018) The BioStudies database-one stop shop for all data supporting a life sciences study. *Nucleic Acids Res* (D1) 4;46

[9] An API is a set of functions and procedures allowing the creation of applications that access the features or data of an operating system, application, or other service.

Figure 1 Intervention logic model for Europe PMC



2.3 Organisation and operation

Each funder provides a financial contribution to Europe PMC, proportional to their total research spending. All partners are also asked to ensure that any open access research they are funding is deposited in Europe PMC.

The European Bioinformatics Institute (EBI) based at the European Molecular Biology Laboratory (EMBL) in Cambridge operates Europe PMC and is funded by a grant, which is managed by the Wellcome Trust. The grant goes mainly towards funding staff, however EMBL-EBI also provides some additional staff time and relevant IT infrastructure in-kind[10].

Key financial decisions surrounding Europe PMC are made by the Funders Committee, which is a subset of all Europe PMC funders. Wellcome is a member of the Funders Committee as it runs the Europe PMC grant while MRC is represented as a core funder providing more than 20% of the overall funding. Other members change over time (every two years) but as a minimum will include one other large funder, one smaller funder, and a funder based outside the UK, with a view to ensuring that Europe PMC caters for the full range of interests and requirements amongst its funders. The Committee meets twice a year, with other funders able to attend as observers. Aside from the Funders Committee, there is also an Annual General Meeting where all funders, the project management team, and the Europe PMC Principal Investigator (from EMBL-EBI) meet.

[10] EMBL-EBI grant application to The Wellcome Trust (2015)

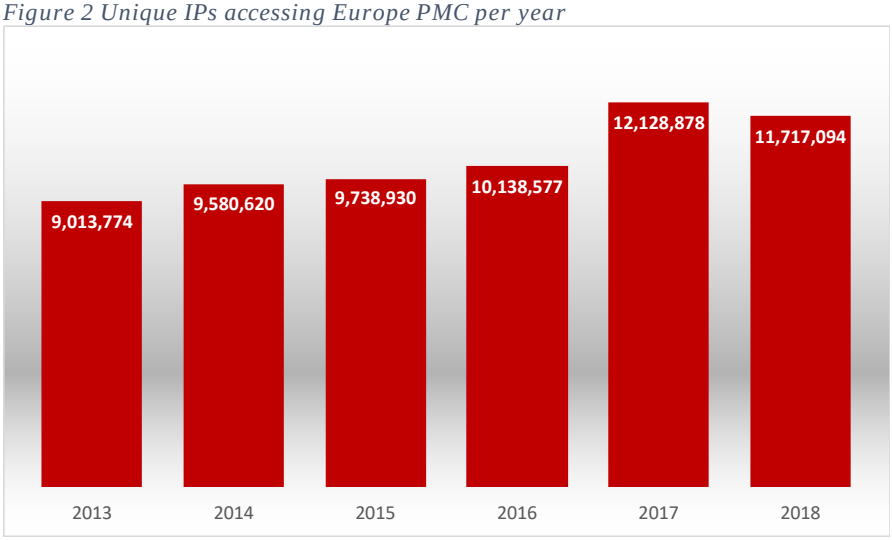
3 Usage of Europe PMC

Users of Europe PMC are largely anonymous and therefore usage has to be estimated via website statistics and user surveys.

3.1 Number and location of Europe PMC users

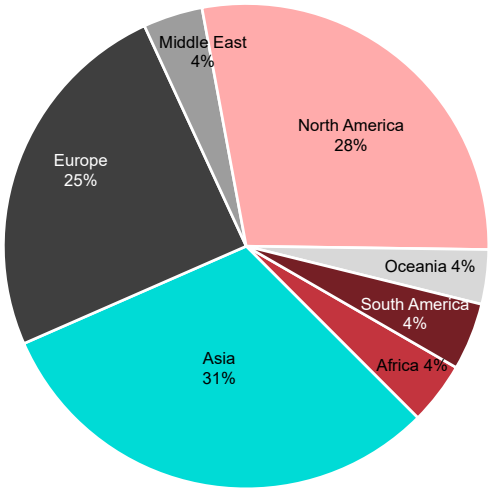
2011 to 2018 saw month-on-month rises in the number of unique IPs accessing the website, going from approximately 500,000 in 2011 to 1.2 million in May 2018[11]. Figure 2, below, shows that the total number of unique IP addresses that accessed the Europe PMC website each year rose from 9 million to almost 12 million between 2013 to 2018. Note that the 2017 data shown is considered unreliable because of suspected bot activity from China.

Figure 2 Unique IPs accessing Europe PMC per year



Source: Europe PMC data from EMBL-EBI

Figure 3 Geographical Distribution of IPs (2018)



Top 10 countries	% IPs
United States	23.2
China	12.7
India	6.9
United Kingdom	6.6
Canada	3.3
Australia	3.1
Brazil	2.6
Germany	2.5
Indonesia	1.6
Italy	1.6

Source: Europe PMC data from EMBL-EBI

[11] Data provided by Europe PMC team at EMBL-EBI

Europe PMC website statistics show that Europe PMC has a very global reach (Figure 3). It is important to note however that the data presented are a snapshot of the distribution in 2018 and may change year on year for both countries and continents. In 2018, the top 3 countries with the largest share of total visits to Europe PMC websites were the US, China and India.

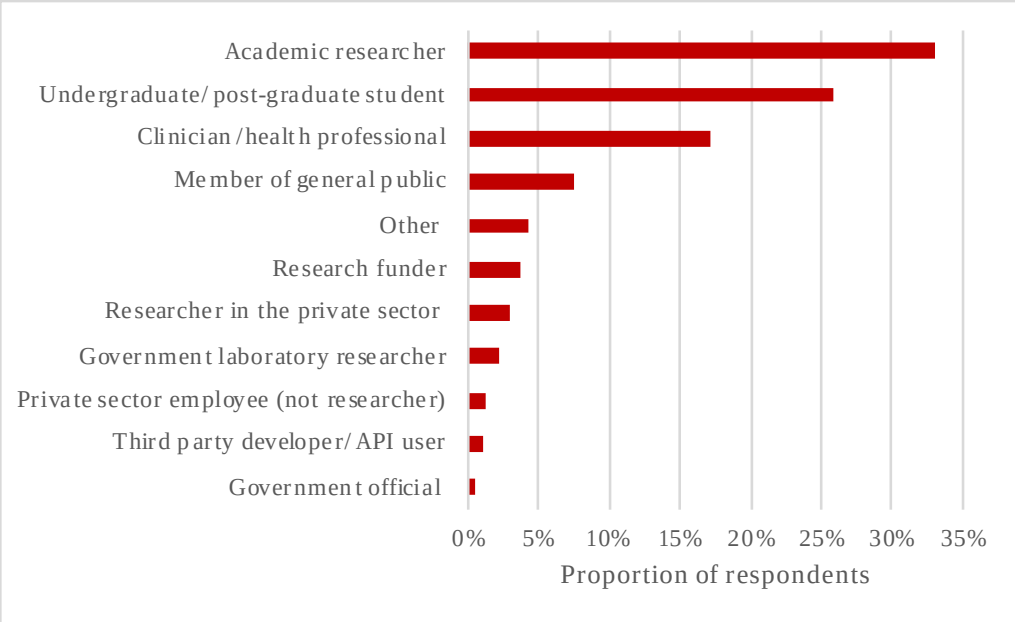
Approximately 12 million unique IPs accessed the Europe PMC website in 2018.

The US, China, India and UK collectively account for almost half of these unique IPs showing that Europe PMC has a global reach.

3.2 Types of users

We conducted a pop-up user survey on the Europe PMC website (termed website survey in this report, see Appendix A and Appendix C), allowing us to estimate the types of users accessing Europe PMC services. Similar to the previous user survey run by EMBL-EBI in 2017[12] (see Appendix E), we found that academic users, students and clinicians/health professionals accounted for the majority of users (Figure 4).

Figure 4 Distribution of website survey respondents by user type (n=958)



Source: Technopolis analysis of website survey

Academic researchers, students and health professionals constitute 76% of Europe PMC users in the website survey.

[12] The user survey was conducted during the 6-week period from 29 June to 8 August 2017. This survey attracted a total of 389 respondents and showed that Europe PMC users are typically researchers (20%) and students (16%) and based in universities (37%). Data presented in Report of Europe PMC Funder Committee Meeting, 9 October 2017.

4 Europe PMC services and functionality

As Table 1 below shows, Europe PMC offers a range of different services relevant for researchers, funders, and developers/API users. Broadly speaking, the services fall into four main categories: (1) providing a repository for research outputs, (2) enabling users to find relevant data and articles, (3) providing the infrastructure to interrogate individual resources, and (4) facilitating third party users who wish to examine or analyse all Europe PMC's resources as a whole (e.g. through API).

Table 1 Europe PMCs main services and functionality

Services / functionality	Description	Potential users[13]
Searchable publication repository	Allows users to search for and download research articles (abstracts and full text), AGRICOLA records[14], NICE clinical guidelines, patents etc. Also has a list of journals that provide full text articles to Europe PMC, including a list of Open Access journals.	Researchers Clinicians Research funders Members of the general public Developers/API users
Links to sequences, molecular structures, citations and data	Users can link directly to DNA and protein sequences and structures in external databases, underlying data sets and citations from the Europe PMC website. Number of citations are also highlighted.	Researchers Clinicians Research funders Members of the general public
ORCID (Open Researcher and Contributor ID) article claiming	Allows researchers to link their Europe PMC articles to their ORCID, helping distinguish them from other researchers	Researchers Clinicians Research funders
Grant finder	Enables users to find grants awarded by Europe PMC funders. Can also be used to find research in a given field, and find potential collaborators	Researchers Clinicians Research funders Public officials Members of the general public
External links services	Enables users to add links from research articles in Europe PMC to related blog posts, news, articles or tools	Other databases and data repositories Funders and services related to publishing
RSS feeds	Enables users to add RSS feeds to their reader	Researchers Members of the general public
SciLite annotations	Allows users to highlight annotations on full text articles	Researchers Clinicians
Europe PMC Plus	Enables users to submit manuscripts (funded in whole or in part by Europe PMC funders), link published papers to grants, and view and export publication and citation data	Researchers Clinicians Research funders
Articles RESTful API	Access publications and related information	Researchers Clinicians Developers/API users
Grants RESTful (Grist) API	Access grant information from Europe PMC funders	Research funders

[13] Here, the user categories specified include both academic and non-academic users including those from the private sector.

[14] AGRICOLA (AGRICultural OnLine Access) is an online database created and maintained by the United States National Agricultural Library of the United States Department of Agriculture. It indexes a wide variety of publications covering agriculture and allied fields including animal and veterinary sciences, entomology, plant sciences, forestry, aquaculture and fisheries, farming and farming systems, agricultural economics, extension and education, food and human nutrition, and earth and environmental sciences.

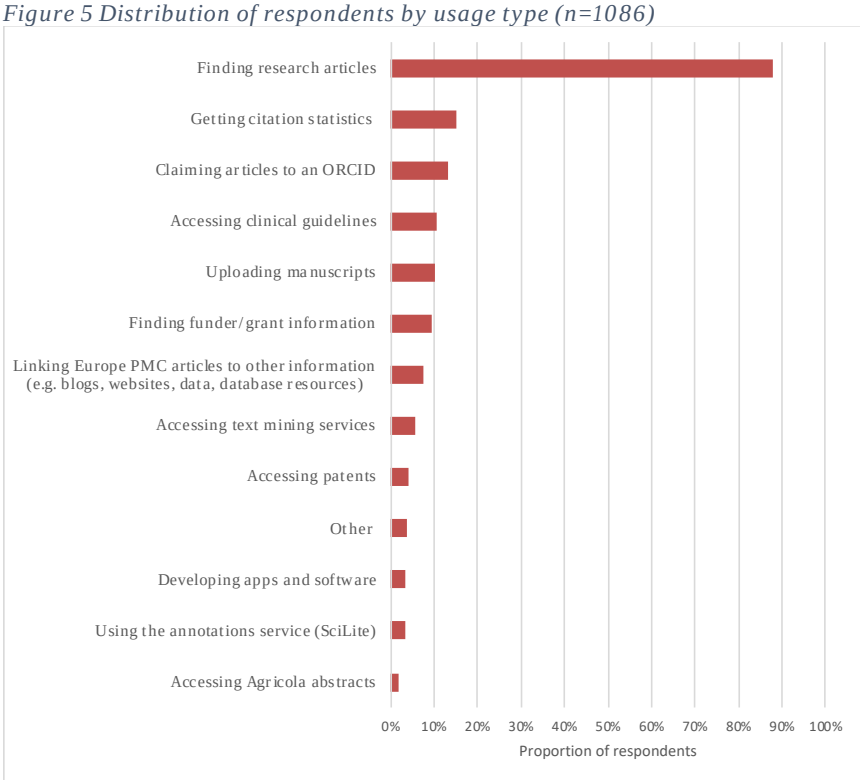
Services / functionality	Description	Potential users[13]
		Policy makers Developers/API users
SOAP web services	Retrieve publication data using Simple Object Access Protocol (SOAP)-based service	Researchers Clinicians Developers/API users
Open Archives Initiative (OAI) service	Gives users access to the metadata of all items in the Europe PMC archive, and full text of some of these items	Researchers Clinicians Developers/API users
Bulk downloads	Provides options to download all Open Access articles, plus metadata of all full text articles in Europe PMC.	Researchers Clinicians Developers/API users
Annotations API	Helps users access annotations contained in abstracts and Open Access articles	Researchers Clinicians Developers/API users
User support	The Europe PMC operations team at EMBL-EBI also provides technical support to users in case of queries through a dedicated helpdesk and the Google API users community Group	All users

Source: Europe PMC website and Technopolis scoping interviews

4.1 Types of usage

Users primarily access Europe PMC to find research articles (see Figure 5). Other uses include meeting funders' open access requirements, linking publications to researcher ORCIDs, and using it as a data source for other similar open access services/platforms (e.g. Unpaywall and OpenAIRE) as well as new apps or software products. Finding full text and open access articles, ORCID linking and checking open access compliance were also the three main reasons why librarians might use Europe PMC (Appendix E.2

Figure 5 Distribution of respondents by usage type (n=1086)



Source: Technopolis analysis of website and email-based researcher surveys

Europe PMC is primarily used to find and download research articles.

Other similar services and developers use the Europe PMC APIs as free and barrier-less resources to the information they need. For instance, Europe PMC operates a data mining pipeline for the Open Targets platform to identify associations between a target protein and a gene from Europe PMC content (abstracts and full text articles). Similarly, Europe PMC is a core data resource for ELIXIR[15], a European initiative to unite leading bioinformatics infrastructures across Europe. ELIXIR helps connect academic and industry users across European member states to knowledge and data generated from publicly-funded research programmes. ELIXIR relies on Europe PMC infrastructure to gather information from publications on how ELIXIR resources are used in research. Europe PMC also enables ELIXIR to connect EMBL-EBI data to national data, in turn helping Member States collect data on funded research. Another example is ContentMine, a software development company that provides text and data mining tools and services, and which uses Europe PMC as its primary source of data (see case study below).

Case study: ContentMine, a text and data mining company

ContentMine, established in 2014, is a not-for-profit text and data mining company specialising in the creation of free, open source tools that allow users to glean information from high volumes of published literature. The software focuses on reducing the time required to undertake lengthy literature reviews and manually extract data from them.

The company provides a number of tools, including the option for clients to commission bespoke solutions based on their specific needs. ContentMine has developed tools allowing users to access information contained within the tables and graphs of published papers, something which traditionally has required manual extraction by the reader. Automatic data extraction can therefore expedite the

[15] https://www.elixir-europe.org

analysis or indeed meta-analysis of the raw data contained within these data sources, making literature reviews much more efficient.

Europe PMC is ContentMine's primary source of open access data. The company chose Europe PMC because it offers a robust, competent and reliable service, allowing ContentMine and its users access to freely available articles, books, patents and clinical guidelines. The rapid download times are also key. The broad-ranging content offered by Europe PMC makes it an ideal source for undertaking pilot searches to scope the quantity and quality of information available on a specific topic, providing researchers with an overview of what has already been published within a research area and allowing them to direct their research better.

ContentMine is based on the Europe PMC Articles API. The process of familiarisation with the API and incorporating it in workflows has taken a significant amount of time and effort. If Europe PMC became unavailable, ContentMine would need to quickly adapt to using an alternative data source, requiring more time and effort for familiarisation and customisation of workflows and additional funds if the alternative data source is not free.

4.2 What users like about Europe PMC

In the surveys and interviews, users identified some unique and helpful features of Europe PMC. Particularly valued aspects included ease of use, good coverage in terms of content and volume, and free access to full-text publications as well as APIs.

Several interviewees stated that Europe PMC fills a gap in the landscape by providing free access to biomedical content, thus supporting the Open Science agenda. It fulfils a critical role in making life science and biomedical content mineable, reusable, interoperable, permanently accessible and discoverable, stressing that Europe PMC offers significantly more than a simple document repository. It provides security and privacy that other services do not necessarily provide. According to one interviewee "*Europe PMC works because it aligns with the European model of multinational collaboration and unifying around a shared infrastructure. Europe PMC is a working model of what the rest of the Open Access community can be working towards.*"

In addition, Europe PMC was seen to provide added value through functionalities such as a good user-friendly interface and layout, ORCID linking, fast download speeds, links to citations, full-text searches, annotations, and access to wider material including NICE guidelines, patents, AGRICOLA records and university theses. Users also highlighted the good quality of the underlying metadata and reading accessibility. Indeed, one academic researcher remarked that they can find more references using Europe PMC because of the ability to do full text searches as well as links to citations, ORCIDs and annotations, making it easier to find related papers.

Individual interviewees also pointed to the robust and competent service provided by Europe PMC, the linking of research with grants, the availability of pre-prints, and the option for green open access (self-archiving by author) as unique features of Europe PMC.

Less well-resourced researchers, such as those from developing countries, citizen scientists and members of the general public without access to paid publisher services particularly appreciated the large volume of full free text articles available through Europe PMC (see case study below). Article access helps them keep abreast of new and existing research in their fields of interest. For some researchers in this category, Europe PMC is just one of several sources for full-text documents including PMC, Google Scholar and ResearchGate. Interestingly, many users seldom go directly to the Europe PMC website, but rather discover links to articles in Europe PMC from search engines or another site. This observation from our survey and interviews chimes with findings from the 2017 survey (see Appendix E).

Case study: Impact of Europe PMC on an academic researcher in a developing country

Researchers form the largest user group of Europe PMC and are therefore a prominent source of impact. To understand the significance of Europe PMC for researchers, we spoke with a senior academic scientist

at a government-funded institute in Northern India who conducts independent research within the field of cellular immunology, while also mentoring PhD students and post-doctoral scientists.

Europe PMC enables this scientist to access scientific publications within his research area and to stay at the forefront of current developments and advances. He also uses the repository to gather recent evidence before undertaking specific research projects, often accessing the platform for up to 7 hours per week. Key benefits include access to relevant articles in an open-access format via a user-friendly interface.

Obtaining academic research funding in India (currently a lower-middle-income economy[16]) is often difficult, due to a lack of available funds. Losing access to the services that Europe PMC provides is highly likely to have a significant impact on the ability of scientists in LMICs to carry out their research. Europe PMC currently allows the academic we interviewed to view a broad range of open access papers that provide the knowledge he requires to stay at the forefront of research in cellular immunology. Paying for a such a service would divert valuable funds away from his research projects and laboratory. For this researcher, the mission of Europe PMC to provide open access to research publications is at the core of its value. Hence, he would like to see awareness of Europe PMC grow in the future.

Users particularly appreciate Europe PMC features such as ease of use, good content coverage, free access to full-text publications and the APIs.

4.3 What users think can be improved

While about 20% of interviewees struggled to identify any gaps or areas for improvement in Europe PMC, a commonly cited perception was that many potential users are not aware of Europe PMC and the added value that it could provide. As such, interviewees advocated greater exposure of Europe PMC through marketing and outreach. In addition, some interviewees highlighted the potential for Europe PMC to play an important role in the context of Plan S[17] – a European initiative that is pushing for all European published research to be made available via Open Access.

A second suggested change was to have more extensive access to full text documents, including from wider disciplines. However, it was acknowledged that such expansion of the service might have cost implications.

The third set of suggestions concerned improving the search functionality. Suggested changes included using Natural Language Processing, allowing searching by first or last authors only, the option of doing full text searches on the main webpage (not just through the advanced search function) and improving the power and precision of searches when using content mining tools. An academic librarian and a third-party developer also recommended better linking of metadata to grants.

Other suggested areas for improvement included the need for a more user-friendly interface for lay people and the option of conducting larger bulk downloads as in Semantic Scholar.

Users feel that Europe PMC and its benefits are not publicised enough.

[16] https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

[17] https://www.scienceeurope.org/wp-content/uploads/2018/09/Plan_S.pdf

5 Comparison of Europe PMC to other services

5.1 Comparator services

A variety of different platforms offer similar services to Europe PMC. While none of these on their own currently provide exactly the same range of services as Europe PMC (see Table 2), collectively they cover most of Europe PMC's services and functionalities. For example, PubMed and PubMed Central cover the same publisher content, citation counts can be obtained from Web of Science or Scopus (although not free), and instead of linking to external databases from the Europe PMC website, users could access the relevant data directly from the external database's own website. The interviews highlighted a lack of awareness and familiarity with Europe PMC services among potential users, particularly researchers. Thus, users might be opting for a multitude of platforms for different uses even though a single platform like Europe PMC might be more suited to their needs.

Table 2 The main comparator services to Europe PMC

Service	Funders	What it offers	How is it different to Europe PMC?	How is it similar to Europe PMC?
PubMed / PubMed Central (PMC)	The United States National Library of Medicine (NLM) and the National Institutes of Health (NIH)	- PubMed consists of a database of citations and abstracts for more than 25 million biomedical articles. - PMC is an electronic full-text archive of over 3 million biomedical and life sciences journals and online books.	PubMed and PMC are separate resources while Europe PMC provides a single search and point of access to PubMed, PMC, and an additional 5 million resources including patent records.	- Allows users to search for and access research articles (abstracts and full text) - Provides access to citation metrics - Provides external links service - Allows users to access content through APIs - Allows bulk downloads
Google Scholar	Google	A freely accessible search engine that indexes the full text or metadata from scholarly literature. It enables searches across many disciplines, and across sources including articles, theses, books, abstracts and court opinions.	Unlike Europe PMC, it does not provide materials on an open access basis, but it is broader in scope. It does not support APIs either.	- Allows users to search for and access research articles (abstracts and full text) - Allows authors to link their publications to their ORCID - Provides web-based user support
OpenAire	European FP7 programme	Infrastructure that aggregates open access publications and research data catalogues, as well as linking them to funding streams It is also built on a fully open source infrastructure framework, enabling third-party users to customise and extend its functionality.	OpenAire can perform advanced research analytics on its platform, targeted at those monitoring funded research.	- Allows users to search for and access research articles (abstracts and full text) - Allows the user to access related publications, data and funders -Allows the author to link their publication to ORCID, associated data sets and funders - Allows users to access content through APIs - Allows bulk downloads - Provides user support via an online helpdesk
Web of Science	Subscription-based (run by Clarivate analytics)	Scientific citation indexing service, drawing on books, peer reviewed journals, original articles and abstracts. Also provides links to full texts.	Unlike Europe PMC, Web of Science is a subscription service. It also covers all disciplines.	- Allows users to search for and access research articles (abstracts and full text) - Links to funder information, related data and citation metrics

Service	Funders	What it offers	How is it different to Europe PMC?	How is it similar to Europe PMC?
				- Allows the user to set up citation alerts - Allows users to access content through APIs - Provides web-based user support
Scopus	Subscription-based (owned by Elsevier)	Abstract and citation database of peer-reviewed scientific journals, books, and conference proceedings Also supports APIs	Scopus is also a subscription service and covers all disciplines. It also has smart tools to help track, analyse and visualise research.	- Allows users to search for and access research articles (abstracts and full text) - Allows bulk download of search results - Allows the user to set up citation alerts - Allows authors to link their publications to their ORCID - Allows users to access content through APIs
ResearchGate	ResearchGate GmbH	A social networking site for scientists and researchers to share papers, ask and answer questions, and find collaborators	Europe PMC has a more diverse range of services. ResearchGate relies on researchers depositing their own research so all open access life science publications are not readily available. Articles do not need to be peer-reviewed prior to upload.	- Allows users to access research articles (abstracts and full text) - Provides citation metrics - Allows the author to link their publication to associated data sets

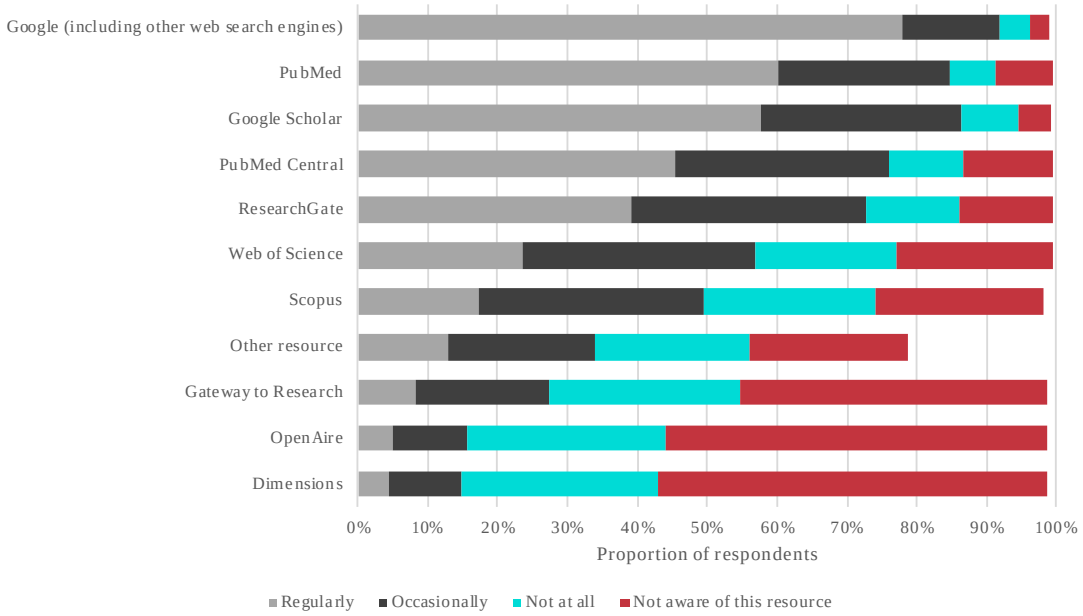
Source: Online research

As a whole, user survey respondents (both website and email-based surveys) most regularly use search engines (e.g. Google search), PubMed, PMC and Google Scholar (see Figure 6). This was true of both Europe PMC users and non-users (data not shown). The 2017 user survey also found that PubMed, PMC, Google and Google Scholar were the most frequently used comparator services (see Figure 31). Thus, Europe PMC appears to provide an integrated solution that is one among many different tools used. Other services that survey respondents also use include publisher websites, Medline, Embase, SciFinder, Cochrane Library, Ovid, Sci-Hub and ScienceDirect.

When we asked librarians which services they promoted or publicised to their users (e.g. through user training, or links on their website), PubMed, Web of Science, Google Scholar and PMC were most frequently mentioned (Figure 7). Europe PMC ranked eight out of the 12 options given. Other services that libraries and/or information portals publicised included EBSCO/CINAHL, Embase, Ovid, Cochrane Library, PsycINFO and Medline.

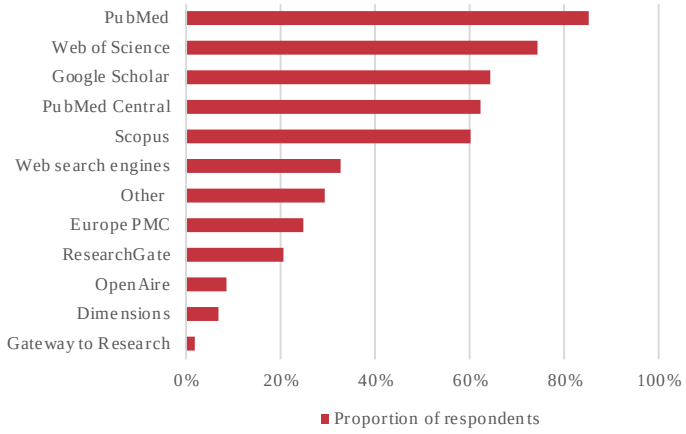
Europe PMC provides an integrated solution that is one among several different tools used to find biomedical content. PubMed, PMC, search engines and Google Scholar are the services that are used most commonly.

Figure 6 Use of other services by user survey respondents (n=1256, includes Europe PMC users and non-users)



Source: Technopolis analysis of website and email-based researcher surveys

Figure 7 Services promoted by libraries or information services (n=137)



Source: Technopolis analysis of librarian survey

5.2 Pros and cons of Europe PMC compared to other services

Interviewees found it difficult to draw a direct comparison between Europe PMC and other services, mainly because the functionalities and business models are rather different. For example, services such as Scopus, Web of Science and Science Direct are commercial, profit-making enterprises and some such as OpenAIRE themselves use Europe PMC.

Nonetheless, according to interviewees, the advantages of Europe PMC lie in its open access provision, more user-friendly interface and better functionalities. For example, seven of the 25 interviewees prefer Europe PMC to PMC for reasons including a better interface, faster search function (about 4X faster according to one interviewee), better availability of DOIs, and a broader range of services. Interviewees from two platforms that partner with Europe PMC highlighted that Europe PMC is nimbler and more responsive to demands for new functionalities compared to PubMed/PMC. One stated that "Europe PMC staff are always thinking about the most modern ways of using data". In contrast, PubMed and

PMC were seen as being less agile in responding to technical demands. Two interviewees (one clinician, one other stakeholder) commented on better functionality and more specific search results from Europe PMC compared to Google Scholar. Europe PMC is also the primary choice for companies such as Impactstory to provide content for their open access services (see case study below).

Case study: How Europe PMC supports Unpaywall

Unpaywall is an open source, free database of open access articles. It is a project of the nonprofit organisation, Impactstory, and harvests open access content from over 50,000 sources including Europe PMC, PMC, publisher sites and university repositories for its users. Unpaywall also has a publication API to allow commercial and non-commercial use of the database similar to what Europe PMC provides. The difference between the two is that while Europe PMC links to articles in PMC, Unpaywall links to articles in PMC, a publisher site or a university repository as appropriate.

Europe PMC is one of Impactstory's preferred sources because of better designed and easier to use metadata, machine interfaces and APIs; features such as links to data citations, PMID to DOI links and author profile pages with links to ORCID; accessibility of the staff; and greater agility than other similar services. In particular, the Europe PMC publication API has some attributes that are not easily accessible in PMC. For instance, Europe PMC tags whether a manuscript is from the author or the publisher, and its .json format is easier to use than the .xml format that PMC provides.

Two examples illustrate how Europe PMC aids Unpaywall. The first concerns how Unpaywall provides information about an open access publication. Any publication that Unpaywall links to is annotated with metadata on the URL location e.g. publisher site or repository, and version e.g. final published version, accepted version or submitted version (not peer-reviewed). The Europe PMC API makes it easy and quick to find this information and put it in Unpaywall's API, which may in turn be used by Unpaywall customers in their own dashboards. The second example is about how Unpaywall provides links to an article pdf. Europe PMC configures links to a pdf in a more user-friendly manner than PMC. Thus, even though a publication may be open access through PMC, Unpaywall directs users to the Europe PMC pdf link. In addition, Europe PMC's pdf viewer is better and simpler to use.

Owing to all its helpful features, Unpaywall constantly uses Europe PMC throughout the day. Even so, it is hard to estimate the time saved as a result of using Europe PMC since there is currently no other good alternative. Nonetheless, without Europe PMC, it would be a lot harder to source the same data from other providers, resulting in Unpaywall coders spending more time completing tasks that are currently routine for them.

Apart from seven interviewees who indicated a preference for using Europe PMC, most of the other interviewees did not indicate a preference for a particular comparator service. However, three individuals (two researchers, one librarian) reported using Google Scholar and PubMed two- to 20-fold more often than Europe PMC, mainly out of habit. A clinician mentioned that he only uses Europe PMC when a publication is not available through PubMed. Some librarians noted (in the survey) that they were unaware of differences between Europe PMC and PMC, with PubMed and PMC more popular among their users.

Some interviewees spoke of the downsides to Europe PMC. One publisher spoke of the inability to perform large data dumps. According to him, Semantic Scholar is better for this purpose, allowing him to download much larger datasets and then locally run code off them. However, this user may have been unaware that it is possible to bulk download full text open access content and other datasets via FTP. Another interviewee, an academic researcher, added that some of the semantic text options in PubMed work better and are more extensive than those currently in Europe PMC. He said that PubMed also benefits from having some slightly more nuanced search terms, such as searching by first authors only.

The main advantages of Europe PMC compared to comparator services are the open access provision, more user-friendly interface and better functionalities such as a faster search function, better availability of DOIs and a broader range of services.

5.3 Europe PMC versus PMC usage statistics

In terms of full text usage, in 2016, the number of Europe PMC hits as a percentage of PMC hits was approximately 5% (744 million versus 32 million), showing PMC's dominance in the field[18]. Similarly, in 2017, Europe PMC accounted for around 3% of the total number of unique user IPs accessing PMC or Europe PMC every month (total number ranging from 23.3 million to 35.6 million unique IPs)[19]. Moreover, out of a total of 1.4 billion articles retrieved across all PMC sites (including Europe PMC) in 2017, Europe PMC accounted for roughly 2% of retrievals[19].

Table 3 Comparison of Europe PMC and PMC usage statistics (June 2017 to November 2018) for Wellcome Open Research, Proceedings of the National Academy of Sciences (PNAS) and PLOS Biology

	Unique user IPs	Total available items	Total items accessed (HTML full text, article PDF, scanned summary and other pages)
Wellcome Open Research			
Average monthly usage in Europe PMC (count)	146	378	312
Average monthly usage in PMC (count)	3,605	237	5,690
Europe PMC usage as % of total usage	3.9%	61.4%	5.2%
Proceedings of the National Academy of Sciences (PNAS)			
Average monthly usage in Europe PMC (count)	22,282	115,011	96,725
Average monthly usage in PMC (count)	625,818	138,189	1,578,964
Europe PMC usage as % of total usage	3.4%	45.4%	5.8%
PLOS Biology			
Average monthly usage in Europe PMC (count)	2,350	4,382	9,972
Average monthly usage in PMC (count)	57,077	4,726	100,556
Europe PMC usage as % of total usage	4.0%	48.1%	9.0%

Source: PMC, PLOS and The Wellcome Trust, analysis by Technopolis

The comparison above shows that in terms of unique user IPs and total items accessed, traffic via Europe PMC is a maximum 9% of total usage across both PMC and Europe PMC. On the other hand, Wellcome Open Research has more items available via Europe PMC than PMC, while the situation is reversed for the other two journals.

Europe PMC is used at least 10 times less than PMC based on a comparison of access to three scientific journals.

[18] Report for Europe PMC Funder Committee Meeting, 9 October 2017

[19] Data from PMC. PubMed usage data not included.

6 Economic Valuation

6.1 Cost of running Europe PMC

We set out to calculate the operating cost of Europe PMC per user per year. Table 4 below represents the breakdown of the anticipated costs of running Europe PMC from 2016 to 2021 in terms of the grant provided.

Table 4 Summary of financial support requested for Europe PMC (2016 to 2021)

Duration	Salaries	Software, provision for engagement, third party costs, miscellaneous	Equipment	Total grant amount
60 months	£ 5,134,739	£ 567,316	0	£ 5,702,055

Source: EMBL-EBI grant application to The Wellcome Trust (2015)

Additional funders have started supporting Europe PMC since the original grant application and are contributing funds to the tune of £96,472 annually. In-kind contribution from EMBL-EBI was estimated to be approximately £5.3 million in 2015[20]. Europe PMC also benefits by getting a large volume of its content in a structured format from PMC free of charge. Hence, its running costs are relatively low.

Using the average number of unique IP addresses that access the Europe PMC website each year as a proxy for the total number of users, we can estimate the cost per user per year from the total costs described below (Table 5).

Table 5 Cost per user per year for running Europe PMC (2016 to 2021)

Total costs per year (£)*	Total costs per year ($)***	Average number of unique IPs per year (2013 to 2018)**	Cost per user per year (£)	Cost per user per year ($)***
2,294,190	3,277,414	10,037,799	0.23	0.33

*Grant amount (Table 4) plus estimated in-kind contributions[20] and contributions from funding partners
**2017 monitoring data not used because of suspected bot activity from China
*** Using OECD's 2017 purchasing power parity (PPP) exchange rate. Thus, the $ amount here is adjusted for what it would cost to buy the same amount of goods or services in the US in USD.
Source: EMBL-EBI grant application to The Wellcome Trust (2015)

Running Europe PMC costs $0.33 per user per year

6.2 Economic valuation of Europe PMC

It is difficult to put a value on access to knowledge, which is one of the main impacts of Europe PMC. However, the economic value of an infrastructure such as Europe PMC can be estimated using proxies. We have examined the value of Europe PMC using two different proxy measures

1. Monetised value of time spent using Europe PMC services – usage value modelling
2. Value users put on Europe PMC services – contingent valuation

Usage Value Modelling

One of Europe PMC's main impacts is enabling access to and usage of a range of resources. Users value the infrastructure by spending time on it. We placed a monetary value on this usage, as opportunity cost, using a methodology with three main parts

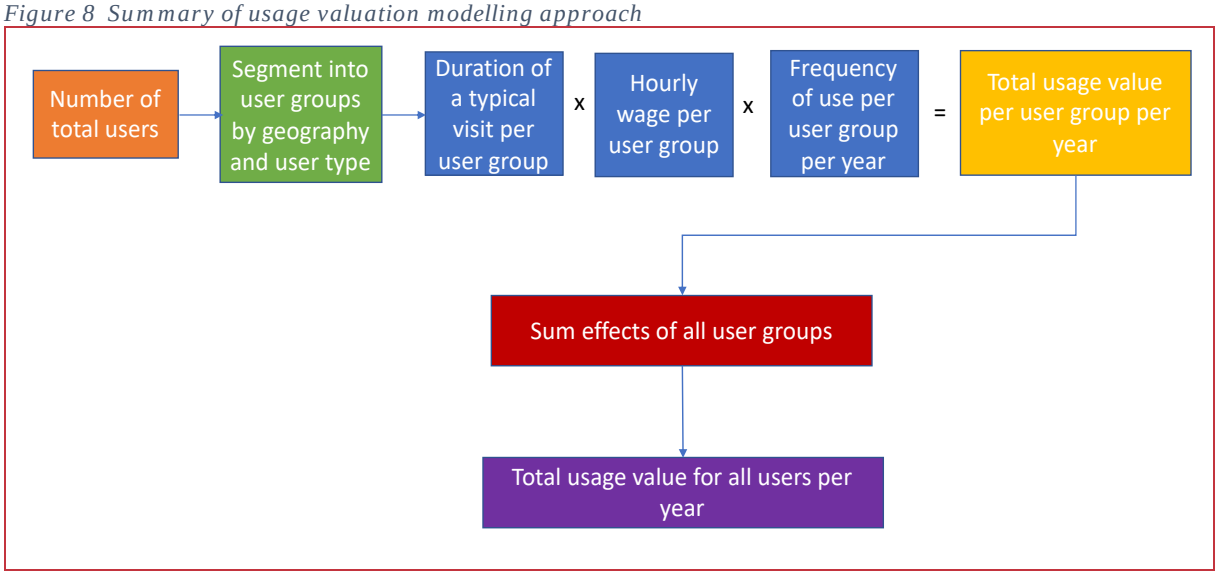
- Dividing all Europe PMC users into different user groups/categories

[20] Europe PMC team at EMBL-EBI (as provided in August 2015 for the Europe PMC grant review)

- Calculating the monetary value of each group's Europe PMC usage (total effect) based on time spent using the service and publicly available wage data
- Summing the effects for each user group to reach a combined value for all Europe PMC users

The flow chart below shows the various steps that are involved in this approach. We discuss each of these steps below, with Appendix F providing more detailed explanations, including the key assumptions used.

Figure 8 Summary of usage valuation modelling approach



Step 1: Segmenting the total users

Initially, we took the total number of unique users (unique IPs) and segmented these by geography i.e. whether they belonged to a developed, developing or transition economy (we used 2018 monitoring statistics as these had country level breakdown, see Figure 3). The exact split of users can be seen in Table 15.

For each geographical segment, we carried out a further segmentation exercise, apportioning each segment to different user types (i.e. academic researcher, clinician, non-academic researcher, general public, research funder, industry and student) based on findings from our website survey which we have assumed to be representative of the typical Europe PMC visitor profile. In other words, our website survey found that 29% of users in developed economies are academic researchers and we have segmented the developed economy segment accordingly.

Based on these assumptions, we have divided our population into 21 segments; the numbers in each of these are detailed in Table 16.

Step 2: Time usage per user group

For each of the 21 segments, we then calculated the median duration of each visit, drawing on information provided by all respondents to our website and email-based researcher survey that indicated they were Europe PMC users. Note that we opted for using self-reported usage times from surveys as it allowed for obtaining consistent data for each user segment. The median time per visit of each user segment is provided in Table 17.

We chose not to calculate our figures using the mean as there are a small number of users whose visits tended to be disproportionately long in time. Using the mean would risk us using an over-inflated average. Similarly, we have not used the mode because the large sample size means that there were very few respondents sharing precisely the same duration per visit.

Step 3: Frequency of visits by each user segment

For each user segment, we calculated the mean number of visits made, drawing again on data for website and email-based researcher survey respondents. For our calculations, we included all respondents that indicated they were Europe PMC users, and who answered the question 'How many times on average do you visit Europe PMC in a month?' We summarise our results for this step in Table 18 with full commentary on the assumptions provided in Appendix F.

Step 4: Usage value per hour

Finally, for each user segment, we determined a suitable proxy median hourly wage by using International Labour Organization statistics[21]. Appendix F provides further details on the assumptions and processes underpinning this step and summary figures are provided in Table 19.

Step 5: Calculating the usage value per segment per year

By multiplying the segment specific figures calculated in steps 1-4, we put a value to the time spent by each user group per year on Europe PMC. This involved:

1. Multiplying the median duration of a visit (Step 2) by the mean visits per month (Step 3) for each user segment. This gave us the average duration of a visit (in minutes) per month.
2. From this, we derived the average duration of a visit (in hours) per year (by dividing by 60 and multiplying by 12) for each user segment
3. We multiplied this figure with the usage value per hour for each user segment (Step 4), giving us the value of Europe PMC usage per user per year
4. Next, we multiplied the usage value per user per year by the number of users in each respective segment (Step 1), which gave use the total usage value per year for each user segment

Table 20, Table 21 and Table 22 show the economic value associated with Europe PMC usage for each of the 21 segments.

Step 6: Calculating the value of Europe PMC usage across all users

Having calculated the usage value for each of the 21 segments, we summed these up to reach a total value of Europe PMC usage per year across all users. As shown in Table 6 below, our calculations indicate that the total usage per year for Europe PMC across all user types and geographies is $1.5 billion.

Table 6 Total net value of Europe PMC usage per year

Geography	Total net value (PPP $ per year)
Developed economies	1,152,490,377
Transition economies	18,792,215
Developing economies	291,681,802
Total	1,462,964,394

Source: Technopolis analysis

Our estimate, however, assumes full attribution of the value generated by accessing and using knowledge to Europe PMC, and that there is no alternative directly comparable service. However, alternative platforms collectively do cover most of Europe PMC's services and functionalities (Table 7). Therefore, we believe it is appropriate to account for some 'deadweight' to this usage value (deadweight being any effects that would have been seen even in the absence of Europe PMC; i.e. the higher the deadweight the

[21] https://www.ilo.org/ilostat/

less unique the infrastructure). In other words, not all of the usage value attached to Europe PMC is unique – some of the value generated could very easily be transferred to another service.

Table 7 Alternatives that replicate Europe PMC services

Services / functionality	Alternative services
Searchable publication repository	PubMed/PMC, Google Scholar, Researchgate (free access) Scopus, Web of Science, Science Direct, Publisher websites (paid access) AGRICOLA, NICE, European Patent Office websites (free access)
Links to sequences, molecular structures, citations and data	DNA and protein sequences and structure databases – Uniprot, ENA, PDBe and BioStudies (literature to data links offered by Europe PMC not replicated) Citations from PubMed and Researchgate (free); Scopus, Web of Science, Dimensions (all paid services)
ORCID (Open Researcher and Contributor ID) article claiming	ORCID, Google Scholar, Scopus
Grant finder	Gateway to Research (free), NIH RePORT (free), Dimensions (paid), other national services
External links services	PMC
RSS feeds	Not available
SciLite annotations	Not available
Europe PMC Plus	PMC (depending on arrangement with funder), institutional repositories (text mining function of Europe PMC not replicated)
Articles RESTful API	PMC (free); Scopus and Web of Science (paid)
Grants RESTful (Grist) API	Dimensions (paid, based on Europe PMC)
Open Archives Initiative (OAI) service	PMC (free); Scopus and Web of Science (paid)

We have used the Scottish Enterprise ready reckoner (see Table 14) to account for deadweight in our usage valuation. However, we know that deadweight levels will differ depending upon the type of use and individual circumstances of the users. For instance, one academic researcher may have many more alternative or comparable services available to them simply because their institution is able to afford access to a more diverse range of platforms. On the other hand, some alternative services such as OpenAIRE, Dimensions and Unpaywall use Europe PMC data themselves. Besides, Europe PMC may represent additional benefits to users (e.g. time saved) through providing an integrated platform for multiple functionalities, a feature not replicated by any another service. As such, we have considered two different scenarios of deadweight: high deadweight (i.e. most Europe PMC services can be replicated exactly elsewhere), and very high deadweight (i.e. almost all Europe PMC services can be replicated exactly elsewhere). These scenarios are considered in Table 8.

Table 8 Usage valuation of Europe PMC under different deadweight scenarios (based on Scottish Enterprise ready reckoner in Table 14)

Scenario	Total usage value per year for all users (PPP $)	Usage value per user per year (PPP $)
0% deadweight	1,462,964,394	125
High deadweight (75%)	365,741,099	31
Very high deadweight (95%)	73,148,220	6

Source: Technopolis analysis

Taking deadweight into account therefore, the total economic value of Europe PMC usage ranges from $73 million to $366 million per year, and $6 to $31 per user per year. If we compare the total usage value per year to the total cost per year (in PPP$), we can determine that for every dollar spent on Europe PMC, the value returned is $22 to $112 for 95% or 75% deadweight, respectively.

For every dollar contributed to Europe PMC, the usage value returned ranges from $22 to $112 depending on deadweight[22].

It can be argued that measuring the value of an infrastructure based on time spent may result in inefficient services being valued higher than a more efficient service that does the same work in a shorter time. While this argument cannot be completely disregarded, considering that there are several alternative services available, we would assume that users would stop using Europe PMC if they found it inefficient. Moreover, ease of use and availability of several features from a single interface are aspects that users appreciated about Europe PMC according to our surveys and interviews. Another point to note is that while we have used the best available data including self-reported data and unique user IP numbers for our valuation, these have their limitations. Self-reported data from surveys is subject to cultural-strategic bias. Moreover, IP addresses do not give a completely accurate estimate of users. For instance, several users may be behind one IP address while a single user may be behind multiple IP addresses if they tend to work across different sites or devices. Automated and robot activities can also account for some unique IP addresses. However, our underlying assumption is that these respective under-estimates and over-estimates will largely cancel each other out, making unique IP addresses a reasonable proxy for total Europe PMC user numbers.

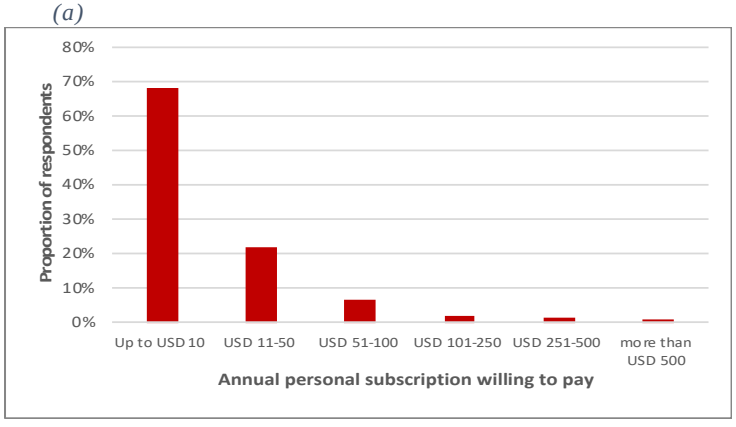
Contingent valuation

Contingent valuation is a way of estimating the value that a person places on a good or service. It involves asking people directly about their willingness to pay (WTP) to obtain a specified service in a hypothetical situation where access is no longer free. In the case of Europe PMC, we can place a contingent valuation by multiplying the mean WTP among users by the total number of users (unique IPs per year).

In the context of this study, we assumed that librarians would be the relevant 'expert group', more able to provide a realistic WTP figure compared to users, owing to their familiarity with subscribing to other databases and services. Therefore, the librarians survey was our primary tool to determine a WTP amount. However, for comparison, we also included a WTP question in the user surveys.

Figure 9 Analysis of willingness-to-pay responses by amount (a; n=856) and common reasons cited for choice (b; n=381)

(a)



(b)

Reason for choice	No. of responses
Low income	104
Believe access should be free	48
Service is worth paying for	48
Other alternative available	43
Low usage	26

Source: Technopolis analysis of website and email-based researcher surveys

[22] This is total usage value per year for all users in Table 8 divided by total cost per year in PPP $ to run Europe PMC as shown in Table 5. As this is a ratio of PPP $ per year to PPP $ per year, the units cancel out each other.

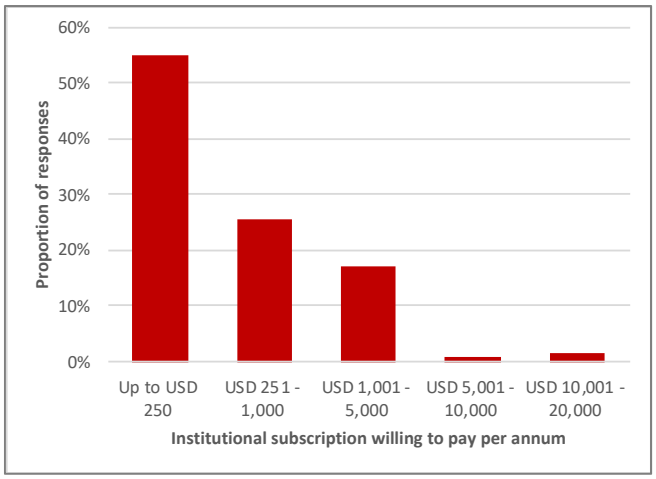
Most users (68%) were willing to pay only up to $10 (the lowest range) to use Europe PMC services. Some respondents also gave explanations for their choice. Reasons cited were low income, belief that access to such services should be free, availability of alternatives, and low usage of Europe PMC. Several respondents stated that ideally they would like the resource to remain free. The mean amount users were willing to pay per year was $28.5, which was calculated using the midpoint of the user's chosen range or exact amount if the user was willing to pay more than $500.

The vast majority of librarians believed that their institutions would be willing to pay $250 or less (the lowest range) towards an annual subscription for Europe PMC (Figure 10). When asked to justify their choice, they most commonly cited low usage and a lack of financial resources (low income/limited budget) within their institution. Several respondents stated that they would prefer not to pay and that they did not see any added value of Europe PMC beyond what is already available through PubMed and PMC. No patterns were observed in librarians' choices based on country and institution type. However, librarians from institutions with 1,000 or fewer life science users were more likely to pick the lowest range.

The mean WTP amount calculated using the WTP amount per user per annum for each institution (midpoint of the range chosen divided by the number of life science users in that institution) is $1.1 per (life science) user per year.

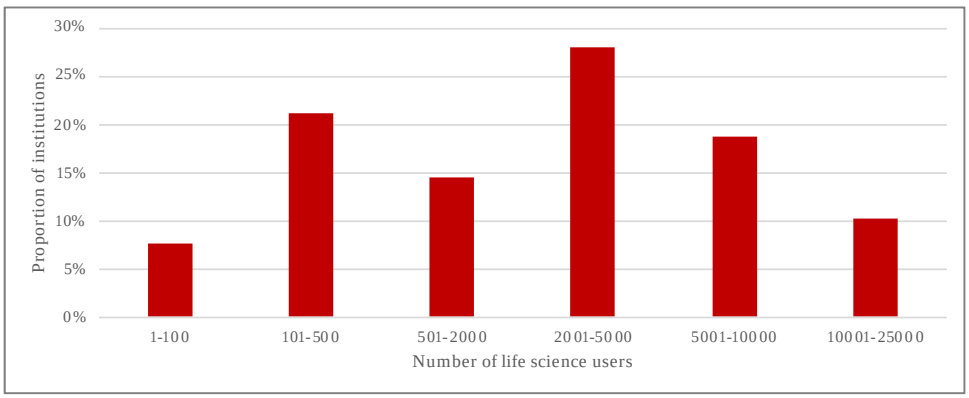
Figure 10 Librarians' responses on their institution's willingness-to-pay amounts (a) and number of life sciences users (b) [n=118]

(a)



Reasons for choice (across 58 responses)	No. of responses
Low usage/no added value	21
Low income/limited budget	12
Other	11
Unfamiliar with service	7
Other alternative available	6
Believe access should be free	5

(b)



Source: Technopolis analysis of librarian survey

The WTP amounts differ considerably between users and librarians. It is important to remember that providing WTPs for complex or unfamiliar tasks can be challenging and may lead to inaccurate results.[23] Furthermore, the perceived valuation of a product is not necessarily stable, with buyers often misjudging the price of something they are currently getting free. Moreover, WTP estimates are constrained by respondents' considerations of their ability to pay (i.e. disposable funds)[24] or availability of current alternatives (e.g. databases and services subscribed to by institutions) and thus responses should be interpreted with caution.

Multiplying the mean WTPs obtained from the user survey and the librarian survey with the mean number of users (unique IPs) in a year, we get a contingent valuation of Europe PMC of $286 million and $11.3 million respectively. Comparing these values to the cost per year of running Europe PMC ($3.3 million), we gather that for every dollar spent on Europe PMC, the value gained is $87 or $3 respectively based on the amounts users and librarians are willing to pay.

It is important to note that contingent valuation is a well-established approach to obtain value for non-market resources. Nevertheless, it has its own shortcomings and one in particular that it relies on stated preferences of survey respondents rather than revealed preferences in binding transactions. Since we inherently need to deal with hypothetical scenarios in this study, there is a risk that survey respondents do not fully understand the context of the question or indeed their answers suffer from cultural-strategic biases.

For every dollar spent on Europe PMC, the value returned is 87 or 3 dollars based on mean willingness to pay quoted by users and librarians respectively[25].

6.3 Other considerations of impact

Time saved by using Europe PMC

While most of the Europe PMC services can be replicated elsewhere, all of the Europe PMC services are not available through a single alternative service. Thus, Europe PMC may enable users to save time by providing multiple services through the same interface. The subject of time savings as a result of using Europe PMC was broached in the interviews. Almost all the interviewees could not estimate any time savings as they had either never used another comparator service for the same task or could not identify another alternative service/platform to perform the same task.

One interviewee from an alternative service mentioned that they had saved around 10 hours of development time using Europe PMC in a particular task. Another interviewee, a third-party developer, stated that Europe PMC's searching function is around 4X faster than that of PMC. However, these statements do not provide any wider information or context that would allow further measurement of time saved by using Europe PMC.

Value for Europe PMC funders

The economic value of Europe PMC for funders was underexplored in this impact review. Based on initial scoping interviews, it became clear that use of Europe PMC by funders is variable, depending not only on the funder but the individual user. Funders only use Europe PMC if they are familiar with its functionalities or they have established a workstream that specifically uses Europe PMC. As such, it was difficult to find a comparator to compare time saved by funders through using Europe PMC rather than another service.

Nonetheless, Europe PMC provides value to its funders through its repository function. In the absence of Europe PMC, they will have to make alternate arrangements to meet their open access commitments

[23] Breidet, C., Hahsler, M., and Reutterer, T. (2006) *A Review of Methods for Measuring Willingness-to-Pay* in 'Innovative Marketing, 2006.

[24] Beagrie, N. and Houghton, J. (2016) *The value and impact of the European Bioinformatics Institute*

[25] This is total contingent valuation per year ($286 million and $11.3 million respectively for users and librarians) divided by the total cost per year in PPP $ to run Europe PMC as shown in Table 5. As this is a ratio, the units cancel each other out.

and this will incur additional costs. Moreover, there are potential savings in having a common, searchable repository where financial contributions are proportional to each funder's research spending, as opposed to creating and maintaining several individual repositories. In addition, there is a non-monetisable value in having a Europe-based platform that is independent of changes in US funding of PMC.

In general, Europe PMC offers functionalities, for example, links to citations, altmetrics and grant data that could potentially be exploited further by funders and would complement the functionalities of other services such as ResearchFish.

7 What would happen if Europe PMC no longer existed?

The majority of interviewees felt that if Europe PMC ceased to exist, it would be a loss to its users and could have a destabilising effect on their workflow. For researchers, students and citizen scientists, including those from developing countries, losing Europe PMC could significantly reduce the volume of research they are able to access and consequently able to conduct. However, this group of interviewees also felt that although Europe PMC was an important resource for them, in the absence of Europe PMC they would use other similar services. A healthcare professional noted that they used PubMed more frequently than Europe PMC, and therefore would not be significantly impacted by loss of access to the service.

A high proportion of the industry users and third-party developers that were interviewed expressed that their work and data pipelines would be more difficult and time consuming to both develop and maintain without Europe PMC. Some of these users highlighted that they may not be able to continue their current work because of the cost and time needed for redevelopment using an alternative data source. In a number of cases, developers have used Europe PMC to source their data and are thus completely reliant on it for their applications to function. In the absence of Europe PMC, they would have to become accustomed to using APIs from other providers and adapt their workflows accordingly, with the element of familiarity and comfort associated with using Europe PMC being lost.

An interviewee from Open Targets indicated that without Europe PMC their established data mining pipeline would be lost. Similarly, an interviewee from ELIXIR noted that they would essentially need to commission PMC to provide a text mining service to sustain their pipeline if Europe PMC ceased to exist. That being said, if Europe PMC and PMC continue to provide the same content, the impact of losing Europe PMC would be diminished. However, these interviewees noted that Europe PMC is more convenient and quicker to use than both PMC and PubMed.

An interviewee from ORCID expressed that the loss of Europe PMC would create a 'noticeable gap' that would likely be occupied by commercial entities. In contrast to these entities, Europe PMC strives to keep information open, free and within the hands of the research community, thereby fulfilling an important need in terms of Open Science, especially for European research funders and researchers. Within ORCID, approximately 1.9 million of a total of 36 million works are attached to Europe PMC as a source, making Europe PMC the third largest data contributor to ORCID and the largest 'disciplinary' contributor[26]. The same interviewee remarked that the impact they are able to have on medical researchers would be diminished without ORCID's links to Europe PMC, and without it, openness in the life sciences would be 'palpably diminished'.

A librarian noted that without Europe PMC, their affiliates would have to use alternative repositories to meet their open access obligations, such as PMC or Zenodo. They also expressed that the loss of Europe PMC would be a significant loss to the Open Access movement more broadly. Furthermore, it was noted that both PMC and Europe PMC are fundamental to providing open access to publicly funded research in both the US and Europe, and that since this type of research is funded by taxpayers, it should be open access. Keeping this in mind, it was stated that the community needs to build on the 'outstanding work' of Europe PMC in making biomedical content available more widely.

Several interviewees noted that unlike PubMed/PMC, Europe PMC is not directly tied to the government. This allows Europe PMC to offer greater stability to the research community. Moreover, Europe PMC is nimble as evidenced by its ability to introduce new services such as indexed preprints relatively rapidly compared to PMC. Previous government shutdowns in the US encouraged interviewees to emphasise the need for a service that replicates PMC content and ensures that there is no disruption to access. Such uncertainties in the continued availability of PMC bolsters the argument in favour of an autonomous repository that is independent of political or funding decisions in the US[27].

Losing Europe PMC would have a destabilising effect on users that build applications from Europe PMC data and the open access movement more broadly. Researchers would be less affected. In the absence of Europe PMC, all users will try to rely on alternatives, but this could result in additional resource costs. Moreover, the loss of

[26] Data from ORCID

[27] Interview

Europe PMC will leave users more susceptible to changes in the priorities of other providers, particularly the US government.

8 Conclusions

Europe PMC has its roots in UK PubMed Central (UKPMC), a free archive and repository for biomedical and life science journal articles. In November 2012, UKPMC was re-named Europe PMC and now has 29 research funders from across the world. It not only contains all PMC content but also provides four broader services: (1) a repository for research outputs, (2) a website to find relevant data and articles, (3) infrastructure to interrogate individual resources, and (4) facilitation of text and data mining as well as ORCID linking and citation tracking. Europe PMC is currently run under grant by EMBL-EBI with a funding committee in place to make key management decisions.

The number of Europe PMC users has been steadily rising over the past five years, with nearly 12 million unique IPs accessing the service in 2018. The US, China, India and the UK collectively accounted for almost half of these unique IPs showing that Europe PMC has a global reach. Our website survey found that academic researchers, students and health professionals comprise over three-quarters (76%) of the user base.

Users are most likely to use Europe PMC in order to find research articles although it is also an important resource for researcher funders looking to see whether publications are meeting open access requirements, and for API users who rely on it for text and data mining. Overall, users particularly appreciate features such as Europe PMC's ease of use, good coverage, the APIs and free access to full-text publications. Nevertheless, several users we spoke with felt that Europe PMC is not sufficiently well-publicised and that not enough people are aware of the potential benefits of using it.

Currently, there are no other platforms that provide exactly the same range of services as Europe PMC. However, alternative platforms collectively do cover most of Europe PMC's services and functionalities. Indeed, for many users, Europe PMC is one of several different tools to find biomedical content. PubMed, PMC, search engines and Google Scholar are also commonly used. However, many view Europe PMC as being better than its alternatives in terms of its open access provision, having a more user-friendly interface, and better functionalities. Nevertheless, usage of Europe PMC remains especially low in comparison to PMC.

Our analysis has also shown that the average annual usage value per Europe PMC user ranges from $31 to $6 per year respectively when a deadweight of 75% or 95% is applied. This assumes that Europe PMC functionalities can be replicated through alternatives for a majority (75% deadweight) or almost all (95% deadweight) of its uses. Considering the running cost is about $3.3 million per year, the usage value returned is $116 or $22 per dollar spent when a deadweight of 75% or 95% respectively is applied.

We also attempted to capture what Europe PMC was worth to its users. We found that users are willing to pay $28 and librarians are willing to pay $1 per user per year. Considering these amounts, the value returned is $87 or $3 for every dollar spent on Europe PMC.

Both these valuations show that Europe PMC represents a very high value to users relative to the investment for funders to run the services. For instance, a valuation of EMBL-EBI data and services using the same two methods demonstrated a value to users equivalent to around 6 times the direct operational cost[28]. That is, for every dollar spent, the value returned was $6. While the contingent valuation of EMBL-EBI data and services in that study was about twice that of Europe PMC (as calculated here based on users' willingness to pay data), the annual operational cost of EMBL-EBI was 24-fold higher than that of Europe PMC.[29] These numbers indicate that the high multiples of Europe PMC valuation are the result of a combination of the high value attributed to Europe PMC services by the global user community with the relatively low cost of running services that are embedded in an established research infrastructure.

[28] Beagrie, N. and Houghton, J. (2016) *The value and impact of the European Bioinformatics Institute*

[29] Contingent valuation based on users' data gave a value of $492 million for EMBL-EBI in 2015, while the annual operational cost was $72 million, using OECD's PPP exchange rate of 0.655 in 2015.

It is however important to stress that the contingent valuation (based on willingness to pay) and usage valuation (based on self-reported usage times) of Europe PMC as a knowledge platform cannot be interpreted as a genuine market price. The approach used may suffer from cultural-strategic biases in stated preferences and thus may lead to overestimation of the value Europe PMC provides.

Europe PMC provides value to its funders through its repository function, enabling them to meet their open access commitments and offering potential savings through having a common, searchable repository funded by several funders. Moreover, it offers potential time savings to users through provision of an integrated platform that combines multiple services available through different providers.

Many of those spoken to over the course of the study have indicated that if Europe PMC ceased to exist, it would cause noticeable problems to several user groups. Poorly resourced researchers (e.g. those from developing countries, citizen scientists, and students) would have a significantly reduced volume of publications that they could access. Some third-party developers spoke of the centrality of Europe PMC to their solutions, meaning that if Europe PMC were to disappear these providers would have to completely rethink and rebuild their solutions, entailing additional time and financial costs. Similarly, current funders would incur costs to replace Europe PMC's repository function. It was also noted that Europe PMC is vital to the Open Access movement and offers a degree of security against changes in the US political climate, which might affect availability of PMC services.

To conclude, Europe PMC presents excellent value for money in terms of the value returned compared to the relatively low costs of running it. While Europe PMC is not completely unique, it offers an integrated, free platform to users. If it were to become unavailable, users who have developed applications using Europe PMC's API and funders who use the repository function would incur additional costs in terms of accessing and learning to use alternative services as well as establishing new workflows. Moreover, Europe PMC contributes greatly to the Open Science agenda and could play an important role in the context of Plan S in Europe. Finally, the protection it offers against changes in the business models and funding decisions of commercial providers and the US government is invaluable as is its nimbleness, which allows rapid introduction of new functionalities such as indexed preprints in response to emerging needs.

Appendix A Approach and Methodology

Our approach was designed to analyse and measure the value of Europe PMC[30] to its funders, the global scientific community and other users, and to deliver the evidence base required to inform strategy decisions for funding of the repository beyond 2021. Its various components are highlighted below.

A.1 Analytical framework

Our finalised methodological approach and data sources to answer the evaluation questions are outlined in the table below. This framework was used to develop the consultation tools (questionnaires for interviews and surveys). Findings from the inception meeting, scoping interviews and some initial desk research fed into defining this methodological approach.

Table 9 Finalised Methodological Approach

Evaluation Question	Methodology	Secondary data sources
What is the value and impact of Europe PMC to funders who support the service and the research community?		
What is the value of the Europe PMC repository (using time as a measure of value) to the research community?	User and Librarian Surveys Secondary Data Analysis Economic Valuation	UK Times Higher Education Salary Survey – academic researcher salary rates
What efficiency savings does Europe PMC provide? What is the value of these efficiencies and who benefits from them?	User and Librarian Surveys Stakeholder Interviews Economic Valuation Case studies Desk research Secondary Data Analysis	UK Times Higher Education Salary Survey Annual Survey of Hours and Earnings (ASHE) – general salary rates World Bank global wage data The Association of Commonwealth Universities wage data
What is the contingent valuation of providing the Europe PMC service and how does this compare to the actual operational costs?	Librarian Survey Stakeholder Interviews Secondary Data Analysis Desk Research	Europe PMC operational cost data
How does the provision and usage of Europe PMC (and its constituent services) compare with comparator services such as PMC and/or OpenAire?	Desk Research User and Librarian Surveys Stakeholder Interviews Secondary Data Analysis Case studies	Monitoring data from Europe PMC and comparator services (e.g. unique IP addresses) Peer-reviewed publications
What would be the impact on the research community if Europe PMC did not exist?	User Survey Stakeholder Interviews Economic Valuation Case studies	

A.2 Data collection

A.2.1 Desk research

Our desk research comprised mainly document and literature review (peer-reviewed and grey literature) and secondary data analysis. We reviewed existing impact evaluations of research infrastructures (but we could identify none for infrastructures similar to Europe PMC), publicly available wage data, Europe PMC website access statistics, previous user survey results, and Europe PMC internal operational and

[30] Levchenko, M et al, Europe PMC in 2017, Nucleic Acids Research, 2018, 46. DOI: 10.1093/nar/gkx1005

financial reports such as funders' business cases, funders' committee meetings and the Europe PMC Principal Investigator's reports. Our findings from the document and literature review have been incorporated into the background section.

In relation to website statistics for Europe PMC and comparator services, we obtained monitoring statistics collected by the Europe PMC team at EMBL-EBI as well as PMC[31]. We also obtained comparative usage statistics for both Europe PMC and PMC for two publications – Wellcome Open Research and Proceedings of the National Academy of Sciences (PNAS) – from the Wellcome Trust and PMC respectively.

A.2.2 Surveys

To capture the impact and value of Europe PMC, we surveyed two main stakeholder groups: users and research librarians using the SurveyMonkey platform. The surveys provided two important types of information: (i) quantitative data for the evaluation of the value of Europe PMC and (ii) qualitative data on views, experiences and perceptions of Europe PMC that were further explored through interviews and case studies.

The core questions for all surveys had some level of commonality, to allow comparison and aggregation across constituencies, even if the invitations, explanatory text and target populations differed in some degree. The surveys were anonymous with the option for respondents to provide their name and contact information for possible follow up interviews. The surveys were initially piloted with 5-10 individuals from the different stakeholder groups to help us refine the questionnaire.

User survey

The actual and potential user population was surveyed through two distinct routes: (i) survey questionnaires emailed to authors of publications stored in Europe PMC ("email-based researcher survey") and (ii) a pop-up survey on the Europe PMC website ("website survey"). Considering that the targeted survey would capture both users and non-users from the research field, while the pop-up survey would capture all types of users including software and app developers, funders and the general public, two separate surveys were developed (see Appendix B and Appendix C).

For the email-based researcher survey, we extracted authors' email addresses where available from Europe PMC for publications published between January and June 2018. Only one email address per publication was included in our sample to avoid surveying too many people from the same institution. Based on the domain name, email addresses were split into groups based on region or country following which surveys were sent out with approximately the same geographical distribution as the Europe PMC website access statistics from 2017. Full breakdown is shown in Table 10 below.

Table 10 Geographical distribution of email addresses contacted for user survey

Group	Relevant domains	Proportion	Total number
China	.cn	29%	4350
US	.edu, .mil, .gov	18%	2700
UK	.uk	6%	900
India	.in	4%	600
Canada	.ca	3%	375
Australia	.au	3%	375
Europe	Relevant country domains minus .uk	6%	900
Asia and Oceania	Relevant country domains minus .cn, .in, .au	6%	900

[31] Data shared were high-level, anonymised summary data and in line with GDPR requirements

Africa	Relevant country domains	6%	900
Latin America	Relevant country domains	6%	900
Generic	.com, .int, .org, .net	14%	2100

In total, the survey was delivered to 15,000 email addresses in three batches. Non-respondents or partial respondents were sent two reminders 7 and 14 days after the first email.

The pop-up survey on Europe PMC's website was implemented with assistance from the Europe PMC Principal Investigator and her team who also posted a link to the survey on the API users community group.

Librarian survey

As research librarians often act as intermediaries on behalf of research users, we expected them to be in a more informed position to provide the study with willingness-to-pay estimates (how much users would be willing to pay for Europe PMC services were they not free) i.e. a monetary value for Europe PMC services. Hence, this population was surveyed separately. The survey questionnaire for librarians is provided in Appendix D. We approached relevant survey respondents through the European Association for Health Information and Libraries (EAHIL); Society of College, National and University Libraries (SCONUL); Scholarly Publishing and Academic Resources Coalition (SPARC); Research Libraries UK (RLUK) and Council of Australian University Librarians (CAUL). We also contacted other national and regional umbrella organisations across the globe, however, we did not get a positive response. The link to the survey was also distributed on the EAHIL and Europe PMC twitter feeds.

A.2.3 Interviews

Scoping interviews

We conducted five scoping interviews with six key individuals involved in funding and delivering Europe PMC. Four of the interviewees were Europe PMC funders and two were members of the Europe PMC team at EMBL-EBI. Among the funders, we interviewed one large charity (Wellcome Trust), one small charity (Marie Curie), one UK public funder (MRC) and one non-UK funder (Austrian Science Fund, FWF).

Topics covered in the scoping interviews included Europe PMC's purpose and remit, its services and how they are used, its benefits and impacts, governance and delivery, and how it compares to other similar services.

User and stakeholder interviews

We conducted a programme of 25 semi-structured interviews to explore the views of a range of stakeholders to fill gaps in the data gathered by desk research and user surveys, and to deepen our understanding (the 'why' and 'how'). The interviewees were identified from the surveys (respondents who have agreed to be interviewed), scoping interviews or desk research. Where possible, we tried to balance the number of interviewees by gender and location (top five countries, global North and South). The interviews were conducted by telephone or Skype where possible (written responses were also accepted). A list of interviewees and interview questions are presented below (see Table 11 and Table 12).

Table 11 List of interviewees

Stakeholder Group	Subtype	Organisation and/or country
Beneficiaries	Researchers	• Germany • India

	Clinicians	• France • Netherlands
	Industry users	• Open Targets
	Students	• US • China
	Third-party developers/API users	• University of Cambridge, UK • Science and Technology Facilities Council, UK • Leiden University Medical Center, Netherlands
	Policy makers/ government officials	• Health Department, Bahrain
	Members of the general public	• Australia
Librarians	Academic librarians	• University of Melbourne, Australia • Vienna BioCenter, Austria
	Healthcare research librarians	• Brighton and Sussex Medical School, UK
Alternative infrastructures / services		• OpenAire • ORCID • Unpaywall
Research funders (not currently funding Europe PMC)		• Canadian Institutes of Health Research (CIHR) • Arcadia Fund, UK
Other stakeholders		• ELIXIR • PLoS • PeerJ • CrossRef
Non-users	Academic Researcher	• Brazil

Table 12 User and stakeholder interview questionnaire

	Introduction
1	Please say a little bit about your organisation and your role within it.
2	[For other funders only] Have you ever funded or thought of funding a repository like Europe PMC? Please explain why. - What current platform do you use to fulfil your commitments for open access?
	Europe PMC services and how they are used
3	How familiar are you with Europe PMC? What do you know about it?
4	Please comment on the range and quality of services provided by Europe PMC. - What are the unique features? - Does it meet your needs? Are there any gaps?
5	How do you / your organisation use Europe PMC? - What specific services do you use?

	- Why use Europe PMC rather than another provider? - Can you provide an estimate of time spent using Europe PMC or time saved? For example, how many minutes do you spend on average per visit or how much time do you save compared to another service per visit? - We are looking to develop case studies to showcase how Europe PMC is used by different stakeholders. In that context, can you provide any interesting examples of use?
6	[for non-users only] Why do you not use Europe PMC? ◦ What would make you start using it?
	Value of Europe PMC
7	Do you / your organisation use services from other similar providers? If so, what specific services do you use and why? - Can you provide an estimate of time spent using these services or time saved? For example, how many minutes do you spend on average per visit or how much time do you save compared to another service per visit?
8	What value/benefits does Europe PMC deliver compared to other similar services? - How do Europe PMC services and functions compare with that of other similar providers? - What are the pros and cons of Europe PMC versus other similar providers?
9	We are looking to develop case studies to showcase the impact of Europe PMC. To that end, can you describe any interesting cases of impact?
10	[For survey respondents only] In the survey, you indicated that Europe PMC was worth between xxx and xxx $per year [use value from survey] to you / your organisation. Can you please explain the reasons for your choice in more detail? - Your choice was at the lower / middle / higher [choose as appropriate] end of the suggested ranges. Is there a specific reason for this choice?
11	What do you think would happen if Europe PMC did not exist?
	Further comments
12	Would you like to make any additional comments?
13	[For comparator services only] Are there any data or documents that you would like to share with us for the evaluation? For example, business case, monitoring report/data, evaluation reports, etc.

A.2.4 Impact case studies

We identified key impacts for various stakeholders through the survey, interviews and client feedback. Based on this information, we created a longlist of eight most interesting examples from which three examples were shortlisted by the client for developing in-depth impact case studies. Examples of novel use of Europe PMC by a developer/API user, unique functionalities exploited by a similar service and use by a researcher in a developing country were chosen for case study development. These impact case studies provide contextual information and illustrate some of the highest impacts of Europe PMC to accompany the quantitative and economic data analysis.

A.3 Analysis

Once data were collected, we analysed the qualitative and quantitative data separately, before triangulating the findings to derive our final conclusions and prepare the draft final report.

A.3.1 Economic Valuation

We estimated the economic value of Europe PMC using proxy measures

1. Monetised value of time spent using Europe PMC services – usage value modelling
2. Value users put on Europe PMC services – contingent valuation

Both methods are described in detail in the next sections.

Usage Value Modelling

One of Europe PMC's main impacts is enabling access to and usage of a range of resources. We placed a monetary value on this usage, using a methodology with three main parts

- Dividing all Europe PMC users into different user groups/categories
- Calculating the effect of each group's Europe PMC usage
- Summing the effect for each user group to reach a combined value for all Europe PMC users

Below, we outline the steps as well as the specific information and assumptions that will feed into the usage valuation. As described in Table 13, the first elements centre on calculating the effects of Europe PMC usage.

Table 13 Calculating the gross effect of Europe PMC usage

Step	Component	Source	Description
1	Number of total users (the user list)	List of unique IP addresses – monitoring data (from EMBL-EBI)	We used the list of unique IP addresses that accessed the Europe PMC website in 2018 as a proxy for the total number of Europe PMC users in a typical year. IP addresses do not give a completely accurate estimate of users. For instance, several researchers may use the same IP address while a single user may appear as multiple IP addresses if they tend to work across different sites. Automated and robot activities can also account for some unique IP addresses. We assumed that these respective under-estimates and over-estimates will largely cancel each other out, making unique IP addresses a reasonable proxy for total Europe PMC user numbers per year. We segmented the users on the basis of whether they were from developed, developing or transition economies. [32]
2	Segmenting users into different user groups (number of users per user group)	Website user survey	We used the results of the website user survey to further segment our user list from Step 1 into different user groups • Research funders • Academic researchers • Non-academic researchers • Industry • Students • Clinicians and health professionals • General public
3	Time per visit per user group	User surveys	Through both the user surveys, we determined the median amount of time each user segment (accounting for different economies and user groups) spent on Europe PMC per visit.
4	Frequency of use per user group	User surveys	Both the user surveys enabled us to calculate the average Europe PMC visits a month for each user segment.
5	Usage value per group	Secondary data (International Labour Organization [ILO])	The different user groups generate a different economic value from their Europe PMC usage. At an individual user level, the economic value of student usage of Europe PMC will be lower than commercial users for example. An additional consideration was the difference in wages across the world. We used wage data available from the ILO for different countries across different economic activities (ISIC Rev. 4 codes[33]) to calculate a median wage per hour for the different economic activities for developed, developing and transition economies. The different economic activities were chosen based on their relevance to the chosen user groups.
6	Total value for each user group	Technopolis calculation	Using the results from step 2-5, we will reach a gross economic value for Europe PMC usage by each respective

[32] United Nations analysis shows which countries fall into each of these categories. See http://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf

[33] https://unstats.un.org/unsd/publication/seriesM/seriesm_4rev4e.pdf

Step	Component	Source	Description
			user group. This means multiplying the total number of users for each group in Step 2 with the figures obtained from Steps 3-5.
7	Total value	Technopolis calculation	By summing the economic value for each user group, we will reach a final figure that establishes the typical total economic value of Europe PMC usage annually.

Contingent valuation

Contingent valuation is a way of estimating the value that a person places on a good or service. It involves asking people directly about their willingness to pay (WTP) to obtain a specified service. In the case of Europe PMC, we can place a contingent valuation by calculating the mean WTP amongst survey respondents.

One option for calculating this mean WTP would be to directly ask Europe PMC users how much they would be prepared to pay for the service if it was not free. However, as claimed by Breidet et al (2006), these types of customer surveys are an inherently inaccurate way of determining WTP. For instance, providing WTPs for complex or unfamiliar tasks can be challenging and lead to inaccurate results.[34] Furthermore, the perceived valuation of a product is not necessarily stable, with buyers often misjudging the price of something they do not use frequently. As such, they advocate asking experts to assess the WTP of customers as they are able to make educated guesses rather than potentially random ones from customers.

In the context of this study, we believed that librarians would be the relevant 'expert group'. We expected them to provide a more realistic WTP figure owing to their familiarity with subscribing to other databases and services. Therefore, the librarians survey was our primary tool to determine a WTP amount. However, for comparison, we also included a WTP question in the user surveys and have calculated mean WTP for this group as well.

Other considerations of impact

Although they did not form core components of our economic model, we also sought to identify impact generated through the following routes in qualitative terms through the interviews and case studies

- Commercial benefits of using Europe PMC, for instance, where Europe PMC has helped enable the development of saleable products
- Time savings achieved by using Europe PMC

Reflections on additionality and the counterfactual

Best practice guidance as set out in HM Treasury's *Green Book* and the Homes and Communities Agency's (HCA) *Additionality Guide* highlights the need to assess the net additional impact of interventions and measures by examining the various components of additionality such as

- Deadweight: the extent to which changes and benefits could have occurred even in the absence of the intervention
- Displacement: the extent to which Europe PMC's positive effects have been offset by disbenefits to its funders' other activities
- Leakage: the extent to which benefits of Europe PMC are realised outside the target area

[34] Breidet, C., Hahsler, M., and Reutterer, T. (2006) *A Review of Methods for Measuring Willingness-to-Pay* in 'Innovative Marketing, 2006.

We have chosen to disregard displacement as none of the funders operate a service themselves that is comparable to Europe PMC. Leakage is not relevant here either as Europe PMC does not place any geographical restrictions on who can use the service.

We estimated deadweight by determining the extent to which Europe PMC users could have achieved similar results through other means (e.g. by using an alternative service). We based this on qualitative feedback on the deadweight associated with Europe PMC usage provided through the surveys and interviews. The HCA's *Additionality Guide* indicates that typical deadweight levels are 24%,[35] and the survey and interviews helped us determine the extent to which it was appropriate to deviate from this average. We drew on Scottish Enterprise's ready reckoner below, choosing deadweight figures whose descriptions most accurately described the feedback provided.

Table 14 Deadweight ready reckoner

Level	Description	Deadweight
None	All of the benefits are as a result of the intervention	0%
Low	The majority of the benefits are as a result of the intervention	25%
Medium	About half of the benefits area as a result of the intervention	50%
High	A high level of the outputs/outcomes are not as a result of the intervention	75%
Total deadweight	None of the outputs/outcomes are as a result of the intervention	100%

Source: Taken verbatim from Scottish Enterprise (2008)[36]

[35] Homes and Communities Agency (2014), *Additionality Guide*, p. 21.

[36] Scottish Enterprise (2008) *Additionality & Economic Impact Assessment Guidance Note: A Summary Guide to Assessing the Additional Benefit, or Additionality, of an Economic Development Project or Programme.*

Appendix B Email-based researcher survey questionnaire

The Wellcome Trust and other Europe PMC funders have appointed the independent research organisation, Technopolis, to carry out a study examining the value and impact of Europe PMC. We would like to understand who uses Europe PMC and how they do so. To that end, we would be grateful if you could complete a 5-minute questionnaire using the link below.

Responses can be provided anonymously. Full details on how the study team will use this information are available at http://www.technopolis-group.com/privacy-policy/.

Please note that questions with an asterisk (*) are important for this study and hence require an answer.

Page 1

About you

Question	Options
1) *In which country are you currently based?	[Dropdown menu]
2) *Which of these best describes you?	[Radio buttons] • Research funder • Academic researcher • Researcher in the private sector (i.e. R&D) • Private sector employee (not researcher) • Third party developer / API user • Clinician / health professional • Government laboratory researcher • Government official (e.g. public health official, civil servant) • Undergraduate / post-graduate student • Other (please specify)

Page 2

How you use Europe PMC

Question	Options
3) *What do you use Europe PMC for? (select all that apply)	[Multiple choice list] • Finding research articles • Claiming articles to an ORCID • Uploading manuscripts • Getting citation statistics for a bibliography/reference list • Finding funder/grant information • Accessing text mining services • Developing apps and software • Using the annotations service (SciLite) • Accessing patents • Accessing Agricola records • Accessing clinical guidelines • Linking Europe PMC articles to other information (e.g. blogs, websites, data, database resources) • Other (please specify)

<table>
<tr><th>Question</th><th>Options</th></tr>
<tr><td></td><td>• I do not use Europe PMC</td></tr>
<tr><td>4) *How often do you use the following services?</td><td>[radio buttons – select one option per row]
<table>
<tr><th></th><th>Regularly</th><th>Occasionally</th><th>Not at all</th><th>Not aware of this resource</th></tr>
<tr><td>PubMed</td><td></td><td></td><td></td><td></td></tr>
<tr><td>PubMed Central</td><td></td><td></td><td></td><td></td></tr>
<tr><td>OpenAire</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Scopus</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Web of Science</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Google Scholar</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Gateway to Research</td><td></td><td></td><td></td><td></td></tr>
<tr><td>ResearchGate</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Dimensions</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Google (including other web search engines)</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Other resource (please specify)</td><td></td><td></td><td></td><td></td></tr>
</table></td></tr>
</table>

Page 3

How much you use Europe PMC

Question	Options
5) *How many times on average do you visit Europe PMC in a month?	[radio buttons] • Less than once a month • 1-5 times • 6-10 times • 11-20 times • 21-30 times • More than 30 times
6) *How many minutes does a typical visit last? Please enter a number.	[single line text] minutes

Page 4

Your experiences of using Europe PMC

Question	Options
7) How helpful do you find Europe PMC for whichever purpose you use it for?	[Radio buttons] • Not at all helpful • Somewhat helpful • Helpful • Very helpful • Extremely helpful • Can't say
8) Please feel free to provide any further details on your experiences of using Europe PMC below (ease of use, how it has impacted your work, features you particularly like, any aspects you would like improved).	[Open text box]

Page 5

In order to help assess the cost-effectiveness of Europe PMC, we would like respondents to estimate what they would be willing to pay for Europe PMC if its services were not free. Please note this question is intended to determine the value of Europe PMC and not to explore a move to charged services. (Costs are indicated in US Dollars; guiding exchange rates can be found at https://www.oanda.com/currency/average).

Question	Options
9) If you were asked to pay an annual personal subscription for Europe PMC, roughly how much would you be willing to pay? Please note this question does not signal any intention to move to charged services.	[Radio buttons] • Up to $10 per year • $11-50 per year • $51-100 per year • $101-250 per year • $251-500 per year • more than $500 per year, please specify
10) Please feel free to explain your choice above.	[Open text box]

Page 6

Thank you for your response. We really appreciate your input so far. If you are willing to participate in a max. 15-minute follow-up interview (by telephone or Skype) with the study team to discuss your experience of Europe PMC and/or similar services further, please provide your contact details below.

Please be assured that your contact details will not be shared outside the study team and will be deleted on completion of the study. Full details on how the study team will handle your data are available at http://www.technopolis-group.com/privacy-policy/

11) Your contact details

Name	[Single line text]
Email address	[Single line text]
Comments	[Single line text]

Appendix C Website user survey

The Wellcome Trust and other Europe PMC funders have appointed the independent research organisation, Technopolis, to carry out a study examining the value and impact of Europe PMC. We would like to understand who uses Europe PMC and how they do so. To that end, we would be grateful if you could complete a 5-minute questionnaire using the link below.

Responses can be provided anonymously. Full details on how the study team will use this information are available at http://www.technopolis-group.com/privacy-policy/.

Please note that questions with an asterisk (*) are important for this study and hence require an answer.

Page 1

About you

Question	Options
1) *In which country are you currently based?	[Dropdown menu]
2) *Which of these best describes you?	[Radio buttons] • Research funder • Academic researcher • Researcher in the private sector (i.e. R&D) • Private sector employee (not researcher) • Third party developer / API user • Clinician / health professional • Government laboratory researcher • Government official (e.g. public health official, civil servant) • Undergraduate / post-graduate student • Member of general public • Other (please specify)

Page 2

How much you use Europe PMC

Question	Options
3) *How many times on average do you visit Europe PMC in a month?	[radio buttons] • This is my first ever visit. • Less than once a month • 1-5 times • 6-10 times • 11-20 times • 21-30 times • More than 30 times
4) *How many minutes does a typical visit last? Please enter a number. *If you are a first-time user, please enter the duration of this session.*	[single line text] minutes

Page 3

How you use Europe PMC

<table>
<tr><th>Question</th><th>Options</th></tr>
<tr><td>5) *What do you use / intend to use Europe PMC for? (select all that apply)</td><td>[Multiple choice list]
Finding research articles
Claiming articles to an ORCID
Uploading manuscripts
Getting citation statistics for a bibliography/reference list
Finding funder/grant information
Accessing text mining services
Developing apps and software
Using the annotations service (SciLite)
Accessing patents
Accessing Agricola records
Accessing clinical guidelines
Linking Europe PMC articles to other information (e.g. blogs, websites, data, database resources)
Other (please specify)</td></tr>
<tr><td>6) *How often do you use the following services?</td><td>[Radio buttons – select one option per row]<table><tr><th></th><th>Regularly</th><th>Occasionally</th><th>Not at all</th><th>Not aware of this resource</th></tr><tr><td>PubMed</td><td></td><td></td><td></td><td></td></tr><tr><td>PubMed Central</td><td></td><td></td><td></td><td></td></tr><tr><td>OpenAire</td><td></td><td></td><td></td><td></td></tr><tr><td>Scopus</td><td></td><td></td><td></td><td></td></tr><tr><td>Web of Science</td><td></td><td></td><td></td><td></td></tr><tr><td>Google Scholar</td><td></td><td></td><td></td><td></td></tr><tr><td>Gateway to Research</td><td></td><td></td><td></td><td></td></tr><tr><td>ResearchGate</td><td></td><td></td><td></td><td></td></tr><tr><td>Dimensions</td><td></td><td></td><td></td><td></td></tr><tr><td>Google (including other web search engines)</td><td></td><td></td><td></td><td></td></tr><tr><td>Other resource (please specify)</td><td></td><td></td><td></td><td></td></tr></table></td></tr>
</table>

Page 4

Your experiences of using Europe PMC

Question	Options
7) How helpful do you find Europe PMC for whichever purpose you use it for?	[Radio buttons]

Question	Options
	• Not at all helpful • Somewhat helpful • Helpful • Very helpful • Extremely helpful • Can't say
8) Please feel free to provide any further details on your experiences of using Europe PMC below (ease of use, how it has impacted your work, features you particularly like, any aspects you would like improved).	[Open text box]

Page 5

In order to help assess the cost-effectiveness of Europe PMC, we would like respondents to estimate what they would be willing to pay for Europe PMC if its services were not free. Please note this question is intended to determine the value of Europe PMC and not to explore a move to charged services. (Costs are indicated in US Dollars; guiding exchange rates can be found at https://www.oanda.com/currency/average).

Question	Options
9) If you were asked to pay an annual personal subscription for Europe PMC, roughly how much would you be willing to pay? Please note this question does not signal any intention to move to charged services.	[Radio buttons] • Up to $10 per year • $11-50 per year • $51-100 per year • $101-250 per year • $251-500 per year • more than $500 per year, please specify
10) Please feel free to explain your choice above.	[Open text box]

Page 6

Thank you for your response. We really appreciate your input so far. If you are willing to participate in a max. 15-minute follow-up interview (by telephone or Skype) with the study team to discuss your experience of Europe PMC and/or similar services further, please provide your contact details below.

Please be assured that your contact details will not be shared outside the study team and will be deleted on completion of the study. Full details on how the study team will handle your data are available at http://www.technopolis-group.com/privacy-policy/

11) Your contact details

Name	[Single line text]
Email address	[Single line text]
Comments	[Single line text]

11) Would you be happy for us to share your contact details with Europe PMC for future user research?

[Radio buttons]

- Yes
- No

Appendix D Librarian survey questionnaire

The Wellcome Trust and other Europe PMC funders have appointed the independent research organisation, Technopolis, to carry out a study examining the value and impact of Europe PMC. We would like to understand who uses Europe PMC and how they do so. To that end, we would be grateful if you could complete a 5-minute questionnaire using the link below.

Responses can be provided anonymously. Full details on how the study team will use this information are available at http://www.technopolis-group.com/privacy-policy/.

Please note that questions with an asterisk are mandatory.

Page 1

About you

Question	Options
1) *In which country are you based?	[Dropdown menu]
2) *Which of these best describes you?	[Radio buttons] • Librarian/information professional employed by an academic institution • Librarian/information professional employed by a public research organisation • Librarian/information professional employed by a private sector research organisation • Librarian/information professional employed by charity/third sector organisation • Librarian/information professional employed by health service organisation • Librarian/information professional employed by a public library • Other (please specify)
3) * Please estimate how many unique users use your library/information services in a year. Please enter a number.	[Open text – single line]
4) Approximately what percentage of the users mentioned above are from the life sciences (including medicine and allied health subjects)?	[Sliding scale – 0 to 100%]

Page 2

Awareness of Europe PMC

Question	Options
5) Which of the following services does your library/information portal publicise (e.g. through user training/support, by providing links on your website, as resources etc.)? Please choose all that apply.	[multiple choice list] • PubMed • PubMed Central • OpenAire • Scopus • Web of Science • Google Scholar • Gateway to Research

Question	Options
	• ResearchGate • Dimensions • Google (including other web search engines) • Other resource (please specify)

Page 3

Your experience of Europe PMC

Question	Options
6) *Do you yourself use Europe PMC in your professional role?	[Radio buttons] • Yes • No

Page 4

Your experience of Europe PMC [only for those who responded 'yes' to Question 6]

Question	Options
7) Please briefly describe what you use Europe PMC for.	[Open text box]
8) In general terms, how helpful do you find Europe PMC?	[Radio buttons] • Not at all helpful • Somewhat helpful • Helpful • Very helpful • Extremely helpful • Don't know / no view
9) Please feel free to provide any further details on your experiences of using Europe PMC below (e.g. ease of use, how it has impacted your work, features you particularly like, any aspects you would like improved).	[Open text box]

Page 5

In order to help assess the cost-effectiveness of Europe PMC, we would like respondents to estimate what their institutions would be willing to pay for Europe PMC if its services were not free. Please note this question is intended to determine the value of Europe PMC and not to explore a move to charged services. (Costs are indicated in US Dollars; guiding exchange rates can be found at https://www.oanda.com/currency/average).

Question	Options
10) If your institution were asked to pay an annual subscription for Europe PMC, roughly how much do you think it would be willing to pay? Please note this question does not signal any intention to move to charged services.	[Radio buttons] • Up to $250 per year • $251 - 1000 per year • $1001 - 5000 per year • $5001 - 10000 per year • $10001 - 20000 per year

Question	Options
	• more than $20000, please specify
11) Please feel free to explain your choice above.	[Open text box]

Page 6

Thank you for your response. We really appreciate your input so far. If you are willing to participate in a max. 15-minute follow-up interview (by telephone or Skype) with the study team to discuss your experience of Europe PMC and/or similar services further, please provide your contact details below.

Please be assured that your contact details will not be shared outside the study team and will be deleted on completion of the study. Full details on how the study team will handle your data are available at http://www.technopolis-group.com/privacy-policy/

12) Your contact details

Name	[Single line text]
Email address	[Single line text]
Comments	[Single line text]

Appendix E Survey analysis

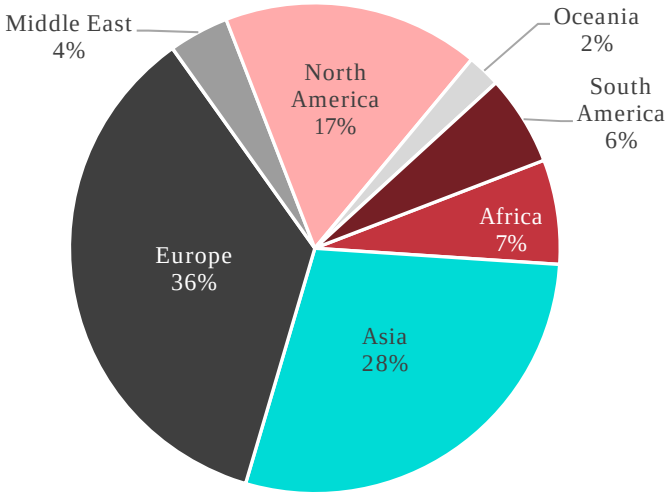
In this chapter, we present the summary findings from the surveys. The analysis includes partial responses provided respondents have answered at least one question about Europe PMC or comparator services. That is, responses of people who had only answered questions on the first page of a survey i.e. the 'About you' section (see Appendices B, C and D) were excluded.

E.1 User surveys

E.1.1 Website user survey

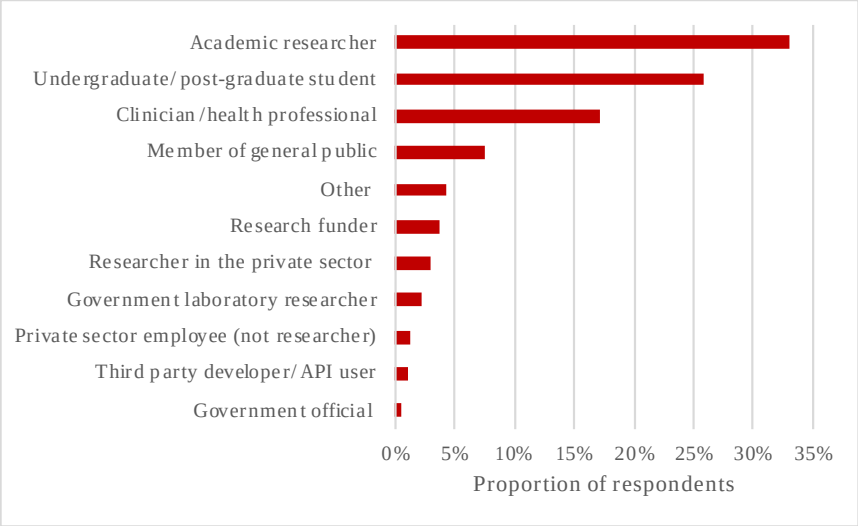
Between 17 September and 2 November 2018, 1162 individuals responded to the pop-up invitation to survey on the Europe PMC website. Of these, we included 958 complete and partial responses based on the inclusion criteria stated above.

Figure 11 Distribution of respondents by geography (a) and user type (b)
(a) n=958



Top 10 countries	% respondents
United States	11.3
India	8.8
China	5.9
United Kingdom	5.7
Brazil	4.0
Germany	4.0
Pakistan	3.7
France	3.7
Mexico	2.7
Canada	2.4

(b) n=958



Europe and Asia together accounted for 64% of respondents with the US, India, China and UK leading the number of responses received country-wise (Figure 11). To note, while the distribution by continent does not mirror the distribution of IP addresses according to 2018 monitoring data (Figure 3), it is broadly comparable to that of the previous survey[37]. However, the top four countries in terms of IP addresses are represented among the top 4 countries for respondents in this survey.

The majority of respondents were academic researchers or students (Figure 11). The 'other' category included librarians, teachers, publishers and editors, and other professionals. Members of the general public included patients and their family members.

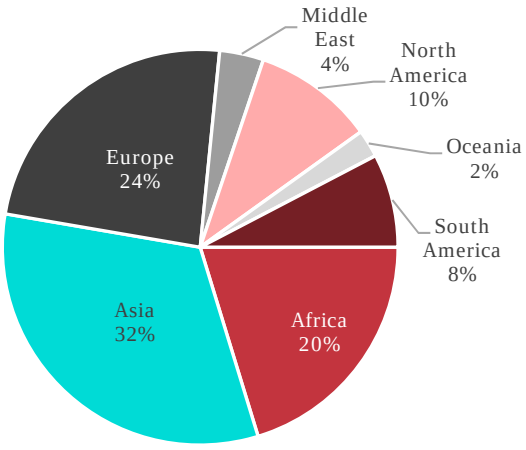
E.1.2 Email-based researcher survey

We received 398 survey responses to our email invitation between 4 September and 2 November 2018, of which all met the inclusion criteria for analysis. Since the invitation was sent out to corresponding authors of publications that appear on Europe PMC, the respondents included a cohort of 176 individuals who do not use Europe PMC.

The continental distribution of the survey respondents reflects neither that of the IP addresses (Figure 3) nor the 2017 survey[17] mainly due to a relatively much larger proportion of African respondents (Figure 30). For instance, South Africa was one of the top 5 countries for responses (Figure 12c). There were some differences in the geographical distribution among users and non-users as well (Figure 12). For example, for China we have a much lower proportion of non-users in our population and for the US we have a much lower proportion of Europe PMC users. Similarly, in the users group we have fewer North American respondents and more Asian respondents compared to the non-users group.

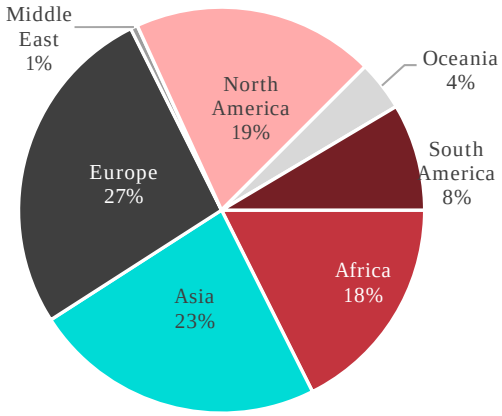
Figure 12 Distribution of respondents by geography

(a) Europe PMC users (n=222)



Top 5 countries (users)				
India	China	UK	US	Brazil
13%	13%	8%	7%	5%

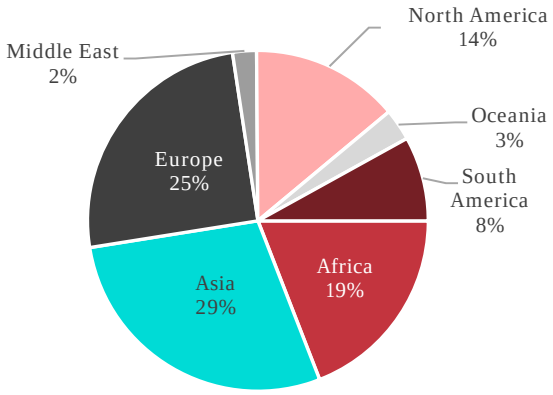
(b) Non-users (n=176)



Top 5 countries (non-users)				
India	US	South Africa	UK	Brazil
14%	14%	6%	6%	5%

[37] Report of Europe PMC Funder Committee Meeting, 9 October 2017

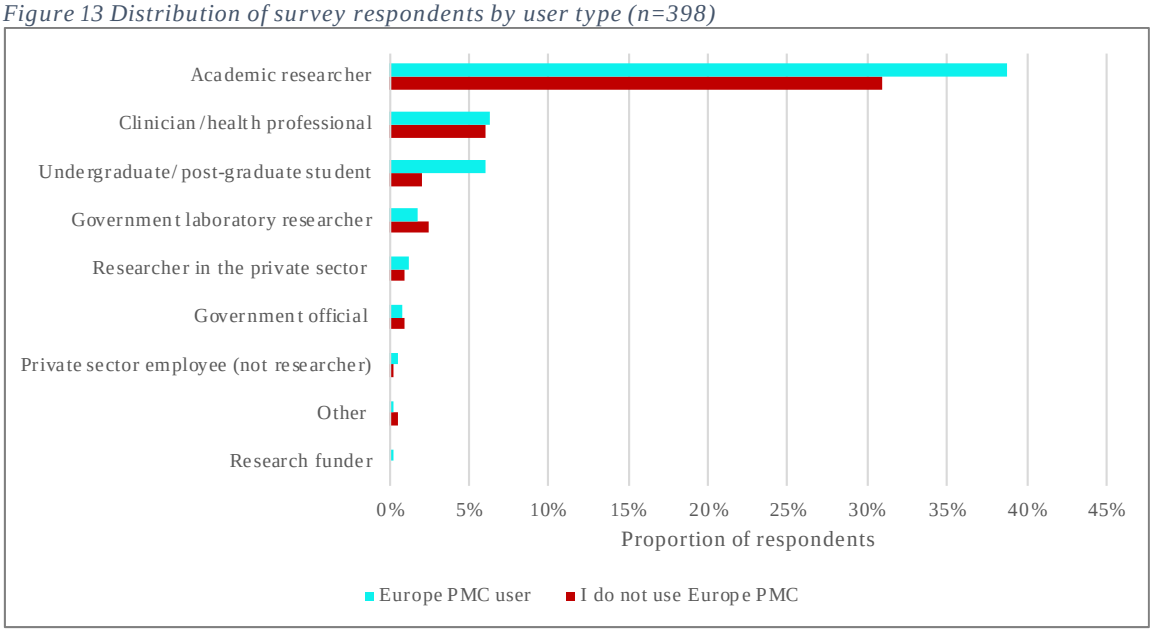
(c) Total (n=398)



Top 5 countries (total)				
India	US	China	UK	South Africa
13%	10%	9%	7%	5%

Nonetheless, across both populations (users and non-users), the majority of respondents identified as academic researchers (Figure 13).

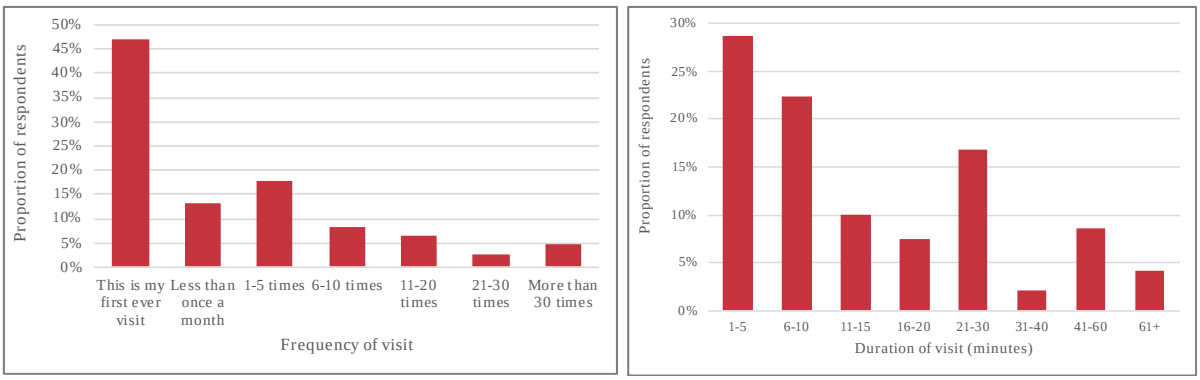
Figure 13 Distribution of survey respondents by user type (n=398)



E.1.3 Frequency and duration of visits to Europe PMC

When asked about frequency of visits, the most common answer was “This is my first ever visit” (47%, Figure 14).

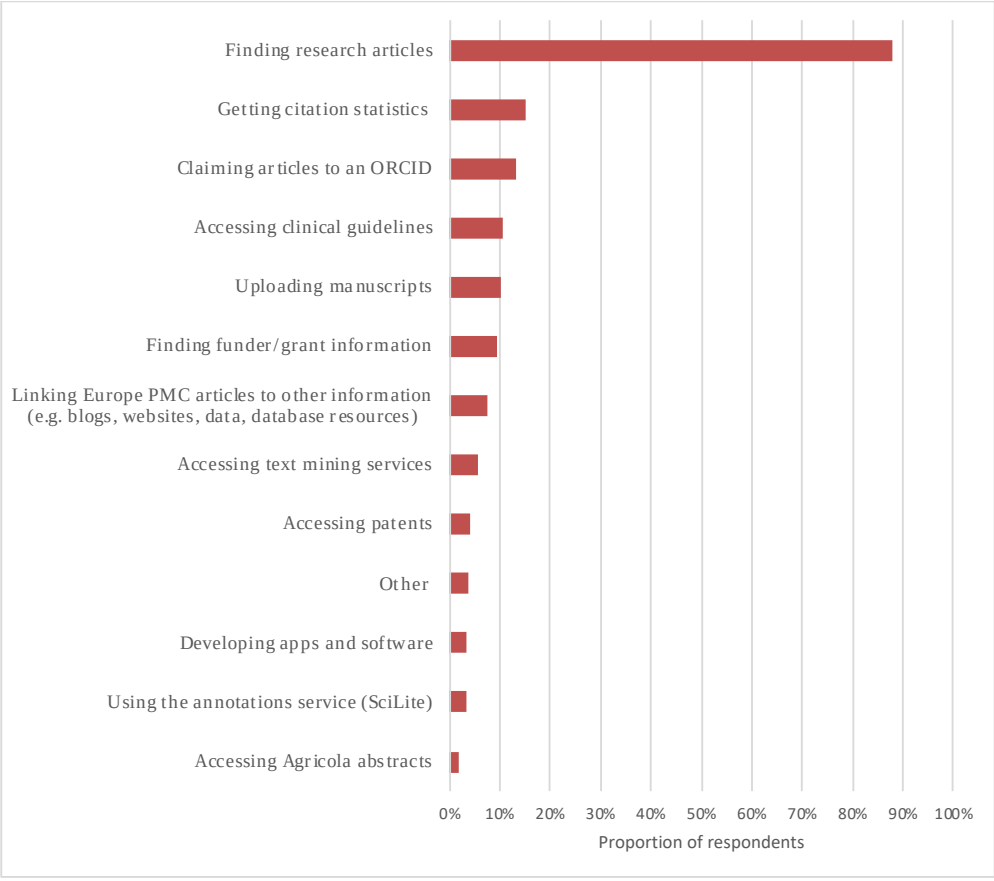
Figure 14 Distribution of respondents by number of visits to the Europe PMC website per month (a) and time spent per visit (b) (merged data from both the website user survey and the email-based researcher survey)
(a) n=1181 *(b) n=1162*



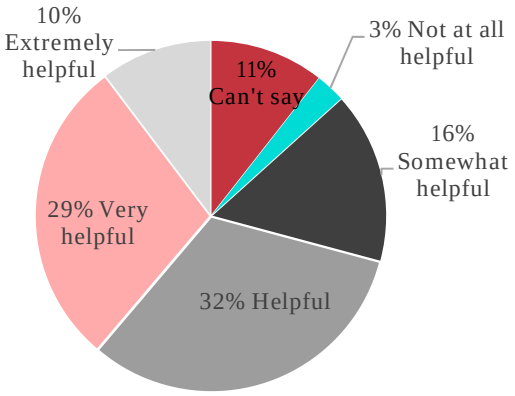
E.1.4 Usage and helpfulness of Europe PMC services

Figure 15 below indicates that Europe PMC is predominantly used for finding research articles with other services used to a much lower extent. Moreover, 87% of respondents find the Europe PMC services useful.

Figure 15 Distribution of respondents by type of use (a) and helpfulness (b) of Europe PMC services (merged data from both the website user survey and the email-based researcher survey).
(a) n=1086

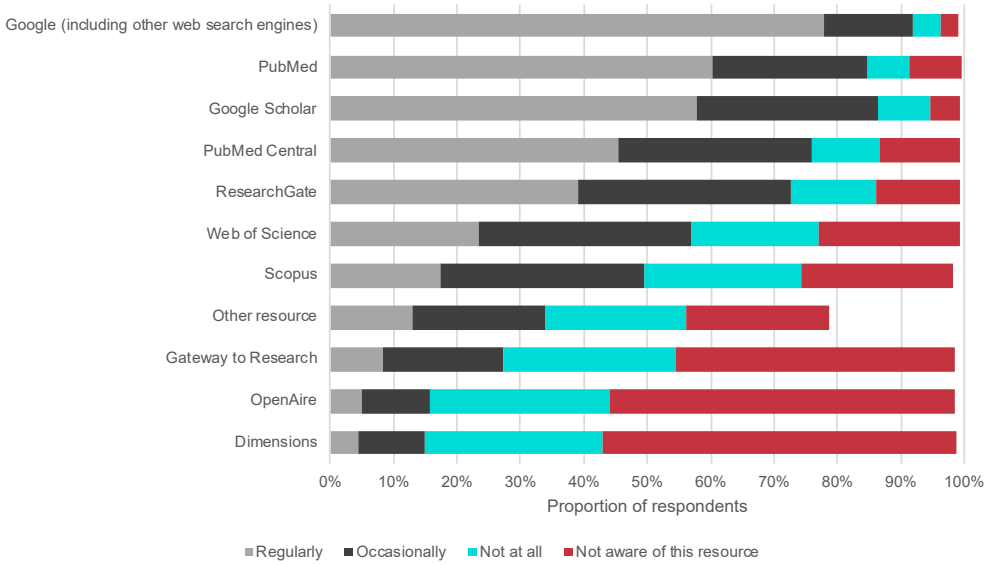


(b) n=1059



As expected from the findings of the 2017 survey[38] (Figure 31), search engines (e.g. Google search), PubMed, PMC and Google Scholar are also used regularly by Europe PMC users and non-users (Figure 16). Thus, Europe PMC is one among many different tools used. Other services that respondents also use but were not included among the options provided included publisher databases, Medline, Embase, SciFinder, Cochrane Library, Ovid, Sci-Hub and ScienceDirect.

Figure 16 Use of other services (merged data from both the website user survey and the email-based researcher survey) [n=1256]



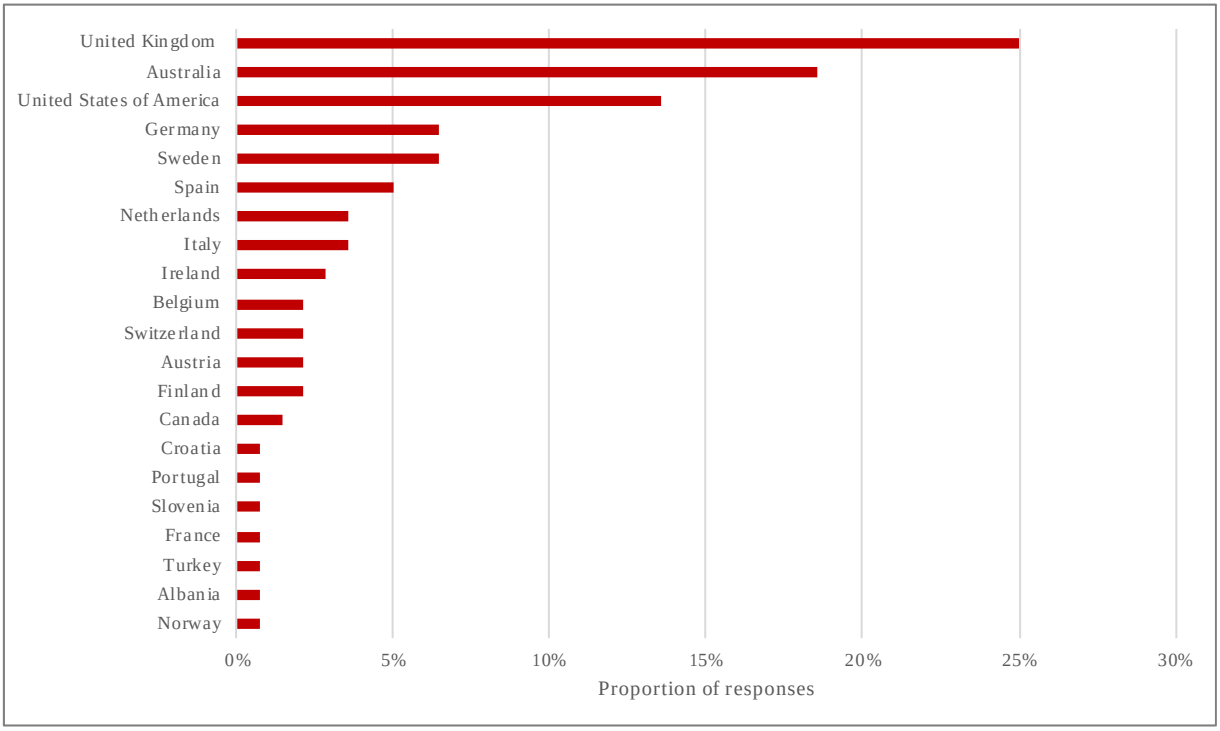
E.2 Librarians' survey

A total of 145 librarians responded to the survey between 4 September and 6 December 2018, 140 of which were included in the analysis. In terms of geographical distribution, the UK and Australia accounted for the greatest number of responses, 25% and 19%, respectively with the vast majority of respondents (68%) belonging to academic institutions (Figure 17).

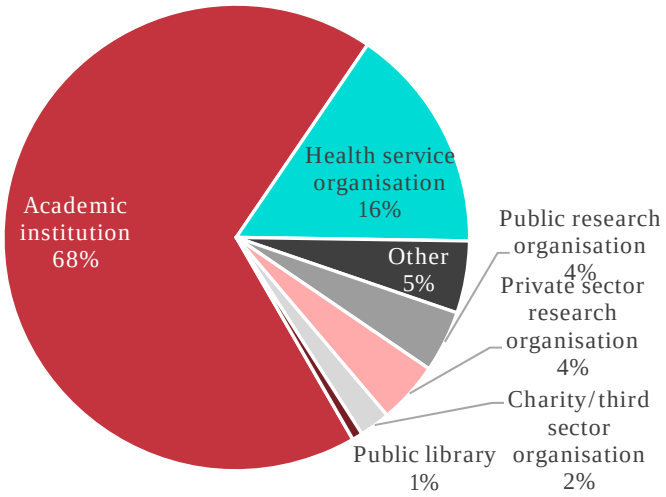
Figure 17 Responses to the librarians' survey by country (a) and organisation type (b)

(a) n= 140

[38] Report of Europe PMC Funder Committee Meeting, 9 October 2017



(b) n= 140



In addition, the organisations that the librarians are affiliated with range from small to very large, evidenced by the fact that librarians are catering to user numbers ranging from less than 500 to above 20,000. In over 40% of cases (n=59) services are being provided almost exclusively to a life sciences audience (90-100% of total users belonging to the life sciences) typically at academic institutions (n=26) or health service organisations (n=21). Figure 18 below shows the number of life science users that respondents are providing services to.

Figure 18 Number of life science users that the respondents are providing services to

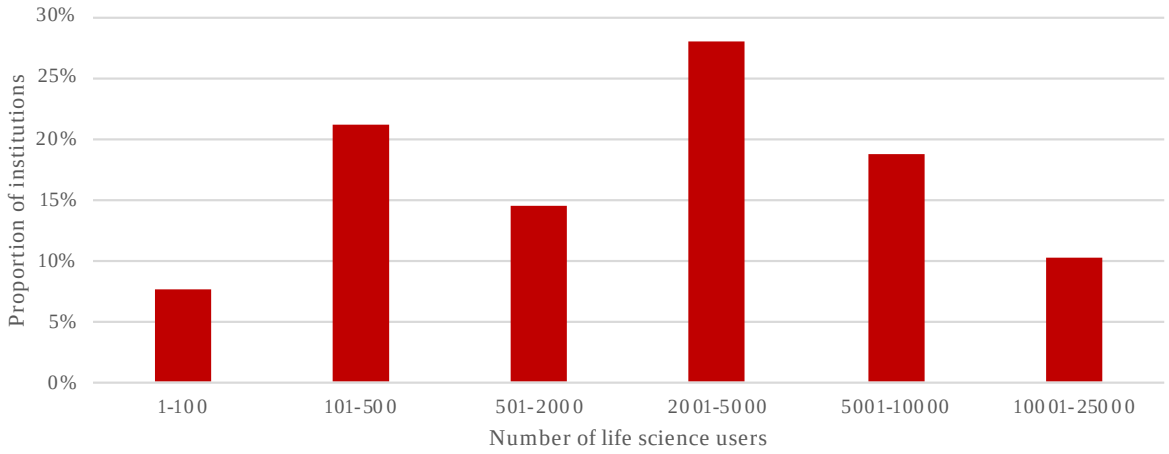
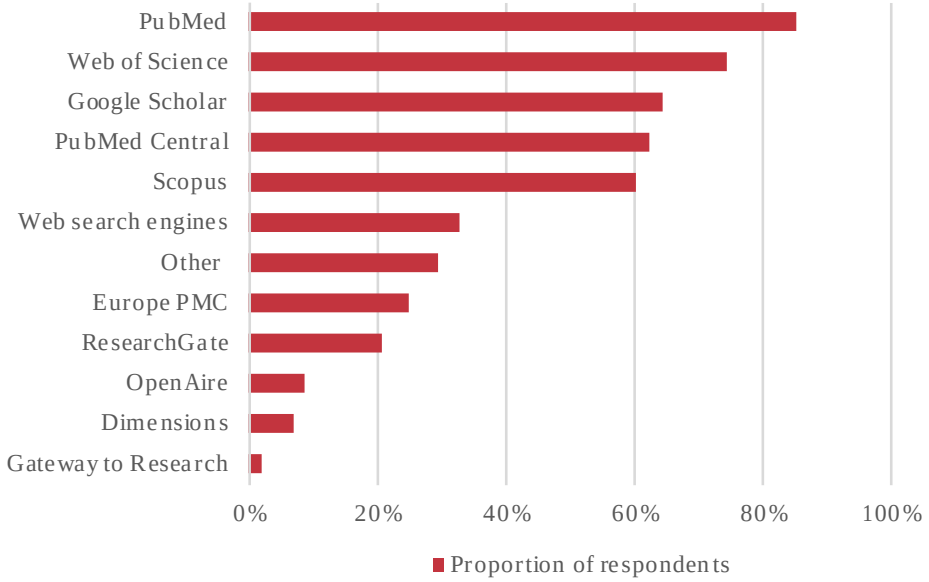


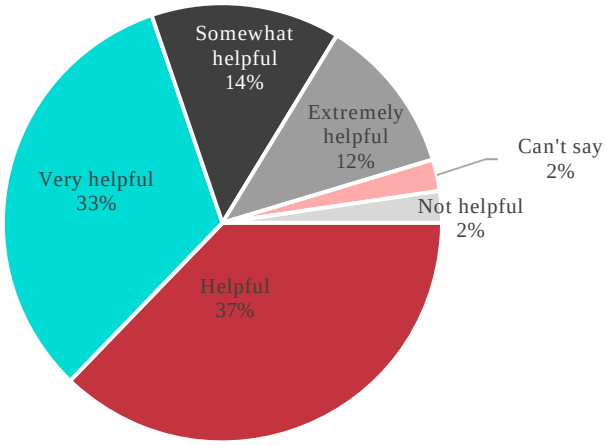
Figure 19 below shows the resources publicised on the respondents' library or information services portal for example via providing user training/support or links to the resource's website. Only 25% of respondents publicised Europe PMC compared to services like PubMed, Web of Science, Google Scholar, PubMed Central and Scopus, which were publicised more commonly.

Figure 19 Services promoted by respondents' library or information services portal [n=140]



Interestingly, 34% of respondents used Europe PMC in their professional role for tasks such as finding articles, ORCID linking and checking open access compliance. 96% of librarians who use Europe PMC find it helpful with 12% and 33% of respondents respectively finding it extremely or very helpful (Figure 20).

Figure 20 Distribution of respondents by helpfulness (n=43)

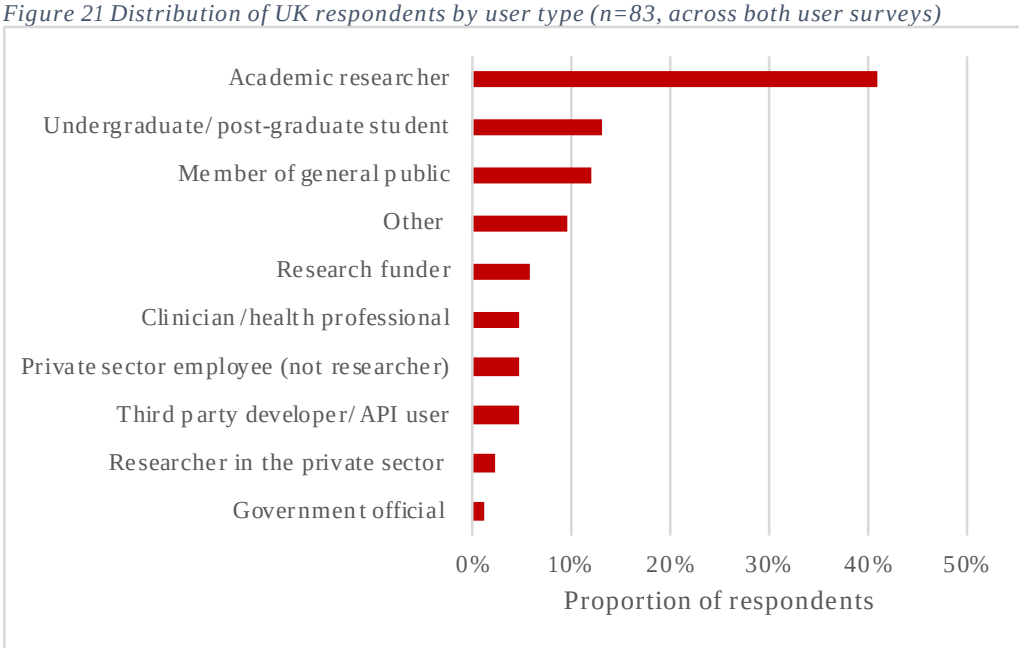


E.3 Analysis of UK survey respondents

E.3.1 User surveys

Across both surveys (website and email-based researcher surveys), there were 83 UK-based respondents with the majority of respondents identifying as academic researchers (Figure 21). Only 10 respondents reported that they did not use Europe PMC.

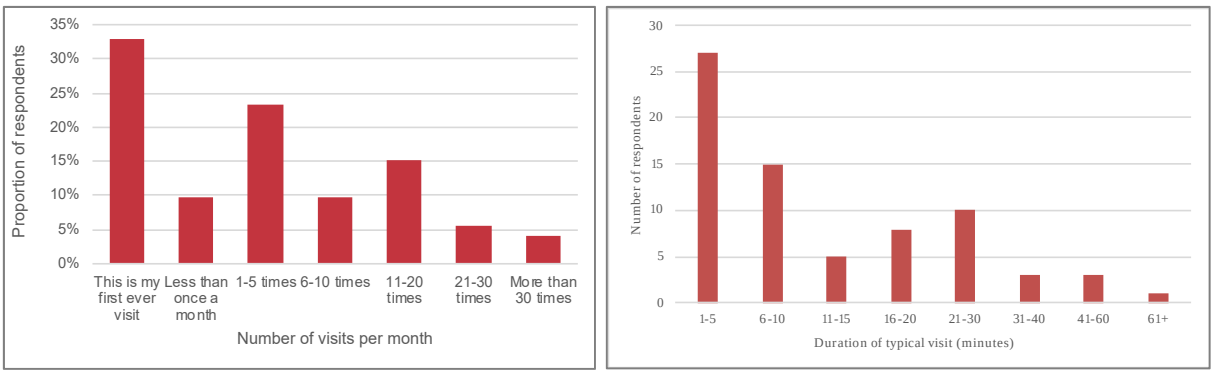
Figure 21 Distribution of UK respondents by user type (n=83, across both user surveys)



The frequency and duration of visits to the Europe PMC website is shown in Figure 22 below.

Figure 22 Distribution of UK respondents by number of visits to the Europe PMC website per month (a) and time spent per visit (b) (across both user surveys)

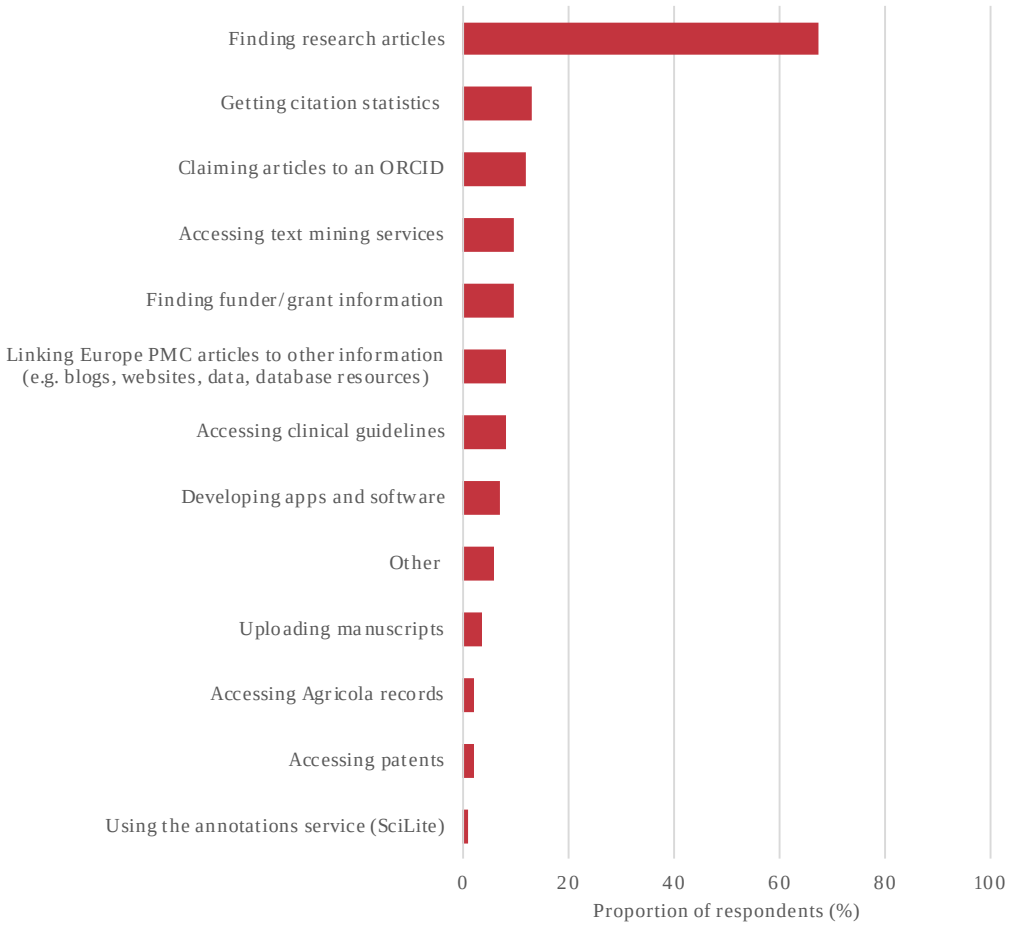
(a) n=73 *(b) n=72*



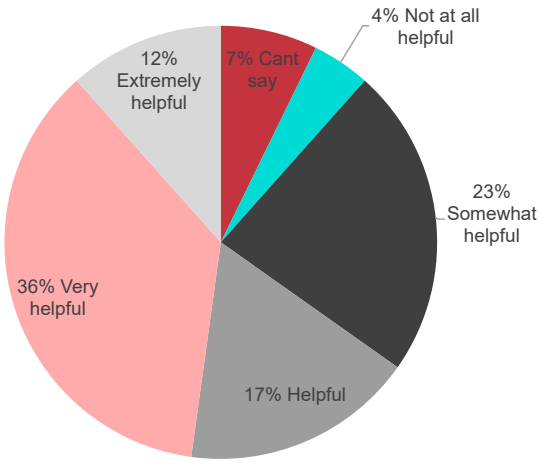
As shown in the overall survey results, UK users predominantly use Europe PMC for finding research articles (Figure 23). Moreover, 88% of respondents find the Europe PMC services useful to some extent.

Figure 23 Distribution of UK respondents by type of use (a) and helpfulness (b) of Europe PMC services (across both user surveys).

(a) n=83

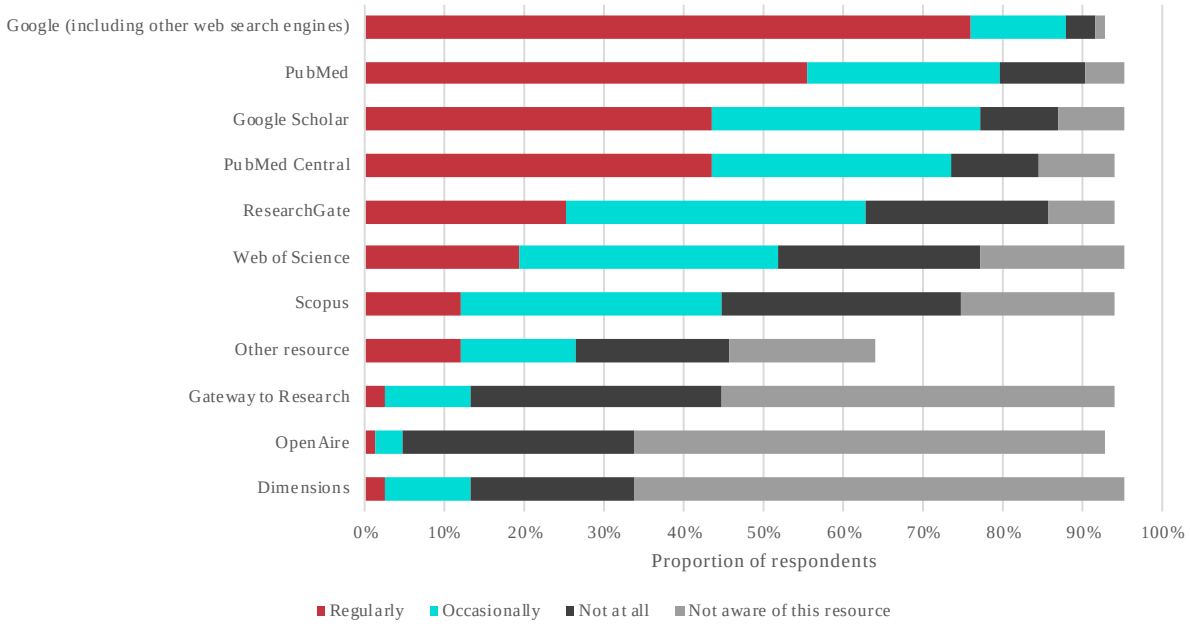


(b) n=69



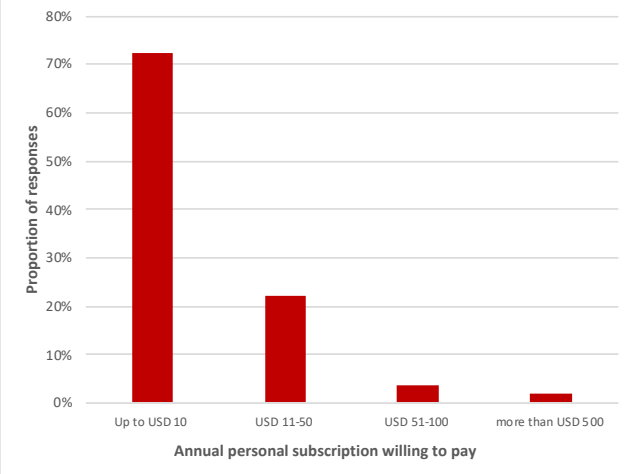
Web search engines including Google, PubMed, Google Scholar and PMC are also used regularly by UK survey respondents (Figure 24).

Figure 24 Use of other services by UK respondents (n=83, across both user surveys)



Most UK users also chose the lowest willingness to pay range (Figure 25). The average willingness to pay was $26 across all UK Europe PMC users (i.e. disregarding the non-users).

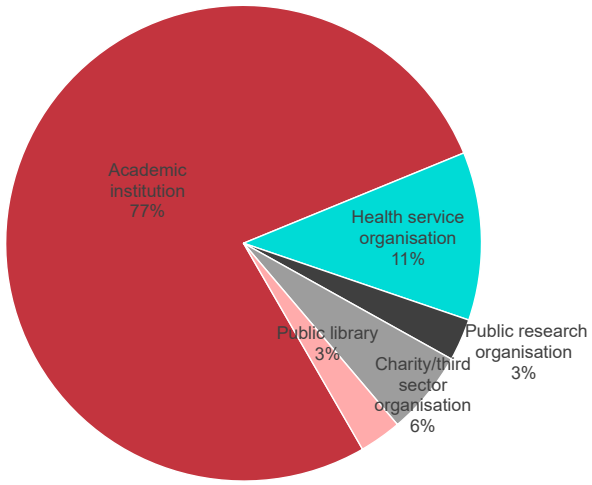
Figure 25 Distribution of willingness to pay amounts for UK respondents (n=54, across both user surveys)



E.3.2 Librarian survey

A total of 35 UK-based librarians responded to the survey. The vast majority of these (77%) belonged to academic institutions (Figure 26).

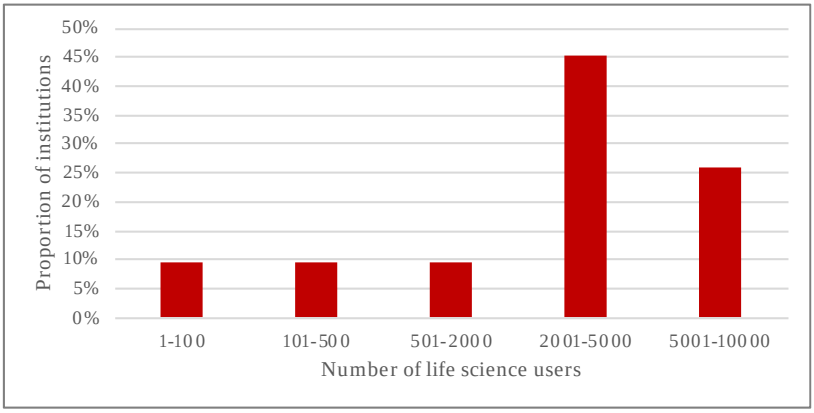
Figure 26 UK responses to the librarians' survey by organisation type [n=35]



In addition, over 70% of the organisations that the librarians are affiliated with had 2,000 to 10,000 life science users. Again, almost half of the librarians (48%) opted for the lowest willingness to pay range (up to $250) while about one-fourth each went for the next two available ranges $251-$1000 and $1001-$5000. Figure 31 below summarises these data. The average willingness to pay amount was $1.3.

Figure 27 Number of life science users (a) and willingness to pay amounts (b) for UK respondents to the librarian survey

(a) n=31



(b) n=31

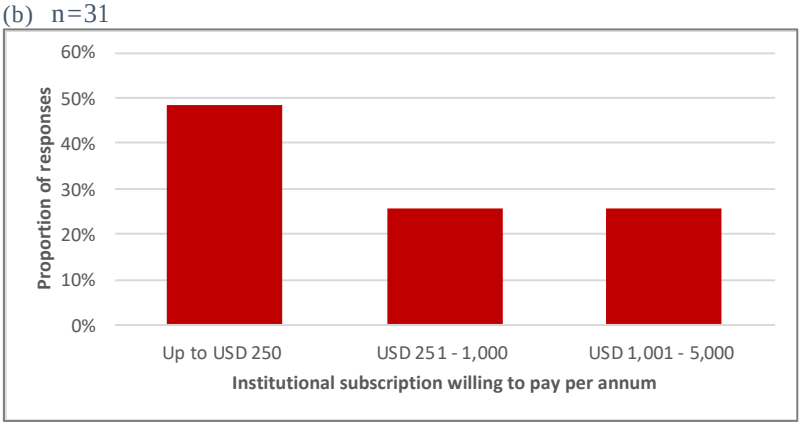


Figure 28 Services promoted by UK respondents' library or information services portal [n=35]

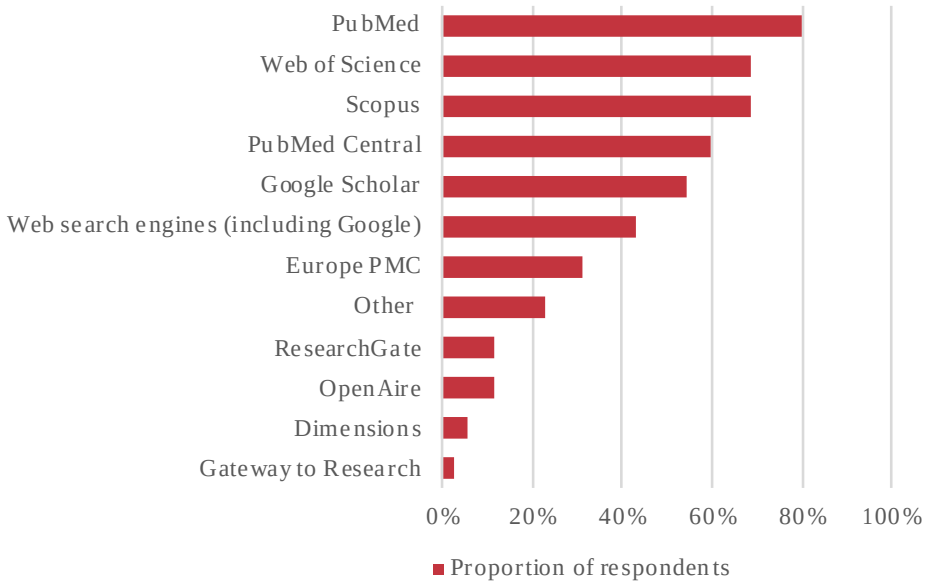
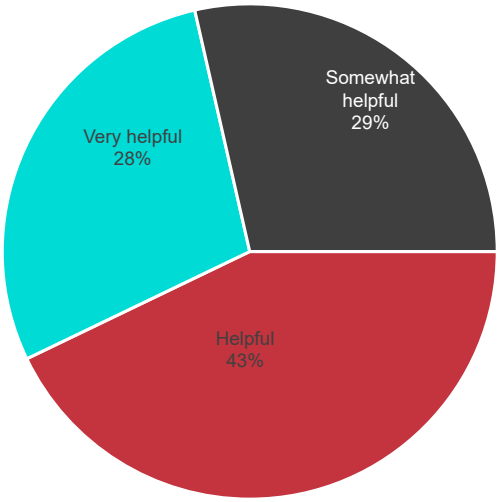


Figure 28 above shows the resources publicised on the respondents' library or information services portal for example via providing user training/support or links to the resource's website. Only 31% of respondents publicised Europe PMC compared to services like PubMed, Web of Science, Scopus, PubMed Central and Google Scholar, which were publicised more commonly.

Interestingly, 46% of respondents used Europe PMC in their professional role for tasks such as finding articles and checking open access compliance. All librarians who use Europe PMC find it helpful with 29% and 43% of respondents respectively finding it very helpful or helpful (Figure 29).

Figure 29 Distribution of UK respondents by helpfulness (n=14)



E.4 Previous Surveys of Europe PMC

User research is an ongoing element of Europe PMC and several studies have been completed in the last 3 years on different aspects, for example, in relation to specific service features and to inform the future roadmap. For this report, we have only referred to information from two of these studies, as described below.

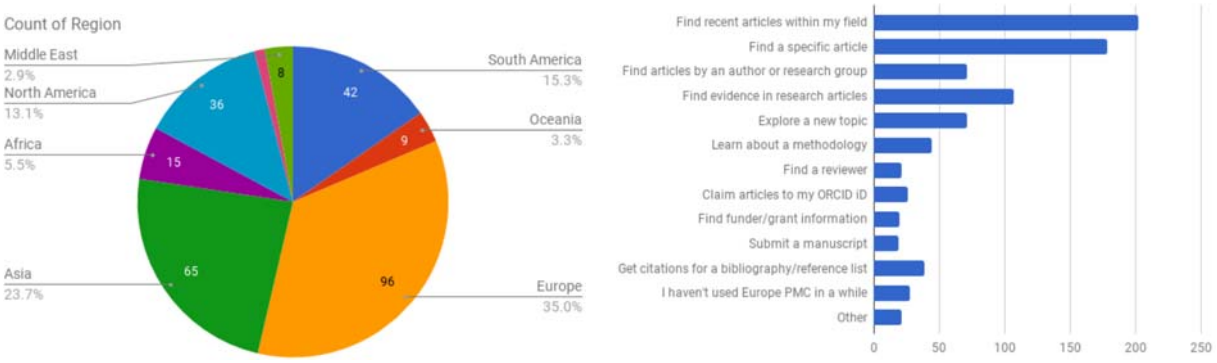
A usability study was undertaken in 2014 to help understand how the Europe PMC website is perceived by its core user communities and potentially to use this feedback in future development of Europe PMC[39]. 15 stakeholder types were consulted which included research scientists, health care professionals, librarians and members of the public. The results suggested that the Europe PMC site was 'reasonably easy to use' and that 'the content was useful'. A number of the issues raised during the survey included the preference to have only a single search results list and to produce a better static content help page. A prominent issue was that users did not fully understand that Europe PMC includes all of the content of PubMed and PMC.

A wider user survey[40] was conducted during the 6-week period from 29 June to 8 August 2017. This survey attracted a total of 389 respondents and showed that Europe PMC users are typically researchers (20%) and students (16%) and based in universities (37%). The majority of users found Europe PMC via a search engine or a link from another website. However, word of mouth also contributed significantly to how people came about the website. Geographical distribution of survey respondents and manner of use are shown in Figure 30 below. The former does not mirror the geographical distribution of the IP addresses shown in Figure 3 above, but it is clear that the most common use of Europe PMC is to find articles.

[39] Europe PMC Annual Report (2014)

[40] Report of Europe PMC Funder Committee Meeting, 9 October 2017

Figure 30 Geographical distribution of Europe PMC survey respondents (left panel) and manner of use (right panel)

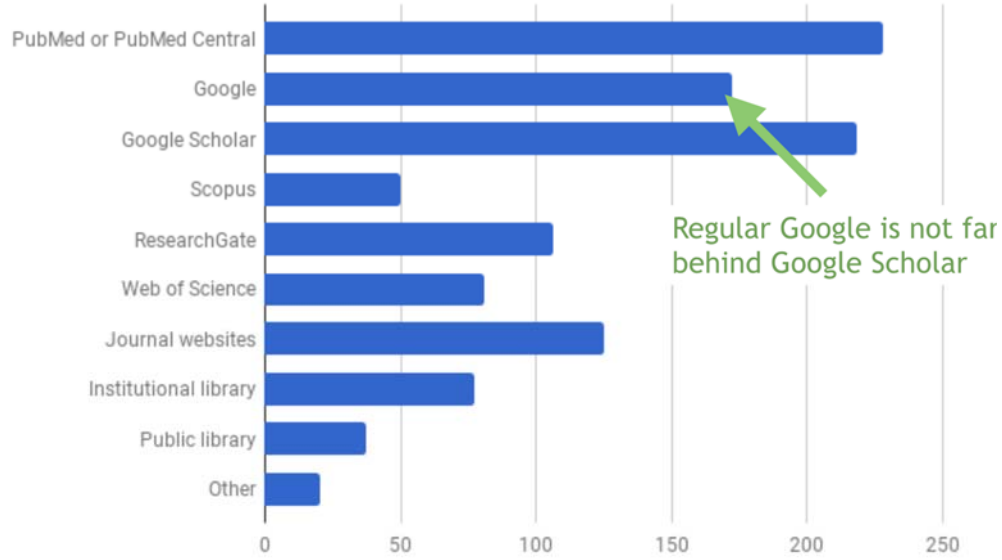


Source: Europe PMC survey (2017)

When asked about what they liked about Europe PMC, users most commonly cited "access to free full text articles" and "ease of use", while what they most disliked was "limited access to free full text articles"[41].

The 2017 Europe PMC survey also investigated the use of comparator services by Europe PMC users. It found that PubMed, PMC and Google Scholar ranked highest (Figure 31).

Figure 31 Use of comparator services by Europe PMC users



Source: Europe PMC survey (2017)

[41] Report of Europe PMC Funder Committee Meeting, 9 October 2017

Appendix F Usage valuation

This appendix sets out the more detailed findings of the usage valuation of Europe PMC. A detailed methodology can be found in Appendix A, setting out all the key steps involved in our assessment.

Taking each of these steps in turn, we establish below the key calculations and assumptions involved in our methodology.

Step 0: The list of unique IP addresses (number of users)

Using Europe PMC monitoring statistics, we know that in 2018, the site was accessed by 11.7 million unique IPs. We considered the number of IPs to be a reasonable proxy for the number of Europe PMC visitors. It does not however, provide a completely accurate representation on user numbers. For instance, several researchers may use the same IP address while a single user may appear as multiple IP addresses if they tend to work across different sites. Automated and robot activities can also account for some unique IP addresses. We assume these respective under-estimates and over-estimates will largely cancel each other out.

Step 1: Segmenting users into different user groups (number of users per user group)

To reach a more accurate view on the economic value of Europe PMC, we estimated the value associated with 21 different user groups or segments. This involved several steps:

- Segmenting users by geography

Europe PMC data provided the country of all IPs that visited the website. Drawing in large part on UN guidance[42], we categorised each country as being either a developing economy, transition economy or developed economy. Using this, we calculated that 53.3% of users are from developed economies, 44.9% from developing economies, and 1.8% from transition economies.

We applied these splits to the total number of users in 2018 (step 1), helping divide users into three segments:

Table 15 Segmentation of Europe PMC users by geography

Country type	No. of users
Developed economies	6,240,565
Transition economies	215,255
Developing economies	5,261,274
Total	11,717,094

Source: EMBL-EBI monitoring data, Technopolis analysis

- Further segmentation by user type

We divided each of the three geographical segments into seven further segments according to user type – giving us a total of 21 segments. Table 16 lists the user categories we used. Drawing on the website survey data concerning the professional background of each respondent, we estimated for each geographical segment, the total share of respondents coming from each user category. We then applied this to our list of users. For instance, in our survey we found that 33.6% of users from developed economies were academic researchers and thus applied this split to the developed economy segment.

[42] See http://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf

Table 16 Breakdown of segments by user type

User category	Developed economies		Transition economies		Developing economies	
	Share of users	Total users	Share of users	Total users	Share of users	Total users
Academic researcher	29.0%	1,809,191	31.4%	67,652	34.0%	1,789,918
Clinician	14.9%	927,497	8.6%	18,450	17.7%	931,119
Non-academic researcher	6.4%	400,770	14.3%	30,751	4.6%	244,080
General public	21.8%	1,362,619	11.4%	24,601	7.2%	379,680
Research funder	2.6%	160,308	8.6%	18,450	5.2%	271,200
Industry	2.8%	171,759	2.9%	6,150	1.7%	90,400
Student	22.6%	1,408,421	22.9%	49,201	29.6%	1,554,878
Total	100.0%	6,240,565	100.0%	215,255	100.0%	5,261,274

Source: Technopolis analysis. 'Share of users' derived from website survey data, and 'total users' derived from EMBL-EBI monitoring data.

Step 2: Time per user group

Using the user survey data (combined data from both surveys), we calculated the median duration of each Europe PMC visit for each segment, as shown in Table 17 below. This has provided the basis for determining the economic value of Europe PMC per user segment.

We have chosen not to calculate our figures using the mean as there are a small number of users whose visits tend to be disproportionately long in time. Using the mean would risk us using an over-inflated average. Similarly, we have not used the mode because the large sample size means that there were very few respondents sharing precisely the same duration of visit.

Table 17 Median duration of each Europe PMC visit (minutes)

User type	Developed economies	Transition economies	Developing economies
Academic researcher	10.0	10.0	10.0
Clinician	10.0	3.0	15.0
Non-academic researcher	8.0	17.5	30.0
General public	10.0	0.5	5.5
Research funder	20.0	0.0	5.0
Industry	5.0	0.0	0.0
Student	10.0	13.5	10.0

Source: Technopolis analysis of Europe PMC user survey

Step 3: Frequency of use by user group

Again, using the user survey data, we have calculated the mean number of visits each user segment makes per month (as shown in Table 18). In making these calculations, we have made a series of assumptions:

- The survey asked respondents to select which range of values best described the number of times they visited Europe PMC per month (e.g. 1-5 times, 6-10 times). We have used the mid-point of each range as a proxy for the precise number of visits made
- We have assumed that anyone stating a use of Europe PMC 'less than once a month' uses it 0.5 times a month
- Where respondents claimed they used Europe PMC 'more than 30 times a month,' we have assumed a value of 31 times per month
- Where respondents stated that their visit to Europe PMC was their 'first ever visit,' we have assumed that such users will only use Europe PMC once a month. In reality, some of these will go on to use Europe PMC more regularly. However, our assumption accounts for the fact that in any given month, there will always be new first-time users (i.e. our survey sample is an accurate snapshot of what is seen in a typical month).

Table 18 Mean number of visits to Europe PMC per month

User type	Developed economies	Transition economies	Developing economies
Academic researcher	6.06	3.75	3.77
Clinician	4.82	1.50	3.28
Non-academic researcher	3.50	12.90	8.84
General public	2.49	1.00	7.98
Research funder	9.71	1.00	4.35
Industry	4.95	-	11.67
Student	3.66	6.21	3.48

Source: Technopolis analysis of Europe PMC user survey

Step 4: Usage value per user group

Since our methodology measures economic value in terms of its opportunity cost, each user group commands a different value from their usage of Europe PMC. A student user for instance will have a lower opportunity cost in their Europe PMC usage than a clinician because of the wage differentials between them. As such, the economic value per user from Europe PMC usage will be higher for a clinician than for a student.

We used secondary data to find the average wage for each user type, shown in Table 19. We express average wages in terms of Purchasing Power Parity (PPP). PPP is a concept that shows the relative cost of buying an identical basket of goods in different countries, stripping out transaction costs, exchange rate differences, and differing costs of living. Wages expressed in PPP therefore provide a more accurate view of the opportunity costs of using Europe PMC worldwide. For example, a worker in a developed economy will tend to have a higher PPP than an identical worker in a developing economy. This is because in the developed economy, the wage commands a higher purchasing power (i.e. can buy more goods and services) than in the developing economy. Using PPP in our analysis therefore allows us to control for the fact opportunity costs will vary by geography.

Table 19 Time valuation per user type (median wage per hour worked- PPP $)[43]

User type	Proxy measure used	Developed economy	Transition economy	Developing
Academic researcher	Wages for 'Education'	22.43	5.67	6.64
Clinician	Wages for 'Human health and social services'	18.70	5.48	5.81
Non-academic researcher	Wages for 'Professional, scientific and technical activities'	20.68	7.87	6.37
General public	Average wage per country	17.89	6.13	4.28
Research funder	Wages for 'Professional, scientific and technical activities'	20.68	7.87	6.37
Industry	Wages for 'Manufacturing' and 'Information and communication'	21.86	7.75	4.87
Student	See note[44]	17.24	5.91	4.12

Source: Technopolis analysis of ILOSTAT data for 2006-2018, related to monthly earnings and hours worked

Step 5: Total economic value for each user group

Using the results from Steps 1-4, we have calculated the economic value associated with each segment's usage of Europe PMC. This involved:

5. Multiplying the median duration of a visit (Step 2) by the mean visits per month (Step 3)
6. From this, we derived the average duration of a visit (in hours) per year for each user segment
7. We multiplied this figure with the total usage value per user group (Step 4), giving us the value of Europe PMC usage per user
8. We multiplied the valuation per user by the number of users in each respective segment (Step 1)

The tables below show the economic value associated with Europe PMC usage amongst each of the 21 segments.

Table 20 Value of Europe PMC usage for all users in developed economies

User type	Median duration of visit (minutes)	Mean visits per month	Amount of time spent using Europe PMC per year (hours)	Usage value per hour (PPP $)	Value from visits per year (PPP $)	Total no. of users	Total value of Europe PMC usage (PPP $ per year)
Academic researcher	10.0	6.06	12.1	22.43	272.03	1,809,191	492,152,345

[43] Wages per hour here calculating by taking the median value for the PPP $ per hour worked in all relevant countries. For each country, calculations are based on the latest available data.

[44] ILOSTATS does not provide a suitable proxy wage for students. Instead we used UK data from the Annual Survey of Hours and Earnings 2018, and the *UK Complete University Guide* (derived from the Higher Education Statistics Agency 2016-17) and calculated the average annual wage for university leavers was 96% that of the average UK annual wage. We have therefore scaled down the average wage per country accordingly to derive an estimated PPP $ for students.

User type	Median duration of visit (minutes)	Mean visits per month	Amount of time spent using Europe PMC per year (hours)	Usage value per hour (PPP $)	Value from visits per year (PPP $)	Total no. of users	Total value of Europe PMC usage (PPP $ per year)
Clinician	10.0	4.82	9.6	18.70	180.24	927,467	167,171,247
Non-academic researcher	8.0	3.50	5.6	20.68	115.83	400,770	46,421,929
General public	10.0	2.49	5.0	17.89	89.26	1,362,619	121,626,681
Research funder	20.0	9.71	3.8	20.68	803.24	160,308	128,765,588
Industry	5.0	4.95	5.0	21.86	108.30	171,759	18,600,938
Student	10.0	3.66	7.3	17.24	126.21	1,408,421	177,751,649
Total						6,240,565	1,152,490,377

Source: Technopolis analysis

Table 21 Value of Europe PMC usage for all users in transition economies

User type	Median duration of visit (minutes)	Mean visits per month	Amount of time spent using Europe PMC per year (hours)	Usage value per hour (PPP $)	Value from visits per year (PPP $)	Total no. of users	Total value of Europe PMC usage (PPP $ per year)
Academic researcher	10.0	3.75	7.5	5.67	42.50	67,652	2,874,893
Clinician	3.0	1.50	0.9	5.48	4.93	18,450	91,036
Non-academic researcher	17.5	12.90	45.2	7.87	355.49	30,751	10,931,611
General public	0.5	1.00	0.1	6.13	0.61	24,601	15,091
Research funder	0.0	1.00	0.0	7.87	0.00	18,450	-
Industry	0.0	0.00	0.0	7.75	0.00	6,150	-
Student	13.5	6.21	16.8	5.91	99.18	49,201	4,879,584
Total						215,255	18,792,215

Source: Technopolis analysis

Table 22 Value of Europe PMC usage for all users in developing economies

User type	Median duration of visit (minutes)	Mean visits per month	Amount of time spent using Europe PMC per year (hours)	Usage value per hour (PPP $)	Value from visits per year (PPP $)	Total no. of users	Total value of Europe PMC usage (PPP $ per year)
Academic researcher	10.0	3.77	7.5	6.64	50.08	1,789,918	89,636,515
Clinician	15.0	3.28	9.8	5.81	57.04	931,119	53,114,450
Non-academic researcher	30.0	8.84	53.0	6.37	337.97	244,080	82,490,910
General public	5.5	7.98	8.8	4.28	37.59	379,680	14,271,246
Research funder	5.0	4.35	4.3	6.37	27.70	271,200	7,512,553
Industry	0.0	11.67	0.0	4.87	-	90,400	-
Student	10.0	3.48	7.0	4.12	28.72	1,554,878	44,656,127
Total						5,261,274	291,681,802

Source: Technopolis analysis

F.1 Analysis findings

Using the methodology set out above, we estimate that Europe PMC is worth $1.5 billion (PPP) a year to its users as a whole, as shown in Table 23. This is equivalent to $125 per user per year, given that Europe PMC has nearly 12 million unique users every year.

Table 23 Total economic value of Europe PMC usage (PPP $)

	Total value per year (PPP $)
Developed economies	1,152,490,377
Transition economies	18,792,215
Developing economies	291,681,802
Total	1,462,964,394

Source: Technopolis analysis

The above values assume a 100% attribution of value to Europe PMC, with no directly comparable services available. We know that while no other platforms provide exactly the same range of services as Europe PMC, alternative platforms collectively do cover most of Europe PMC's services and functionalities. Based on the Scottish Enterprise ready reckoner (Table 14), we could assume a 75% deadweight if we consider a high level of the usage value is not exclusively dependent on Europe PMC or a 95% deadweight if a very high level of the usage value is not dependent solely on Europe PMC being available. On applying these weights to the findings above, the total economic value of Europe PMC usage ranges from $73 million to $366 million, and a per user value of $6 to $31 per year for 95% and 75% deadweight respectively.

technopolis |group| United Kingdom
3 Pavilion Buildings
Brighton BN1 1EE
United Kingdom
T +44 1273 204320
E info@technopolis-group.com
www.technopolis-group.com