

**Institution: Durham University**

**Unit of Assessment: 31 (Classics)**

**Title of case study: Open Access to Ancient Greek and Latin through *Diogenes* (CS1)**

### 1. Summary of the impact

*Diogenes*, created solely by Peter **Heslin**, is a freely distributed, open-source programme which enables access to all the major databases of classical Greek and Latin texts that have been in public circulation since the mid 1980s. *Diogenes* has had a significant and lasting impact on the education and cultural life of many of its tens of thousands of users. Some of these are professional classicists, who utilize it for both research and teaching. But a much larger part of the user population consists of students and non-academic readers of ancient Greek and Latin. *Diogenes* makes available to them the whole corpus of classical literature in the original languages. It also provides integrated morphological tools and lexica to support the needs of both language learners and more advanced readers. *Diogenes* has also had a significant and enduring impact on the movement towards open access publishing of digital resources for classics worldwide.

### 2. Underpinning research

*Diogenes* is both the outcome of classical research and a classical research output in its own right:

**a) Research on Latin epic.** **Heslin** first became aware of the shortcomings of available digital resources while conducting research for his first monograph, *The Transvestite Achilles*. His research on Latin epic prompted the creation of *Diogenes*, and the book and the programme evolved in tandem. For example, a query using the archaic, and now obsolete, *Pandora* programme to find all places in Greek and Latin literature in which the island of Scyros is mentioned (a question fundamental to *The Transvestite Achilles*) produced results which were incorrect and incomplete. Further research showed that the error was due to improper omission of passages where the searched word was hyphenated in the printed editions on which *Pandora* depends. To give another example, **Heslin** needed to identify parallels for particular noun-adjective combinations, but found that this could not be done reliably by *Pandora*.

**b) *Diogenes* as research.** *Diogenes* itself is a work of original research, and was submitted as such in RAE 2001. The databases of classical Greek and Latin texts, which were first developed in the 1970s on mainframe computers, use an idiosyncratic, ad-hoc encoding which involves extensive manipulation of individual ones and zeroes, rather than, say, self-documenting tags, as in XML. In the absence of any documentation, the only way to establish the significance of the stream of apparently meaningless numbers by which the databases encode the metadata of the classical texts (e.g. book and line numbers, lemmata, complex page layout of scholia) was by a process of reverse-engineering. This entailed reading the opaque computer code of the databases in order to identify the features of the classical text which were embedded there. This could only be done by a researcher equally versed in binary data encoding and in the conventions of the most challenging of printed classical texts, from documentary papyri to scholia. Furthermore, integrating the *Perseus* morphological and lexical tool into *Diogenes* required the skills of a software developer who was intimately familiar with the quirks of Greek and Latin morphology and lexicography. This research is expressed and embodied in the computer code and internal documentation of *Diogenes*, which is downloaded by every user of the programme.

Initial work on reverse-engineering the Latin and Greek databases began in 1999 and continued apace after **Heslin** was appointed at Durham in 2000 as a Lecturer in Classics. This initial phase of development, in which new features were gradually added as further aspects of the databases were decoded, continued until version 1.0 was released in 2003. The next phase of research and

development focused on finding an infrastructure that would enable easy installation by non-technical users on three different platforms: Windows, Mac and Linux. This arrived with the 3.0 release of *Diogenes* in 2007. The most recent phase involved the large task of integrating open-source morphological and lexical tools from *Perseus*; this was done later in 2007 and led to version 3.1.

**c) Dissemination of research insights.** As **Heslin** was developing *Diogenes*, he articulated and disseminated the research insights of his work in a series of publications. Reverse-engineering the specifics of version E of the *Thesaurus Linguae Graecae* (TLG) CD-ROM led to a substantial review article on that database. An equally substantial review article on an early version of the *Thesaurus Linguae Latinae* (TLL) CD-ROM was motivated by the urgency of preventing that project from making some of the same mistakes as those of other classical databases. A podcast (*Diogenes: Milestones and Morals*: <http://www.digitalclassicist.org/wip/wip2008-12ph.html>) outlines the history of the interaction between classical research and computer science, and maps out the future directions which research at Durham is pursuing to create more sophisticated and freely-distributable corpora of ancient texts.

### 3. References to the research

1. Peter **Heslin**, *The Transvestite Achilles: Gender and Genre in the Achilleid of Statius* (Cambridge University Press 2005).
2. Peter **Heslin**, *Diogenes* (1999-present). Source-code with documentation is included with the downloaded application, which is available via a link on the website: <http://www.dur.ac.uk/p.j.heslin/Software/Diogenes/>.
3. Peter **Heslin**, Review article on the *Thesaurus Linguae Latinae*, Third electronic edition. *Bryn Mawr Classical Review* (2006.02.19): <http://bmcr.brynmawr.edu/2006/2006-02-19.html>.
4. Peter **Heslin**, Review article on *Thesaurus Linguae Graecae*, CD-ROM Disk E. *Bryn Mawr Classical Review* (2001.09.23): <http://bmcr.brynmawr.edu/2001/2001-09-23.html>.

The quality of the research listed above is indicated by the peer-review processes which led to publication and submission to RAE 2001 (item 2) and RAE 2008 (items 1 and 4), and by positive reviews of item 1 in several leading journals in the field.

### 4. Details of the impact

The reach of *Diogenes* is best demonstrated by the download statistics. As of 31 July 2013, version 3.1.6 of *Diogenes* (released on 22 October 2007) has been downloaded from the official site 91,011 times. This is certainly a considerable under-estimate of actual use, because *Diogenes* is a redistributable open-source software. Users are encouraged to pass it on to others freely in accordance with the license (it is also included with several Linux distributions), so many users of *Diogenes* will not have obtained it from the official download site. The download statistics do, however, reflect one particular aspect of the impact population: its international character. The UK is merely in 7th place in the number of downloads, well behind USA, Spain, Italy, Mexico, Brazil and Greece, in that order. The user population of *Diogenes* is mixed. It includes many scholars; when the American Philological Association, the professional body of North American classicists, withdrew support for its obsolete *Pandora* programme, it officially endorsed *Diogenes* as the replacement in its Newsletter (August 2005). As one Royal Holloway classicist said on a public e-mail list in 2008, '*Diogenes* has emerged as far and away the best tool for the job while commercial and funded rivals have fallen away.' *Diogenes* is also used on a considerable scale by students at all levels, and by members of the non-academic public who read ancient Greek and Latin for their own interest and pleasure. It is not possible to quantify the relative size of these segments of the mixed user population, but it is clear that the order of magnitude of downloads, not to mention free distributions, dwarfs the population of professional classical researchers worldwide.

The significance of *Diogenes* is threefold:

**(a) Free access to classical texts.** For students and members of the public without access to a university library or without an institutional subscription to expensive on-line resources, *Diogenes*

## Impact case study (REF3b)

provides free access to the massive extant corpus of ancient Greek and Latin literature which was encoded on now-archaic databases that continue to circulate widely via person-to-person copying and peer-to-peer network file sharing. Unsolicited e-mails indicate that *Diogenes* is used by a variety of users who may have no other access to Greek and Latin texts: '*Diogenes* I use outside of any formal educational setting--it's just for my own use. It is extremely useful for reading Latin texts: to determine quickly word meanings, parse forms, or check syllable length. I'm taking up Greek (again), and it will surely be just as useful for that, too' (9/9/2012). '[*Diogenes*] has been very helpful with some words that are hard to find in the Greek lexicons and dictionaries that are available to someone who is not a scholar or a theological student. I've been teaching myself ancient Greek as I translate the New Testament' (20/4/2012). 'I have been extensively using *Diogenes* for 6 years, all through my undergraduate career at the University of Chicago, into ... grad school ..., and now for private study outside of any curriculum. I firmly believe it is the best tool ... for studying Greek ... I cannot understate the value of such a resource' (11/3/2013).

*Diogenes* is widely used in countries where libraries may be poorly stocked, and where classical editions and lexica are prohibitively expensive. It is notable that the top ten countries for downloads include Mexico (4th place), Brazil (5th), Columbia (9th), and Argentina (10th), ahead of European countries with more developed traditions of studying the classics, such as France (11th). Many users from developing countries have expressed their gratitude: '*Diogenes* ... makes my Greek readings a lot easier ... I study and teach classical Greek here in Porto Alegre, Brazil' (24/10/2009). From Poland: '*Diogenes* has been my one and only tool for working on classical texts ever since I first installed it ... I find it truly amazing' (11/03/2009).

**(b) Morphology and lexica for language learners.** The morphological analysis and lexical look-up tools in *Diogenes* support active, self-directed, and independent engagement with the ancient languages on the part of secondary school students, undergraduates, postgraduates and non-academic users. Clicking on a word in *Diogenes* instantly generates a morphological analysis from *Perseus*, and corresponding definitions from the standard Greek and Latin lexica of Liddell-Scott-Jones and Lewis-Short. This makes it much easier to look up a word in these advanced tools than in even the smallest beginner's dictionary. *Diogenes* therefore helps students to appreciate that translating words from Latin and Greek is not a simple matter of one-to-one mapping, but of negotiating complex, non-congruent semantic boundaries. Many language teachers use *Diogenes* for the very easy and intuitive way in which it integrates advanced lexicographical tools. A language teaching officer in Classics at Cambridge University writes: 'At my induction meetings I recommend that all my students – whether undergraduate or graduate – download and use the programme as a matter of course; they routinely report: "it changed my life!" – as it has mine. I imagine that a large proportion of our student body here uses *Diogenes* regularly' (5/10/2012). A professor in the USA writes: '*Diogenes* is ... open all the time—it's a wonderful teaching tool and I have all my students use it' (16/10/2011). Not only beginners, but students at all levels around the world use *Diogenes*: 'I'm a student from Catania University ... and I would thank you for developing this fantastic open source program' (24/4/2009). '[*Diogenes* is] a great help for my studies in Classics at Munich university (LMU)!' (25/10/2011). 'I'm a young Portuguese classicist, and just wish to say thank you for your kind contribution to knowledge. You developed a wonderful tool that really helps those [who] seek to better understand the ancient world and its literature.' (28/05/2009). Self-directed learners of Latin and Greek rely on *Diogenes* for morphological tools and lexica. Hundreds of postings to *TextKit*, a public internet forum frequented by independent learners of Greek and Latin, recommend it to newcomers. Some of these users even adapt it for their own purposes, for example, setting up the browser Firefox so that a click will command *Diogenes* to parse and define a Greek or Latin word found on any website on the internet. Another user, who describes himself as 'a professional computer geek who has been interested in classical languages his entire life' has incorporated *Diogenes* into the website he created with a selection of resources for 'self-teaching amateur classicists'. Students and teachers of Latin and Greek at all levels rely on *Diogenes*.

**(c) Influencing open access in digital classics.** *Diogenes* was the first large open-source project in Classics, and other projects have followed its example; this sharing of resources has benefitted the common good. It brought to light the undocumented structure of the CD-ROM databases published by the Packard Humanities Institute (PHI) and the *Thesaurus Linguae Graecae* (TLG). *Diogenes* pioneered the advanced functionality which is now also available via the on-line TLG,

**Impact case study (REF3b)**

such as integrated morphological analysis with lexicon look-up and morphologically aware searching. This means that even those who access classical texts through other digital interfaces have benefitted from the example set by *Diogenes*.

Of greater lasting significance, however, is the current movement to create and release new digital resources under open access licenses which will ensure that they remain perpetually free. The *Perseus* project at Tufts University followed *Diogenes*' example when it made its texts and tools available under an open access license. This in turn permitted *Diogenes* to integrate *Perseus*' morphological analysis data and digitized lexica. *PhiloLogic*, at the University of Chicago, has combined texts from *Perseus* and re-used some source code from *Diogenes*. More recently, Harvard's Center for Hellenic Studies (CHS) has launched several open access initiatives, in which they have acknowledged *Diogenes* as the trailblazing model. The director of CHS writes: '*Diogenes* has embraced the concept of open access for a tool that is exceptionally important and effective for students at all levels of competence in Ancient Greek, and it has done so in a way that has and will continue to inspire others to do the same.' On a more practical note, one of the collaborators on digital projects at the CHS writes: 'without your open-source tools, I can honestly say that important parts of the *Homer Multitext* project would probably be a decade behind where we are now.' *Diogenes* has thus provided both a model for other projects of the principles of open access for digital classics, and freely available tools which those projects have been able to use and build upon. In June 2013 **Heslin** built an extension to *Diogenes* to bring the CD-ROM databases into the modern era by converting them to an easily understandable XML-based format. The DigiLibLT project in Vercelli, Italy, intends to distribute the PHI database of Latin texts, after being converted to XML by *Diogenes*, from their website under an open access, Creative-Commons license. For the ways in which *Diogenes* has been influencing other open access digital projects at Durham, see REF 3a.

**5. Sources to corroborate the impact**

1. Download figures: the *Diogenes* website is hosted by Durham University, but the software itself is distributed via SourceForge, which is a very large website catering for the distribution of open-source software. The externally audited figure of 91,011 given above for the number of downloads for version 3.1.6 was taken from the publicly accessible SourceForge project page for *Diogenes*: 69,602 (Windows) + 16,972 (Mac) + 4,437 (Linux) = 91,011. See <http://sourceforge.net/projects/diogenes/files/diogenes/3.1.6/> The nation-by-nation download data is publicly available on that same site.
2. The quotations from individuals are from private e-mails to Heslin, which are available in the audit file.
3. The quotation from the Royal Holloway classicist is from a public message sent to a Classics e-mail list: <http://lsv.uky.edu/scripts/wa.exe?A2=ind0810e&L=classics-l&T=0&F=&S=&P=60>.
4. The TextKit forum for independent language learners can be found at <http://www.textkit.com>; a Google search on the TextKit forum for posts mentioning 'Diogenes' gives approximately 1,170 results (31 July 2013). For the person who integrated *Diogenes* with Firefox, see <http://www.textkit.com/greek-latin-forum/viewtopic.php?f=2&t=9764#p75516>.
5. For the integration of *Diogenes* into a public website with resources for independent language learners, see <http://aoidoi.org/diogenes/>.
6. For the APA's announcement of the obsolescence of Pandora and recommendation of *Diogenes* as the alternative in its Newsletter (August 2005, vol. 28:4, p. 9), see [http://apaclassics.org/images/uploads/documents/newsletters/August\\_2005.pdf](http://apaclassics.org/images/uploads/documents/newsletters/August_2005.pdf).