# Wellcome Trust APC spend (2014-15): an explanation of the data

## Overview

In March [2014](http://blog.wellcome.ac.uk/2014/03/28/the-cost-of-open-access-publishing-a-progress-report/) and March [2015](http://blog.wellcome.ac.uk/2015/03/03/the-reckoning-an-analysis-of-wellcome-trust-open-access-spend-2013-14/) the Wellcome Trust published details of its open access spend for the year previously, **as reported by UK institutions and the Trust’s Major Overseas Programmes in receipt of an OA block grant** (see: <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Guides/WTX036803.htm>). The aim of this exercise was to make the debate around the costs of open access publishing more evidence based.

In 2014 six research funders established the Charity Open Access Fund ([COAF](http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Charity-open-access-fund/)) in order to provide a single funding mechanism to cover APC’s. So this year (March 2016), we are publishing the data for 2014-2015 Charity Open Access Fund (COAF) open access spend. As in 2015 in addition to simply reporting spend, we have run the data through our “compliance monitoring” tool (developed for us by [Cottage Labs](http://cottagelabs.com/)) to help us determine programmatically whether the paper is in the Europe PMC repository and what licence (if any) is attached to the article.

The dataset *only* includes information when an APC was levied. If an author has self-archived a paper, or the publisher provides a gold OA service (but makes no charge), this information is not included in this dataset.

Equally, data are not included in cases where a researcher (based at an institution not in receipt of an OA block grant) received a supplement to their grant to cover OA publishing costs.

## The dataset

The dataset contains the following elements:

### PMCID/PMID/DOI

To help the Trust identify whether a work is in PMC/Europe PMC, we ask institutions to cite the PMCID (or the PubMed ID if a PMC ID is not known). We also ask for the article DOI. Where possible, any missing IDs have been inserted by the automated compliance checking tool.

### Publisher

This lists the name of the publisher, as reported by the institution. We do not impose any name authority control on this field, so the same publisher is listed in different ways by different institutions. For example, OUP, Oxford University Press, and O.U.P. This year we have not tried to normalise this data.

As this data is self-reported by institutions in receipt of an OA grant it may contain errors (e.g. the wrong publisher may be listed).

### Journal name

This lists the name of the journal (which published the COAF-funded work), as reported by the institution. Again, we do not impose any name authority control on this field, so the same journal is listed in different ways by different institutions. For example, PNAS, Proceedings of the National Academy of Sciences, P.N.A.S. This year we have not tried to normalise this data.

As this data is self-reported by institutions in receipt of an OA grant it may contain errors (e.g. the wrong journal may be listed).

**ISSN**

The automated compliance checking tool supplemented the data provided by the institutions with the journal ISSN where it was able to identify it in Europe PMC.

### Article title

This is the title of the article, as reported by the institution.

### Cost

The cost listed is the figure which the institution has reported as the full cost of the APC. These costs should exclude any page or colour charges which the publisher may levy.

The cost quoted, includes VAT where applicable. All costs have been converted into £ sterling.

As you look through the data you will see (for example) different prices cited for articles published in the same journal. When these differences are small, it can be assumed that this reflects fluctuations in currency prices. Where the differences are more significant it suggests either an error in the data being reported or additional costs (e.g. page and colour charges) have been included,

**Full text in Europe PMC?**

The automated compliance checking tool checks whether the full text of the article appears in Europe PMC.

If it does, it will be marked as TRUE, if not, it will be marked as FALSE.

**Accepted author manuscript?**

The automated compliance checking tool checks whether the version of the paper in Europe PMC is the Accepted Author Manuscript or the final published version. If an APC has been paid the final version of the paper should be in Europe PMC, otherwise it is not compliant.

If the article is the author manuscript, it will be marked as TRUE. If it is the final version it will be marked as FALSE. If it was not possible to identify its status it will be marked as “unknown”.

**Licence**

Institutions are not currently asked to provide licence details. The automated compliance checking tool attempts to detect the licence for the article in three places in the following order: first, it checks the XML of the article in Europe PMC, second, it checks the Europe PMC web page, finally, it checks the publisher’s site.

If the licence can be identified as a Creative Commons licence it is marked as such. If another licence can be found it will be marked as “non-standard-licence”. If no licence information can be found it will be marked as “unknown”.

**Licence source**

This column reflects where the licence information was found, as described above.

**Journal Type**

If the automated compliance checking tool can find the journal in [DOAJ](http://doaj.org/) it will be marked as “oa”, if it does not, it will be marked as “hybrid”.

**Wellcome Trust compliant?**

The automated compliance checking tool calculates whether an article is compliant based on whether the full text is in Europe PMC **and** the detected licence is CC-BY or CC0. As APCs were paid for all articles in this dataset compliance has been manually set to false for all accepted author manuscripts.

## Comment

Notwithstanding the issues with the data (described above) we hope this this data will be of use to help better understand the cost of OA publishing.

A blog summarising the findings from the data can be found at: <http://blog.wellcome.ac.uk/>

Robert Kiley

3rd March 2016