

# UK PubMed Central explained in Nucleic Acids Research Paper

Martin Fenner, Gobbledygook

November 14, 2010

Last Tuesday Nucleic Acids Research published a nice paper describing the UK PubMed Central (UKPMC) database (McEntyre 2010). UKPMC was started in 2007, the enhanced version described in the paper was launched January 2010. In November 2009 I published an interview with Phil Vaughan, the senior author of the paper. The paper talks about the specific enhancements done to PubMed Central, including an integrated search of PubMed and PubMed Central, “Cited by information” and semantically enriched content generated by text mining.

Content profile for PMC and UKPMC. Figure 3 from McEntyre JR et al.

Thanks to Duncan Hull we know that PubMed currently contains information about 20 million papers (Twenty million papers in PubMed: a triumph or a tragedy?). About 10% of these papers are available as fulltext from PubMed Central. What wasn’t clear to me and what I learned from the paper is that only 194,000 papers, or 1% of PubMed content, are from the PMC Open Access Subset (and that includes papers with a non-commercial OA license). All these papers are available as download or via PMC-OAI service. Although 1.8 million papers (the majority digitized back issues) can be freely downloaded as full-text from PubMed Central, they carry a publisher copyright and can’t be reused for research purposes (e.g. full-text mining) without an explicit permission from the publisher.

## References

McEntyre JR, Ananiadou S, Andrews S, Black WJ, Boulderstone R, Buttery P, et al. UKPMC: a full text article resource for the life sciences. Nucleic Acids Research. 2010 November; DOI: <http://doi.org/10.1093/nar/gkq1063>.