

Title: Life as a Palaeontologist: Academia, the Internet and Creative Commons

Author(s): Ross Mounce^{*1}

Volume: 2

Article: 10

Page(s): 1-10

Published Date: 01/12/2012

PermaLink: <http://www.palaeontologyonline.com/articles/2012/life-as-a-palaeontologist-academia-the-internet-and-creative-commons/>

IMPORTANT

Your use of the Palaeontology [online] archive indicates your acceptance of Palaeontology [online]'s Terms and Conditions of Use, available at <http://www.palaeontologyonline.com/site-information/terms-and-conditions/>.

COPYRIGHT

Palaeontology [online] (www.palaeontologyonline.com) publishes all work, unless otherwise stated, under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license.



This license lets others distribute, remix, tweak, and build upon the published work, even commercially, as long as they credit Palaeontology[online] for the original creation. This is the most accommodating of licenses offered by Creative Commons and is recommended for maximum dissemination of published material.

Further details are available at <http://www.palaeontologyonline.com/site-information/copyright/>.

CITATION OF ARTICLE

Please cite the following published work as:

Mounce, Ross. 2012. Life as a Palaeontologist: Academia, the Internet and Creative Commons
Palaeontology Online, Volume 2, Article 12, 1-10.

Life As A Palaeontologist: Academia, the Internet and Creative Commons

by Ross Mounce^{*1}

Introduction:

The results of scientific research can be of interest to experts and non-experts alike. This is perhaps especially true for palaeontology, which captures public interest — but obtaining access to this information is sometimes difficult, even for scientists. Taking a rather different tack from previous Palaeontology [online] articles, I'm going to provide a brief overview of how the Internet has changed and is significantly changing palaeontology and academia in general, helping to open up research for the greater benefit of science and society.



FIGURE 1 — SIR TIM BERNERS-LEE SENDS A MESSAGE AT THE LONDON 2012 OLYMPICS.

When Sir Tim Berners-Lee helped to invent the World Wide Web more than 20 years ago, he did it 'for everyone' (see Fig. 1), and to this day he still campaigns to maintain open web standards. Had he patented his technology, or asserted restrictions on it, [the world would certainly be a very different place](#) and you probably wouldn't be reading this website on a freely accessible network. The very fact that it is open and royalty-free is part and parcel of the tremendous success of the Internet.

For professional and amateur palaeontologists, access to information on the Internet is extremely valuable. Being able to see pictures of specimens, data from measurements and 3D scans saves significant effort being otherwise duplicated — not to mention time and travel costs given that fossils are scattered all over the world.

It is important to highlight the difference between free and Open. If digital content is 'free' to access, as is, for example, the website of the newspaper *The Guardian*, it can be viewed without paying money, but the viewer does not get any further legal permissions unless specifically stated. By contrast, with Open content or data, "anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike" (see [opendefinition.org](#)).

The vast majority of Open content on the Internet, including Palaeontology [online], is now provided under a particular set of legal licences created by [Creative Commons](#). These allow authors of digital works to clearly and simply make explicit the ways in which they allow their work to be shared, reused, remixed and redistributed. There are only three Open Creative Commons licences, each denoted by its own logo:

Creative Commons Zero (CCO)

CREATIVE COMMONS ZERO (CCO)

The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighbouring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

Creative Commons Attribution (CC BY, or CC-BY)

CREATIVE COMMONS ATTRIBUTION (CC BY, OR CC-BY)

If you alter, transform, or build upon this work, you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Creative Commons Attribution Share-Alike (CC BY-SA)

CREATIVE COMMONS ATTRIBUTION SHARE-ALIKE (CC BY-SA)

If you alter, transform, or build upon this work, you may distribute the resulting work only under the same license to this one AND you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

The most recent authoritative survey, in December 2010, estimated that there were [more than 160 million different](#) works on the Internet that used one of these three licences. Some websites specifically track the usage of CC licences; for example, more than [59 million photos](#) have been made Open on the photo-sharing website Flickr.

Science has also been hugely transformed in the past decade by digital openness (see Video 1).

VIDEO: http://www.youtube.com/watch?v=hZAcTNFzF-s&feature=player_embedded

VIDEO 1 — SCIENCE COMMONS EXPLAINS HOW CREATIVE COMMONS ARE TRYING TO MAKE SCIENCE BETTER. IT'S DIRECTED BY JESSE DYLAN AND IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE LICENCE.

Creative Commons licensing has been an integral part of the success of open-access publishing under the CC BY license. "CC has provided a strong, consistent signal that you can use openly published research to do with what you want," says Mark Patterson, director of publishing at the open-access publisher Public Library of Science (PLOS). "Because CC licenses are created by experts and have a solid legal foundation, they have become the gold standard in open access publishing" (adapted from [The Power of Open](#) 2011, made available under a [Creative Commons Attribution License CC BY](#) licence).

Open Access:

If you are unfamiliar with or unsure of what Open Access means in relation to scholarly works, then watch Video 2. It's truly brilliant. It not only succinctly defines the issue, but also conveniently provides much of the historical background, politics, facts and figures.

VIDEO: http://www.youtube.com/watch?v=L5rVH1KGBCY&feature=player_embedded

VIDEO 2 — OPEN ACCESS EXPLAINED VOICED BY NICK SHOCKEY AND JONATHAN EISEN, AND ANIMATED BY JORGE CHAM IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE.

I find it hugely embarrassing to tell you that even academics at institutions with very good libraries don't have access to everything they need. Academic mailing lists such as [DML](#) are awash with requests from researchers looking for PDF files of academic papers, as are social-media networks such as Facebook and [Twitter](#). Most of the time, these requests aren't for 'difficult to obtain' ancient literature, but rather for twentieth or twenty-first century papers that have to be paid for. I won't name names, but I vividly remember an example on Twitter recently, when a researcher at the University of Oxford asked if anyone had access to a particular paper. I, a researcher at the University of Bath, tried and failed; a researcher from the University of Cambridge tried and failed; so did a researcher from Bristol University. Only many hours later did a researcher from Royal Holloway, University of London, manage to get access to this cursed paper. Access to scientific literature is a hugely inefficient 'postcode lottery' and disadvantages academics at less wealthy institutions.

No wonder, then, that academics who choose to make their works Open Access receive more views, downloads and citations than those who publish in subscription-only journals, as has been shown in many different peer-reviewed studies (see suggestions for further reading).

Despite this, the majority of academic research is still made available on the Internet only behind a paywall. This applies as much to palaeontology as to any other discipline. For every Open Access palaeontology-publishing journal such as [Acta Palaeontologia Polonica](#), [Palaeontologia Electronica](#), [PLoS ONE](#) and [ZooKeys](#), there are scores more profit-making subscription-access journals.

Much of the research in these journals is ultimately paid for by the public, through tax-fuelled government funding or charities. Yet for decades, subscription-access publishers have been monopolizing this public resource and charging increasingly exorbitant fees to access it (see Video 3).

VIDEO: http://www.youtube.com/watch?v=GMIY_4t-DR0&feature=player_embedded

VIDEO 3 — THIS VIDEO SARCASTICALLY EXPLAINS THE BIND RESEARCHERS PUT THEMSELVES IN WHEN THEY SUBMIT THEIR WORK FOR PUBLISHING IN A FOR-PROFIT SUBSCRIPTION-ACCESS JOURNAL. IT IS BY ALEX HOLCOMBE AND IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE.

With Open Access journals, everyone can gain the benefit and pleasure of reading every last word for free, forever. With subscription-access journals, unless you have a paid subscription or are affiliated with an institution that subscribes, it is likely that you will be denied access to most articles. But Open Access isn't just about reading; crucially, it also enables royalty-free reuse. Teachers and

educators can print off all or portions of as many Open Access papers as they wish without incurring legal costs. Subscription-access science effectively bars itself from classrooms and other educational settings — copy a 2-page paper for 50 students to read and it could cost you US\$4,905! Even single-access, single-paper charges can be outrageous: [who would pay US\\$113 plus tax to read a 5-page paper?](#)

The situation in biomedical research is seriously unethical, and has had [documented negative impacts on public health](#), although in recent years levels of access have improved. I would argue that the prevalence of paywall-restricted access also harms palaeontology.

It is hugely naive to assume that only academics benefit from access to academic works (as documented at [WhoNeedsAccess?](#)). Palaeontology has a long and undeniable history of ‘amateur academics’ making brilliant scientific contributions. For example, the nineteenth-century fossil collector [Mary Anning](#) not only found many outstanding fossils, but also contributed to the understanding of their anatomy: she was the first person to correctly interpret some of the strange objects that she had found as [coprolites](#) (see Fig. 2). There are thousands, if not millions, of potential amateur palaeontologists in the world, whose enthusiasm lies untapped and whose knowledge remains underdeveloped — and the situation is exacerbated because most of the world’s scientific literature is hidden behind paywalls on the Internet.



FIGURE 2 — MARY ANNING KNEW WHAT THESE WERE: FOSSIL FAECES! © ELECTROBLEME.

Examples of valuable ‘amateur’ contributions to science enabled and emboldened by Open Access include school students such as Jack Andraaka, who this year [discovered a novel method of cancer](#)

[detection](#) aided in part by access “to free online scientific research”. I believe that we would have more Andrakas in this world if cutting-edge science wasn’t so actively restricted.

A few academics (mostly more senior ones in my experience) seem fearful of Open Access; it represents a change in the way things are done. Academia is notorious for its conservatism. Many years ago, there were all sorts of strange criticisms levelled at Open Access — for example, that peer-review standards are lower in Open Access journals — but most if not all of these have been proven wrong over time. I refer interested readers to the Wikipedia section on that topic for more. It is of historical interest only.

There are many different and valid arguments for Open Access; in my opinion, [the arguments work best when combined](#). Open Access provides many benefits so I am overjoyed that the United Kingdom has [mandated](#) that from 1 April 2013, all its publicly funded research output must be made Open Access. There are concerns over the cost, but when weighed against the global benefits, I am sure that it will be worth it. I could write my entire article on Open Access alone — some people have written whole books on the subject — but I think that it is worth moving on to other areas of digital openness in academia.

Open Data:

This year, the Open Knowledge Foundation granted me a Panton Fellowship to promote the [Open Data movement](#), which is starting to gain momentum in science – as good scientific practice requires independent analysis and access to data. The Panton Principles were launched in 2010 as a guide for how scientists should publish their data online (see Video 4).

VIDEO: http://www.youtube.com/watch?v=JmN7gTGkJwA&feature=player_embedded

VIDEO 4 — INTRODUCTION TO THE PANTON PRINCIPLES © SOPHIE KERSHAW, ALASTAIR KAY, ROSS MOUNCE (2012), LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION LICENCE CC-BY-3.0. PANTON PRINCIPLES BY PETER MURRAY RUST, CAMERON NEYLON, RUFUS POLLOCK, JOHN WILBANKS (2010-02-09).

Open principles are even more important for data than for publication of results, because one of the major values of data is in its reuse, not mere presentation. The permission to reuse, reanalyse, resample and remix scientific data is vital to science. The trouble is that when research data does get shared online alongside an associated academic article, it is often in rather rudimentary ‘supplementary information’ files, which are not fit for purpose. They are often in formats such as Word documents and PDFs, which can’t be used for reanalysis: no statistical analysis software ever created accepts PDF files as an input, yet for the convenience of publishers this is how the supporting data for innumerable papers is made available online.

This regrettable oversight (which, to give them their due, many publishers are now fully aware of and working to change) creates a barrier to reuse that can prevent independent verification of the results and analyses presented in scientific papers, or make them much harder. Much of my PhD research involves trying to reuse and reanalyse previously published data in novel ways to gain new insight about the evolution of fossil animal groups. I have learnt from painful experience the inefficiencies how research data has been made available. I have tried to [raise awareness](#) of the

problems and have helped to research and write papers [encouraging best practice](#), but there is much still to do.

Independent verification is important: it can uncover major problems with the original analysis. The peer-reviewed scientific literature is almost certainly hiding [tens of thousands of errors and examples of misconduct](#).

Considerable hard data have emerged on the scale of misconduct. A metastudy (D. Fanelli *PLoS ONE* 4, e5738; 2009) and a detailed screening of all images in papers accepted by *The Journal of Cell Biology* (M. Rossner *The Scientist* 20 (3), 24; 2006) each suggest that roughly 1% of published papers are fraudulent. That would be about 20,000 papers worldwide each year.

– Colin Macilwain, *Nature*

QUOTE FROM COLIN MACILWAIN PUBLISHED IN NATURE

Making data more immediately reusable and openly licensed would improve transparency and remove some of the barriers to independent reanalysis, for the greater good of science. At least one major publisher is now proposing [Open Data as the default](#) standard for data supporting a scientific publication. Exceptions from the Open norm could of course be made for justifiably sensitive data such as patient medical data, or information on the locations of rare animals, plants or fossils but the point is – these have to be justified as sensitive before they can be excepted from the (Open) norm.

Most [reputable evolutionary-biology journals](#) now require that the data behind research papers be archived on the Internet in repositories such as [Dryad](#), [MorphoBank](#) and [Figshare](#), to ensure unhindered and quick access for all readers. Palaeontology journals, on the whole, haven't yet made this change, although *the Journal of Vertebrate Paleontology* did say in a 2011 editorial that it was "**considering** following other key journals in instituting a policy requiring that data supporting results presented in a publication be archived in a public repository".

One might think that by politely emailing the authors of studies, one would be able to gain access to their research data. Alas, in [my own](#) and [others'](#) experience, this system doesn't always work well. Authors can be away on holiday, in the field or just plain unwilling to share (incidentally, [willingness to share data is correlated with the quality of the results](#) - 'dodgy' or 'weak' data are less likely to be shared), and email addresses change as authors move between institutions, making it hard to even contact them in the first place. Of course, authors can even be deceased — sometimes making it impossible to obtain the underlying research data from a published study. Emailing the author takes away precious time from both parties, and doesn't work well if one is performing a meta-

analysis and needs data from many different studies. For all these reasons and more, it is far more transparent, efficient and sensible to share research data as Open Data in a public archive for the benefit of everyone.

Open Source code:

Research presented in a publication can be broken down into three parts: the paper describing the work done; the data; and the statistical methods and manipulations applied to the data to provide the results and the basis of the conclusion. I have already talked about Open Access in relation to the first part and Open Data in relation to the second, but this third part is often the most perniciously non-Open.

Modern research is often very complex — many of the simple observations and discoveries have been made already. We need to use complex techniques and methods to make further groundbreaking insights. This is often done by computer, either in generalist computing environments such as [R](#) or [MATLAB](#), or in specialist programs such as [TNT](#), [PAUP*](#) or [SPIERS](#). Yet many of these programs are not Open Source — it is not possible to check what the code that runs them is doing, so scientists have to take it on trust (sometimes with erroneous results).

Science Code Manifesto

SCIENCE CODE MANIFESTO BY PALPHY

Thus, in 2011, the [Science Code Manifesto](#) was released. It notes:

“Software is a cornerstone of science. Without software, twenty-first century science would be impossible. Without better software, science cannot progress.

But the culture and institutions of science have not yet adjusted to this reality. We need to reform them to address this challenge, by adopting these five principles:

- **Code** All source code written specifically to process data for a published paper must be available to the reviewers and readers of the paper.
- **Copyright** The copyright ownership and license of any released source code must be clearly stated.
- **Citation** Researchers who use or adapt science source code in their research must credit the code’s creators in resulting publications.
- **Credit** Software contributions must be included in systems of scientific assessment, credit, and recognition.
- **Curation** Source code must remain available, linked to related materials, for the useful lifetime of the publication.”

To date, nearly 1,000 people have endorsed this manifesto and is still open for signatures. Regardless of numbers, the principles of this manifesto are vitally important to help to avoid [controversies](#) around the public perception and transparency of science.

Open Educational Resources and MOOCs:

In this last section, I would like to mention perhaps the most exciting impact of openness and the Internet: the ability to educate humanity *en masse* for a mere fraction of the cost of ‘traditional’ educational courses. Open educational resources (OERs) are freely accessible, openly formatted and openly licensed documents and media that are useful for teaching, learning, education, assessment and research purposes (see Video 5).

VIDEO: http://www.youtube.com/watch?feature=player_embedded&v=-xGRztrWv-k

VIDEO 5 — THE OERS — OPEN EDUCATIONAL RESOURCES BY INTHEACADEMIA IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE.

Aggregator sites such as [OERcommons](#) list more than 50,000 such resources, of which more than 20,000 are ‘Science & Technology’ related. This corpus is growing day by day. As well as classifying content by subject, it is also lists content by school age suitability (Primary, Secondary, Post-secondary) and type (Audio Lecture, Full Course, Lecture Notes, Activities). Examples of high-quality openly available content can be seen below.



UK universities are getting in on the act, with OER repositories of Open content now available online from the University of [Oxford](#), the University of [Leicester](#), [University College London](#), the University of [Bath](#), the University of [Nottingham](#) and more.

Audio and video resources and lecture slides are fantastic rich content to share openly. It takes a long time to prepare a good, high-quality lecture course, and sharing it online for others to use, adapt and remix could result in a lot of time saved. In addition, it could ensure that top-quality material can be delivered where it is needed most, for example to less-privileged institutions, or even at home as part of Open University-style distance-learning courses.

To put the emerging innovation of free, open online courses in a financial context:

Americans collectively owe US\$914 billion in student loans (source: [Federal Reserve Bank of New York](#)). Average tuition costs alone are \$28,500 per year. Despite this, in autumn 2012, a record 21.6 million students are estimated to be attending US colleges and universities, an increase of about 6.2 million since autumn 2000.

In the United Kingdom, many fees for domestic undergraduate courses are now set at £9,000 per year. At the moment, UK universities and colleges have only [2.5 million](#) enrolled students, and these numbers seem to be relatively stable.

So there is serious money to be made in the higher-education market. OER-style online delivery of undergraduate courses has been touted as both a disruptive opportunity and a [threat to some brick and mortar universities](#).

Online courses called MOOCs (massively online open courses) are attracting the interest of both for-profit and non-profit groups. Many aren't strictly Open, but they do tend to have at least partly free content and do not require enrolment at a traditional university. Examples are [Udacity](#) (for-profit), [Udemy](#) (for-profit), [Coursera](#) (for-profit), [EdX](#) (non-profit) and [Peer to Peer University](#) (P2PU, non-profit).

For the moment, this content tends to be restricted to computer-heavy subjects such as maths, physics and computer science, which lend themselves to online teaching. However, there are courses coming online for a broader variety of subjects every week. Coursera has more than 1.9 million enrolled students this year, although this figure cannot be directly compared with traditional courses because MOOCs have high dropout and non-completion rates, with people tending to dip in and out quite freely. Traditional courses, with their high fees, are less lightly started and discarded!

With respect to palaeontology, it is obvious that online-only delivery couldn't provide everything required — field trips and study of specimens are conceivably much richer experiences in person than in virtual learning environments. However, I think that before too long, we will see a full MOOC offered for highly theoretical palaeontological subjects such as [phylogenetics](#) and statistical analyses. The American Museum of Natural History in New York already offers [MOOC-like courses](#), but they are not free – \$495 + \$25 registration fee if you're interested, hence they're neither MOOCs nor OERs.

Conclusions:

The Internet has forever changed the way academia operates — even palaeontology!

Open Access has been mandated in the United Kingdom, with Europe soon to follow (not to mention that Latin America already largely embraces Open Access, and that Australia is getting there too).

Open Data is beginning to be recognized as similarly important, with many funders such as the US National Science Foundation, the UK Biotechnology and Biological Sciences Research Council and the UK Natural Environment Research Council now actively tightening up their data-sharing rules.

Open-source code and fully reproducible science are next on the agenda for wide-scale change going into 2013.

Huge investment is currently being made to further develop OERs and MOOCs to enable higher education online for everyone. Very early days here, but keep watching.

My take on it is that we should all be grateful to Creative Commons for legally enabling free and open enterprise without the hassle of restrictive copyright terms that would prevent much societal good. The organization is especially brilliant considering its small size — even now, at perhaps its largest, it has only 21 full-time staff members. The power of Open should not be underestimated!

Suggestions for further reading:

On Creative Commons:

Creative Commons is [nearing its tenth birthday](#) (7–16 December) and has just elected a new [scientific advisory board](#). This will ensure that it continues to serve and enable the needs of science as best it can.

On Open Access:

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T. & Harnad, S. 2010. Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLoS ONE* 5: e13636. doi:[10.1371/journal.pone.0013636](https://doi.org/10.1371/journal.pone.0013636)

Mike Taylor has an excellent set of tutorials explaining terms and concepts in Open Access starting [here](#).

On Open Data:

The Open Knowledge Foundation has an excellent [post](#) that explains Open Data in the wider context, it's much bigger than just academia – financial, transport, environment, weather and cultural data too.

On Open Source Software:

Ince, D. C., Hatton, L., and Graham-Cumming, J. 2012. The case for open computer programs. *Nature* 482: 485–488. doi:[10.1038/nature10836](https://doi.org/10.1038/nature10836) — *If you can't access this paper, [please drop me an email](#).*

¹ PhD Candidate and [Open Knowledge Foundation](#) Panton Fellow | Biology and Biochemistry Department, University of Bath, Claverton Down, Bath, BA2 7PY, UK.