

# Europe PMC: a gold and green OA repository

---

Promoting open access to research outputs

*Making 100% of scientific publications available under open access:  
how can that objective be achieved?*

5<sup>th</sup> Couperin's Open Access meeting – January 24/25, 2013

Robert Kiley, Wellcome Trust  
r.kiley@wellcome.ac.uk

**wellcome**trust

## Agenda



Europe  
PubMed  
Central

1. Brief overview of Europe PMC
2. Discuss how content is added to the repository
3. Discuss Wellcome experience of working with publishers
4. Consider costs of “gold”, OA publishing
5. Highlight some of the unique functionality Europe PMC offers
6. Invite other research funders to join Europe PMC

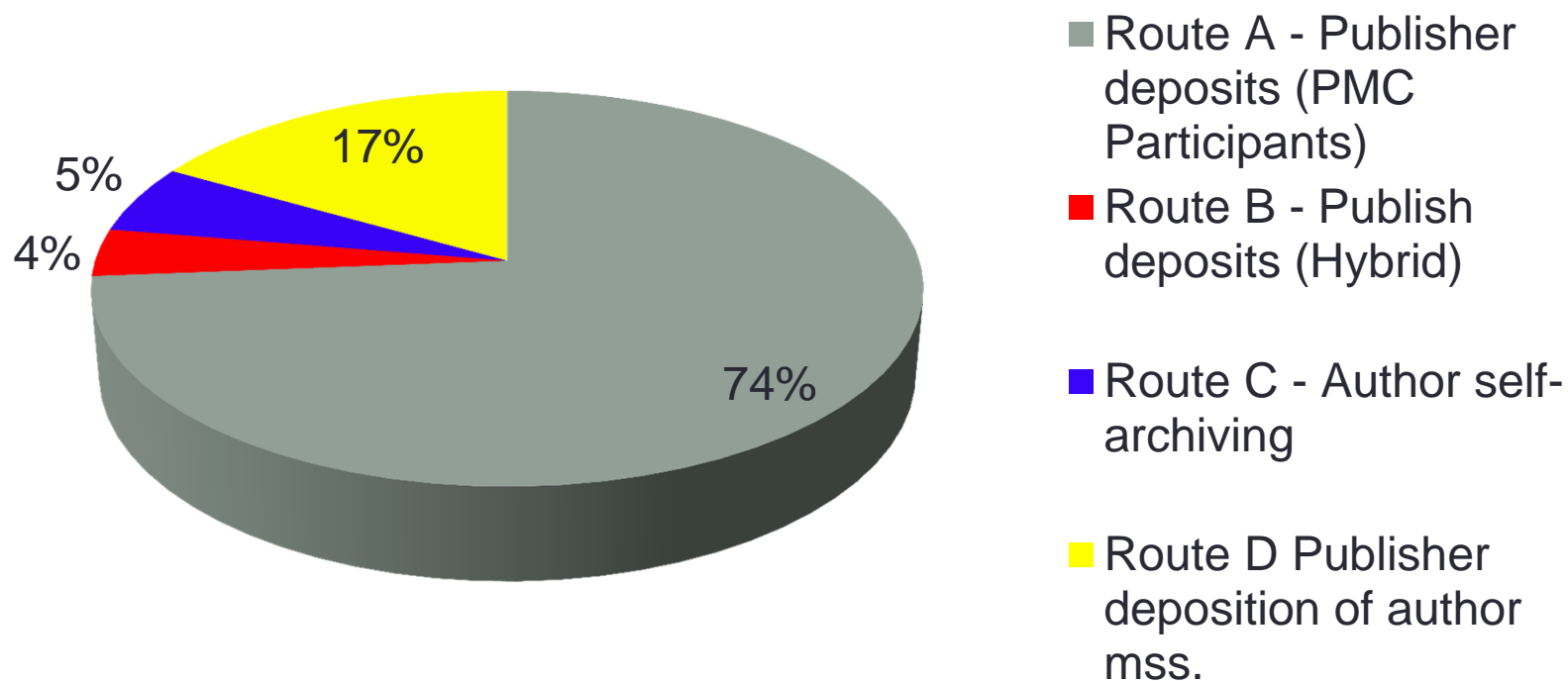
## Europe PubMed Central (Europe PMC)

- A service of the [Europe PMC Funders' Group](#) working in partnership with [European Bioinformatics Institute](#), [University of Manchester](#) [British Library](#) in cooperation with the NCBI.
- It includes content provided to the [PubMed Central International archive](#) by *participating publishers*
- 19 funders mandate deposition of research outputs into Europe PMC
  - including ERC, MRC, Wellcome Trust, CR-UK, FWF
- Provides free access to over 2.6m full text, peer-reviewed research papers, + 27m bibliographic records derived from PubMed, Agricola, Patents etc..

## How content is added to (Europe) PMC

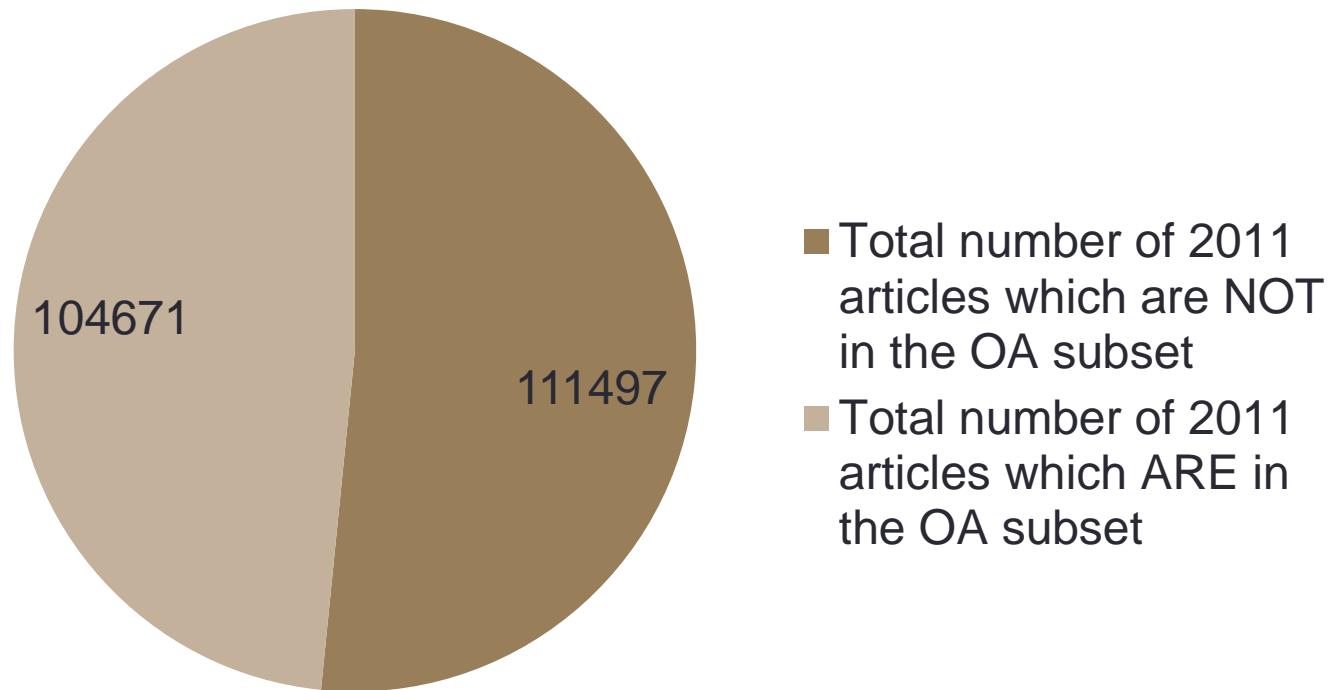
- Route A: By **publishers** who make all their content available at the time of publication
  - (e.g. *PLOS, BMC, eLife*)
- Route B: By **publishers** who make individual articles available at the time of publication
  - (e.g. articles hybrid in hybrid journals )
- Route C: By **authors** who self-archive the author manuscripts in PMC/Europe PMC
  - (e.g. articles published in *Science*)
- Route D: By **publishers** who deposit the peer-reviewed manuscript (authors also have to “sign-off the deposit)
  - (e.g. *Elsevier articles for NIH, NPG content for Wellcome [and others]*)

## How content is added to (Europe) PMC – by numbers



- 216,000 articles published in 2011, available in PMC
- 96% of this content deposited by publishers

## 2011 articles – how many articles have reuse licences?



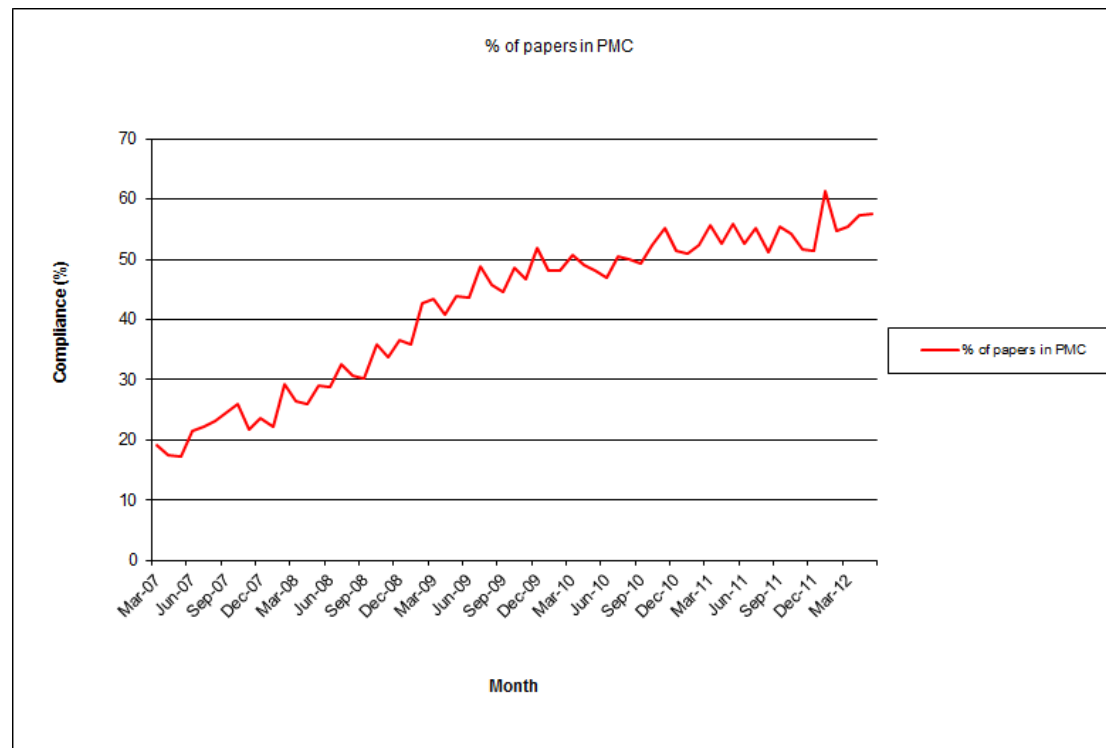
- 48% of content added to PMC in 2011 was included in OA subset

## Publishers: Wellcome experience

- Wellcome has worked with publishers to specify services when OA fee is paid
  - Deposition of final version in PMC, in XML
  - Licence to facilitate re-use. From April 2013, must use CC-BY
- Also worked with publishers (e.g. NPG) when “green” version made available
  - Licence attached to each AAM deposited
  - Articles included in OA subset and available for download (in XML) via OAI and FTP
    - Link to [article](#) in Europe PMC or via [OAI](#) or via [FTP](#)
- But...also have to deal with publisher “take down” requests
- Set up “Publishers Panel” to discuss OA and scholarly publishing

## Wellcome Trust funded content in Europe PMC

- Compliance with OA mandate still disappointing ( $\approx 60\%$ )
- 90% of this content is deposited by publishers
  - 10% of WT content is ingested via the self-archiving route



**Compliance – around 55%-60%**



# Strengthening Wellcome Trust policy (June 2012)

- Specific sanctions for non-compliance:
  - withholding final payment on grants, pending assurance papers listed on final reports are compliant
  - requiring previous Trust-funded papers to be compliant before any funding renewals or new grants awards are activated
  - discounting non-compliant Trust-funded papers as part of a researcher's track record



## Cost of OA publications (1)

- Data published by University of Edinburgh<sup>1</sup> shows that over the last 4 years, average APC (for Wellcome Trust funded research papers) is **£1741**

Year	No.of articles	Total cost	Mean APC
2007/08	19	£32,712.85	£1,721.73
2008/09	38	£58,648.88	£1,543.39
2009/10	49	£90,304.78	£1,842.95
2010/11	67	£124,359.79	£1,856.11
2011/12	87	£146,687.10	£1,686.06
Total	260	£452,713.40	£1,741.21

Table 1: High-level details of the Wellcome Trust open access expenditure at UoE

1. Gold Open Access: Counting the Costs. *Ariadne*, 2012. <http://www.riadne.ac.uk/issue70/andrew>

## Cost of OA publications (2)

- Analysis of OA spend (for Trust-funded research) from 30 UK universities shows that in 2011-12:
  - Some 1770 papers published under an author-pays model
  - Total cost to Trust of £3,181,278
  - Average APC = **£1797** (close to the Edinburgh data, of £1741; £56 difference)
- If 100% of Trust funded research incurred via an APC, total annual spend would be **£8,985,000**
  - i.e. £1797 x 5000 [number of papers attributed to Trust in an average year]
- Equates to **1.2%** of Trust's annual research spend
  - Annual research spend in 2011-12 was £746m
    - £9m / £746m = 1.2%
    - If our spend had been, say, £600m, then OA costs equal 1.5% of our research spend

## Benefits of a Europe-wide, subject-based repository

- Greater visibility of research findings
  - All articles in a single repository, fully accessible from PubMed and Google (as well as Europe PMC)
- Research process enhanced
  - Innovative features - citation networks, articles integrated with underlying data-sets; text mining
- Greater compliance with funder mandate
  - Publishers can (and do) deposit in PMC
- Integration with Institutional Repositories (IR's)
  - Content can be shared with IR's
  - Integration with ORCID

# Visibility of research outputs

Google search results for "ap1903 blood brain barrier". The search bar shows the query and a magnifying glass icon. Below the search bar are tabs for "Web", "Images", "Maps", "Shopping", "More", and "Search tools". The results show "About 2,230,000 results (0.46 seconds)". The first result is a link to a paper: "A general chemical method to regulate protein stability in the ...". This result is highlighted with a red box. Below it are other results including a PDF titled "ARGIN1™ Regulated Homodimerization Kit - ARIAD ..." and a Wikipedia entry for "Blood-brain barrier". At the bottom of the page is the Google UK logo.

Google Scholar search results for "p53". The search bar shows the query and a magnifying glass icon. Below the search bar are tabs for "Articles", "Legal documents", "Any time", "Sort by relevance", and "Sort by date". The results show "About 1,210,000 results (0.03 sec)". The first result is a link to a paper: "WAF1, a potential mediator of p53 tumor suppression". This result is highlighted with a red box. Below it are other results including "p53 mutations in human cancers", "The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation", "The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53", and "Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms". At the bottom of the page is the Google UK logo.

# Research process enhanced – innovative services (1)

Europe PubMed Central

gwa studies

Results

Results: All citations (12564) | Full text articles (15766) | Sort by: Relevance | Date | **Times Cited**

- Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.**  
(PMID:17554300) Free resource  
Wellcome Trust Case Control Consortium  
Nature [2007, 447(7145):661-678]  
**Cited: 2806 times**
- PLINK: a tool set for whole-genome association and population-based linkage analyses.**  
(PMID:17701901) Free resource  
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC  
Am J Hum Genet [2007, 81(3):559-575]  
**Cited: 2732 times**
- Principal components analysis corrects for stratification in genome-wide association studies.**  
(PMID:16862161)  
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D  
Nat Genet [2006, 38(8):904-909]  
**Cited: 1275 times**
- Finding the missing heritability of complex diseases.**  
(PMID:19812666) Free resource  
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM  
Nature [2009, 461(7265):747-753]  
Cited: 993 times

**Sort results by times cited**

**View “Cited by” and “Cites the following” data**

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.  
(PMID:17554300)

Abstract Citations BioEntities Related Articles

**Cited By** displaying 25 of 2806 citations • Web of Science® times cited (2963)-subscription required

- A second generation human haplotype map of over 3.1 million SNPs.**  
(PMID:17943122)  
International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam

Show all items

**Cites the following** displaying 118 of 118 citations

- Genome-wide association studies for common diseases and complex traits.**  
(PMID:15716906)  
Hirschhorn JN, Daly MJ.  
Nat. Rev. Genet. [2005]
- Evaluating coverage of genome-wide association studies.**  
(PMID:16715099)  
Barrett JC, Cardon LR.  
Nat. Genet. [2006]

# Research process enhanced – innovative services (2)

The screenshot shows the EvidenceFinder interface with a search for 'malaria'. The results page displays three articles with highlighted text mining findings:

- Structure-function of falcipains: malarial cysteine proteases.** (PMCID:PMC3317066) Free resource. Pandey KC, Dixit R. J Trop Med [2012, 2012:345195]. Cited: 0 times. Extracts relevant to your question: Malaria caused by **Plasmodium falciparum** is responsible for 880,000 deaths per year worldwide [1].
- Antimalarial plants of northeast India: An overview.** (PMCID:PMC3326788) Free resource. Shankar R, Deb S, Sharma BK. J Ayurveda Integr Med [2012, 3(1):10-16]. Cited: 0 times. Extracts relevant to your question: Malaria is caused by **single-celled protozoan parasites called Plasmodium** and transmitted to man through the Anopheles mosquito.
- Development and optimization of a novel 384-well anti-malarial imaging assay validated for high-throughput screening.** (PMCID:PMC3247113) Free resource. Duffy S, Avery VM. Am J Trop Med Hyg [2012, 86(1):84-92]. Cited: 0 times. Extracts relevant to your question: However, the actual number of RBC per ml of culture in a standard Malaria culturing protocol can be quite variable caused by **multiple factors such as cell lysis caused by parasite replication, disturbance of compacted RBC suspension, and pipetting accuracy of 100% compact RBC.**

On the right side of the interface, a sidebar titled 'Ask a question' lists various queries related to malaria, with 'What causes Malaria? (840)' highlighted in red.

**Evidence Finder - text mining technologies**

**Acromine - disambiguation tool**

*Refine your query* (suggestions by Acromine)

- cardiopulmonary resuscitation
- cytochrome P450 reductase
- contraceptive prevalence rate
- cardio-pulmonary resuscitation
- computer-based patient record
- clinical prediction rule
- Central Population Register
- civil personal registration
- Continuous Plankton Recorder
- computerized patient record
- clinical pregnancy rate
- conjugate peak refinement
- calibrated population resistance
- clinical partial response
- complete pathological response
- after cardiorespiratory resuscitation
- Central Person Registry
- progesterone receptor
- Chasing, Preoccupation and Risked Relationships
- Cumulative percent revision
- critical perfusion rate
- Central Person Register
- Central Population Registry
- C-reactive protein

## Integration with IR's

- Already good working relationship between Europe PMC and initiatives such as Driver and OpenAire
  - Metadata from Europe PMC is harvested by [Driver](#)
  - OpenAire use the xml full text to mine content for FP7 funding information
- Content can be pulled from Europe PMC to an IR
  - OAI interface available (<http://europepmc.org/oai.cgi?>)
- Application under development that uses the Europe PMC web service to push metadata of full text articles (in OA subset) to Repository Junction (RJ)
  - RJ will then push articles to the relevant university IR
- Win-win situation
  - IR's can still showcase the outputs from their institution
  - Content becomes more easily discoverable through Europe PMC



## Policy initiatives

- UK – Finch and RCUK
- Increasingly strong drive to support OA – particularly the European Commission statement which set expectation for policies at national level

- *If Europe is to reap the benefits of wider access to scientific research results, clear policies are needed both national and European*

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/era-communication-towards-better-access-to-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf)

But

- *National initiatives and practices are still fragmented, thus preventing the European Union from realising its full research and innovation potential*

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/background-paper-open-access-october-2012\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/background-paper-open-access-october-2012_en.pdf)

- A Europe-wide repository will help realise these ambitions

## Invitation to join Europe PMC: rules of engagement

- Funder to develop/modify its OA mandate to specify that all funded, life sciences research, should be made available through the Europe PMC repository
- Funder policy is clarified such that if funds are used to cover OA publishing costs, then the *publisher* is required to deposit the paper in Europe PMC at the time of publication
  - This builds on the work already done by Trust and others
  - Ensures that articles published in hybrid journals, are made accessible and minable through Europe PMC
- Funder makes a contribution (based on its research spend) to contribute to the running costs of Europe PMC

## Indicative costs (per annum)

Funder	Annual Research Spent	Contribution to Europe PMC running costs (% of total costs)	Contribution to Europe PMC running costs (£/€) per annum
Small	£50m	2%	£14k/€17k
Medium	£175m	7%	£48k/€58k
Large	£650m	25%	£176k/€212k

*Plus contribution to the R&D development budget*

# Conclusion

- The Europe PMC Funders have developed a function-rich and well-used repository
- Europe PMC is:
  - cost effective
  - will help to increase visibility of research you fund
  - will provide you with the means of implementing your OA policies
- Funders invited to join Europe PMC and help shape the onward development of the service

