

A life online

Darwin is the latest eminent scientist to get an online archive. How do these undertakings change our understanding of history, asks Henry Nicholls.

When exactly did Charles Darwin first use the iconic and ideology-laden phrase "survival of the fittest"? On page six of the first issue of the first volume of the first edition of *The Variation of Animals and Plants Under Domestication*, published in 1868. The phrase only makes it into *On the Origin of Species* in that book's fifth edition, published the following year.

If you are James Moore, or Janet Browne, or Adrian Desmond, or David Quammen, or any of the other people who have written biographies of Darwin and analyses of his thought, you probably already knew that. Now, though, you don't need years of scholarship to dig out as many such intriguing nuggets as you can imagine. With the launch on 19 October of Darwin Online¹, answering this kind of question becomes a doddle.

The service makes the complete works of Darwin — in the form of both scanned-in pages and searchable text — freely available to anyone with an Internet connection. It also contains a handful of his key unpublished manuscripts, such as the jottings he made during the voyage on *The Beagle*. This kind of resource, allowing in-depth interrogation of a scientist's entire body of work, is becoming increasingly common. Its proponents believe that it will do more than just cut down on bike rides, train trips or even flights to visit physical archives; they expect it to inject new life into historical scholarship.

The ambitious idea for Darwin Online came to John van Wyhe, a historian of science at the University of Cambridge, UK, back in 2002. Although there were plenty of Darwin's works on the web, he recalls, they were spread across many sites with no obvious editorial standards. "It was just utter chaos." Given the many other great men of science (for they are almost exclusively men) who boast digital archives of their work and correspondence, this was a gap that needed to be filled (see 'Best of the rest', opposite). "It just didn't seem right that other figures should have these rich databases and not Darwin," says van Wyhe.

Van Wyhe aims to add every edition and translation of Darwin's published work to the website by 2009, the bicentenary of his birth and the 150th anniversary of the publication of *On the Origin*. The ultimate plan is to scan in and transcribe a mountain of some 35,000 Darwin manuscripts dispersed around the world, including drafts of *On the Origin*, descriptions of his experiments and even his domestic

accounts. The aim is to host everything Darwin ever wrote except for his private letters, which the long-running Darwin Correspondence Project, also in Cambridge, is dealing with on its own website².

By the standards of scholarship in the humanities, as opposed to those of the well-equipped lab, these projects do not come cheap. Van Wyhe has a grant of £286,000 (US\$532,000) from Britain's Arts and Humanities Research Council (AHRC), although he will need more to finish off the project as he wants it. The Newton Project, which aims to be the definitive, open-access repository for Isaac Newton's writing and correspondence, has received grants from the AHRC that total £900,000. This funding should be enough to transcribe, footnote and publish online his vast collection of theological writings (almost half of what he wrote) and more than half of his unpublished optical writings by 2010. As with Darwin Online, the long-term vision is to publish everything Newton ever wrote — although coding up the mathematics will be a particular challenge, notes Robert Iliffe, the historian from Imperial College London who acts as the project's editorial director.

Iliffe is adamant that the venture offers serious value for money. In the past, thousands of pounds would be thrown at producing a print edition of Newton's work, with only a few institutions prepared to stump up the cash to buy a copy. "The online medium brings with it such a huge audience," he says.

Heavy hitters

That's certainly the experience at the Einstein Archives Online³. This site currently boasts more than 3,000 scanned-in pages from almost 1,000 of Einstein's scientific and non-scientific works. Even though he wrote almost exclusively in German, there is huge interest in these documents, says Diana Buchwald, director of the Einstein Papers Project at the California Institute of Technology in Pasadena. When the Einstein Archives Online went live back in 2003, there were nearly 250,000 unique user sessions in the first five days, she says.

Darwin Online is ready for a similar surge of activity as both supporters and distorters of Darwin's ideas make use of the ability to search through the published material. There are several items that should be of particular interest to Darwin fans, notably the previously unpublished *aides-memoires* he made as *The Beagle* explored the Galapagos Islands. The original notebook has been lost, probably



stolen in 1983, but the University of Cambridge's library has a microfilm of it. It is also possible to inspect all his published illustrations; the visually impaired can download audio files of all the texts.

The creationist faithful would do well to take a look, says van Wyhe. "If people feel so strongly about Darwin, they should actually take the time to read his own words rather than relying only on the interpretations of others." Even if this doesn't convert them to evolution by natural selection, it should expose the popular misconception that Darwin had an anti-Christian agenda, he says. "This was not what he was about," says van Wyhe. "He was simply a scientist trying to explain how the world works."

The same cannot be said for the defiantly irreligious Francis Crick who, enraged by the decision of Churchill College, Cambridge, to build a chapel, wrote a letter to the college's namesake Winston enclosing £10 towards the building of a brothel to go with it. Unfortunately, this letter is in the Churchill archive rather than the Crick archive now being catalogued at the Wellcome Library in London. As



W. FERRELL/ISTOCK.COM; P. JACKMAN/ISTOCK.COM

Contemporary history raises other challenges, most notably that of copyright. In many instances, it's simply not clear who owns it. "You may own the physical letter, but the person that wrote it owns the copyright," says Peter Hirtle, intellectual-property officer for Cornell University Library in Ithaca, New York. In an archive with many correspondences, "you have potentially hundreds and hundreds of copyright owners". Consequently, publishing outgoing and not incoming letters is the only practical option for many correspondence projects.

But even if near-complete, searchable archives are to remain the privilege of the long dead, they still offer new scholarly possibilities. Last year, Albert-László Barabási, professor of physics at Harvard University's Dana-Faber Cancer Institute in Boston, and a colleague used data from the Darwin Correspondence Project and Einstein Papers Project to compare how long these men took to reply to incoming letters⁵. The ability to use the dates of documents to map a scientist's writing and correspondence patterns might give new insights into how he organized his life and work, says Barabási. As van Wyhe will be able to tell something about who is reading which of Darwin's texts online, research into the researchers will be possible, too, through studies of the way the resource is used. Will everyone look for survival of the fittest, or will people dig into ever more intriguing byways? What grandeur may be found in this new view of a life? ■

Henry Nicholls is a freelance science writer in London.

1. <http://darwin-online.org.uk>
2. www.lib.cam.ac.uk/Departments/Darwin/index.html
3. www.alberteinstein.info
4. Ridley, M. *Francis Crick: Discoverer of the Genetic Code* (HarperPress, 2006).
5. Oliveira, J. G. & Barabási, A.-L. *Nature* 437, 1251 (2005).

yet, only a sliver of the Crick archive is accessible online, by means of the website of the US National Library of Medicine. "It's a fantastic new tool not available to previous generations of researchers," says Matt Ridley, who has just published a biography of Crick⁴. "I longed for even more correspondence to be there."

Contemporary conundrums

But Crick and other contemporary giants raise archival problems that Darwin and Newton avoid. "Indexing becomes more difficult as the corpus gets larger," says Darwin Stapleton, executive director of the Rockefeller Archive Center in New York, whose father named him after the great nineteenth-century naturalist. The archive of a contemporary scientist such as Crick's colleague James Watson may contain more than a million items — including e-mails, spreadsheets, slide presentations, audio recordings and TV appearances, all of which present their own indexing challenges. "What can be done with the Darwin collection cannot be done with a twentieth-century figure," says Stapleton.

Best of the rest

In addition to the online archives for big hitters such as Darwin, Einstein and Newton, there are several other projects intent on revealing the writings of famous scientists.

The Robert Boyle Project (www.bbk.ac.uk/boyle) hosts about a fifth of Boyle's writings, while his 'work diaries' with annotated transcriptions benefit from virtual page-turning wizardry at www.livesandletters.ac.uk/wd.

Jean-Baptiste Lamarck's archive (www.lamarck.cnrs.fr) contains digital photographs of the great biologist's 19,000-strong herbarium collection as well as scanned-in pages from his manuscripts and transcriptions of much of his published work.

The Panopticon Lavoisier (<http://moro.imss.fi.it/lavoisier>) presents much of Antoine-Laurent Lavoisier's published and unpublished

chemical work, hundreds of photographs of his laboratory instruments and mineral collection, and a catalogue of his own library.

The Thomas A. Edison Papers Project (<http://edison.rutgers.edu>) holds scans of 180,000 digital images from the 5 million pages contained in the Edison archive.

The collection of Eva Helen and Linus Pauling papers (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/index.html>) contains some 50,000 items, of which around 100,000 have been digitized. Not all of these are available online.

The US National Library of Medicine (<http://profiles.nlm.nih.gov>) has a repository for a selection of works from twentieth-century leaders in biomedical research and public health, including Francis Crick, Oswald Avery, Joshua Lederberg and Barbara McClintock.