

Determining Textbook Cost, Formats, and Licensing with Google Books API: A Case Study from an Open Textbook Project

Eamon Costello,
Richard Bolger,
Tiziana Soverino,
and Mark Brown

ABSTRACT

The rising cost of textbooks for students has been highlighted as a major concern in higher education, particularly in the US and Canada. Less has been reported, however, about the costs of textbooks outside of North America, including in Europe. We address this gap in the knowledge through a case study of one Irish higher education institution, focusing on the cost, accessibility, and licensing of textbooks. We report here on an investigation of textbook prices drawing from an official college course catalog containing several thousand books. We detail how we sought to determine metadata of these books including: the formats they are available in, whether they are in the public domain, and the retail prices. We explain how we used methods to automatically determine textbook costs using Google Books API and make our code and dataset publicly available.

INTRODUCTION

The cost of textbooks is a hot topic for higher education. It has been reported that by 2014 the average student spent \$1,200 annually on textbooks.¹ Another study claimed that between 2006 and 2016 the costs of college textbooks increased over four times the cost of inflation.² Despite this rise in textbook costs, a survey of more than 3,000 US faculty members (“The Babson Survey”) found that almost every course (98 percent) mandated a textbook or related study resources.³ One response to the challenge of rising textbook costs is open textbooks. Open textbooks are a type of open educational resource (OER). OERs have been defined as “teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and repurposing by others. Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other tools, materials, or techniques used to support access to knowledge.”⁴ OERs stem from the principle that access to education is a human right and that, as such, education should be accessible to all.⁵ Hence an open textbook is made available under terms which grant legal rights to the public, not only to use, but also to adapt and redistribute. Creative Commons licensing is the most prevalent and well-developed intellectual property licensing tool for this purpose.

Open textbook projects aimed at promoting publishing and redistributing open textbooks, both in digital and print formats, have been growing. For example, the BCampus project in Canada began in 2012 with the aim of creating a collection of open textbooks aligned with the most popular subject areas in British Columbia.⁶ The project has shown strong growth, with over 230 open digital textbooks now available and more than forty institutions involved. A significant recent

Eamon Costello (eamon.costello@dcu.ie) is Assistant Professor, Open Education at Dublin City University. **Richard Bolger** (richard.bolger@dcu.ie) is Lecturer at Dublin City University. **Tiziana Soverino** (tiziana.soverino@dcu.edu) is Researcher at Dublin City University. **Mark Brown** (mark.brown@dcu.ie) is Full Professor of Digital Learning, Dublin City University.



development in open textbooks occurred in March 2018, when the US Congress announced a \$5 million investment in an open textbook initiative.⁷

In addition to helping change institutional culture, and challenge attitudes to traditional publishing models, one of the most oft-cited benefits of open textbooks is cost savings. According to the College Board's Survey of Colleges, the average annual cost to US undergraduate students in 2017 for textbooks and materials was estimated at \$1,250.⁸ This figure is remarkably close to the aforementioned figure of \$1,200 a year, as reported by Baglione and Sullivan. However, there is little known about the monetary face value of books that students are expected to buy, beyond studies based on self-reported data. Students themselves in the US have attempted to at least open the debate in this area by highlighting book price disparities.⁹ Nonetheless, they only report on a very small number of books, and the College Board representing on-campus US textbook retailers have disputed their results for this reason, claiming that they have been selective in the book prices they have chosen.

Hence this study seeks to address the gap that exists in knowledge about the true cost of textbooks in higher education. This is in the context of a wider research project we are conducting on open textbooks in Ireland.¹⁰ Determining the cost of books is not straightforward as books can be new, used, rental, or digital subscription. However, the cost of new books does set a baseline for other forms, particularly rental and used books. Our aim here is hence to start with new books, by analyzing costs of all the required and recommended textbooks of one higher education institution (HEI) in Ireland.

The overarching research question this study sought to address is: What is known about the currently assigned textbooks in an Irish university? The sub-questions were:

- RQ1: What is the extent of textbooks that are required reading?
- RQ2: What are the retail costs of textbooks?
- RQ3: Are textbooks available in digital or e-book form?
- RQ4: Are textbooks available in the public domain?

The next section outlines our methodology and how we sought to find answers to these questions.

METHODS

In this section we describe our approach, the dataset generated, and the methods we used to analyze the data. We identified a suitable data source comprising the official course catalog of a HEI in Ireland with more than ten thousand students. In the course catalog faculty give required and recommended textbook details for all courses. This information is freely accessible on the website of the HEI; the course catalog is powered by a software system known as Akari (<http://www.akarisoftware.com/>). Akari is a proprietary software system used by several HEIs in and outside Ireland to create and manage academic course catalogs. The course team gained access to a download of all books recorded in the database of the course catalog (Figure 1). In this catalog, fields are provided for lecturers to input information for students about books such as title, International Standard Book Number (ISBN), author, and publisher. Following manual and automated data cleansing, 3,014 unique records of books were created. Due to the large number of books, at this stage we sought a programmatic solution for finding out more information about these books.

Module Details		Module Resources	
Title:	Interaction Design APPROVED	Essential Book Resources	
Long Title:	Interaction Design	Bryce L. Allen 1996, <i>Information tasks</i> , Academic Press San Diego [ISBN: 0120510405] written and edited by Stuart K. Card, Jock D. Mackinlay, Ben Shneiderman 1999, <i>Readings in information visualization</i> , Morgan Kaufmann Publishers San Francisco, Calif. [ISBN: 1558605339] edited by Usama M. Fayyad ... [et al.], <i>Advances in knowledge discovery and data mining</i> [ISBN: 0262560976] Jennifer Fleming, 1999, <i>Web Navigation: Designing the User Experience</i> [ISBN: 1565923510] edited by Christian Gram and Gilbert Cocton [sic] 1996, <i>Design principles for interactive software</i> , Chapman & Hall London [ISBN: 0412724707] edited by Robert Jacobson 1999, <i>Information design</i> , MIT Press Cambridge, Mass. [ISBN: 0262600358] Nicholas V. Iuppa 2001, <i>Interactive design for new media and the Web</i> , Focal press Boston [ISBN: 0240804147] edited by Morten Kyng and Lars Mathiassen 1997, <i>Computers and design in context</i> , MIT Press Cambridge, Mass. [ISBN: 9780262112239] Kevin Mullet, Darrell Sano 1995, <i>Designing visual interfaces</i> , SunSoft Press Englewood Cliffs, NJ [ISBN: 0133033899] Donald A. Norman 2002, <i>The design of everyday things</i> , Basic Books New York [ISBN: 0465067107] Peter Morville and Louis Rosenfeld 2007, <i>Information architecture for the World Wide Web</i> , O'Reilly Sebastopol, CA [ISBN: 0596527349] Ben Schneiderman 2009, <i>Designing the User Interface: Strategies for Effective HumanComputer Interaction</i> [ISBN: 9780321601483] Valery I. Frants, Jacob Shapiro, Vladimir G. Voiskunskii 1997, <i>Automated information retrieval</i> [ISBN: 0122661702] Karen Sparck Jones (Editor), Peter Willett (Editor) 1997, <i>Readings in Information Retrieval</i> [ISBN: 1558604545] Rory Stuart 1996, <i>The design of virtual environments</i> , McGraw-Hill New York [ISBN: 0070632995] Edward R. Tufte 1997, <i>Visual explanations</i> , Graphics Press Cheshire, Conn. [ISBN: 0961392126] Tay Vaughan 2008, <i>Multimedia</i> , McGraw-Hill New York [ISBN: 0072264519] Susan Weinschenk, Pamela Jamar, Sarah C. Yeo 1997, <i>GUI design essentials</i> [ISBN: 0471175498] Richard Saul Wurman, Peter Bradford (Editor) 1997, <i>Information Architects</i> [ISBN: 978-0823064557]	
Module Code:		This module does not have any article/paper resources	
Credits:	10	This module does not have any other resources	
NFQ Level:	8		
Field of Study:	Multimedia		
Valid From:	2017/18 (Sep 2017)		
Module Delivered In	1 programme(s)		
Administrator:			
Module Coordinator:			
Module Department:	60 - COMMUNICATIONS		
Module Description:	Interdisciplinary in content and structure, the module provides students with an understanding of the principles and practice of organizing digital information using graphic user interface design.		
Learning Outcomes			
On successful completion of this module the learner will be able to:			
L01	Demonstrate knowledge and critical evaluations of information design and human centred design		
L02	Recognise the significance of the graphical in information design; paradigms and genre as structuring devices; navigation and the simplification of complex structures; site architecture, wayfinding and mapping.		

Figure 1. Course Catalog Screenshot.

We initially thought that ISBNs might prove the best way to accurately reconcile records of books. However, many ISBNs were incomplete or mistyped. Moreover, many instructors simply did not enter an ISBN. Given the capacity for errors in the data—for instance, some lectures simply entered “I will tell you in class” in the book title field—we required a tool that could handle fuzzy search queries, e.g. cases where a book title or author were misspelled. The tool we selected was the Google Books Application Programming Interface (API).¹¹ This API provides an interface to the Google Books database of circa thirty million books. The service, like the main Google search engine, is forgiving of queries that are mistyped or misspelled. Hence, we constructed a query based on a combination of author name, book title, and publisher. Following experimentation, we determined that these three search terms together allowed us to find books with a high degree of accuracy whilst also accounting for possible spelling errors.





Figure 2. System Design.

We then wrote a custom JavaScript middleware program deployed in the Google Cloud Platform. This program parsed the file of the book search queries, passed them to the Google Books API as search requests and saved the results. The API returned results in JavaScript Object Notation (JSON) format. JSON is a modern web language for describing data. It is related to JavaScript and can be used to translate objects in the JavaScript programming language into textual strings. It is used as a replacement for XML as it is arguably more human readable and is considerably less verbose. We then imported this JSON into a MongoDB database to filter and clean the data, before finally exporting them to Excel for statistical analysis. MongoDB is a document store database that natively stores objects in the JSON format and allows for efficient querying of the data.

The Google Books API provides some key metadata on books aside from the usual author, publisher, ISBN, edition, pages, etc. as it gives prices for selected books. Google draws this information from its own e-book store which contains over three million books and a network of resellers who sell print and digital versions of the books. In addition to price, Google Books also contains information on accessible versions of books, digital/e-pub versions, PDF versions, and whether the book is in the public domain. We have published a release of this dataset and all of our code to the software repository GitHub. We then used the Zenodo platform to generate a digital object identifier (DOI) for the code.¹² One of the functions of the Zenodo platform is to allow for code to be properly cited and referenced. We published our code in this way for others interested in replicating this work in other contexts. In the next section we will provide an analysis of the results of our queries.

RESULTS

After extracting and processing the data from the course catalog and Google platforms, we obtained 3,030 unique course names and in these courses we found over 15,414 books listed.

Required versus Recommended Reading

From the course catalog data, we found that 11,022 (71.5 percent) books were required readings and the remaining 4,392 (28.5 percent) were recommended.

Upon cleaning and removing duplicates and missing data, we identified 3,014 books that could be queried using the Google Books API. Querying the API returned results for 2,940 books, i.e. it found 97 percent of the books and only seventy-four books could not be found. The Google Books API returns information in JSON format. Figure 3 below shows an example of the JSON information returned for one book.

```

{
  "volumeInfo" : {
    "title" : "Psychiatric and Mental Health Nursing",
    "authors" : [
      "Phil Barker"
    ],
    "industryIdentifiers" : [
      {
        "type" : "ISBN_13",
        "identifier" : "9781498759588"
      },
      {
        "type" : "ISBN_10",
        "identifier" : "1498759580"
      }
    ],
    "imageLinks" : {
      "smallThumbnail" :
"http://books.google.com/books/content?id=btSOCgAAQBAJ&printsec=frontcover&img=1&zoom=5&edge=curl&source=gbs_api"
    }
  },
  "saleInfo" : {
    "isEbook" : true,
    "retailPrice" : {
      "amount" : 62.39,
      "currencyCode" : "USD"
    }
  },
  "accessInfo" : {
    "publicDomain" : false,
    "pdf" : {
      "isAvailable" : true
    }
  }
}

```

Figure 3. Sample of book information returned by Google Books API.

Digital Formats and Public Domain License

Figure 4 shows the numbers of PDF (1,219) and e-book (1,016) versions of books reported to be available. Eight hundred and fifty-four were available in both PDF and e-book format. From the



total of 2,940 individual books listed their availability was as follows:

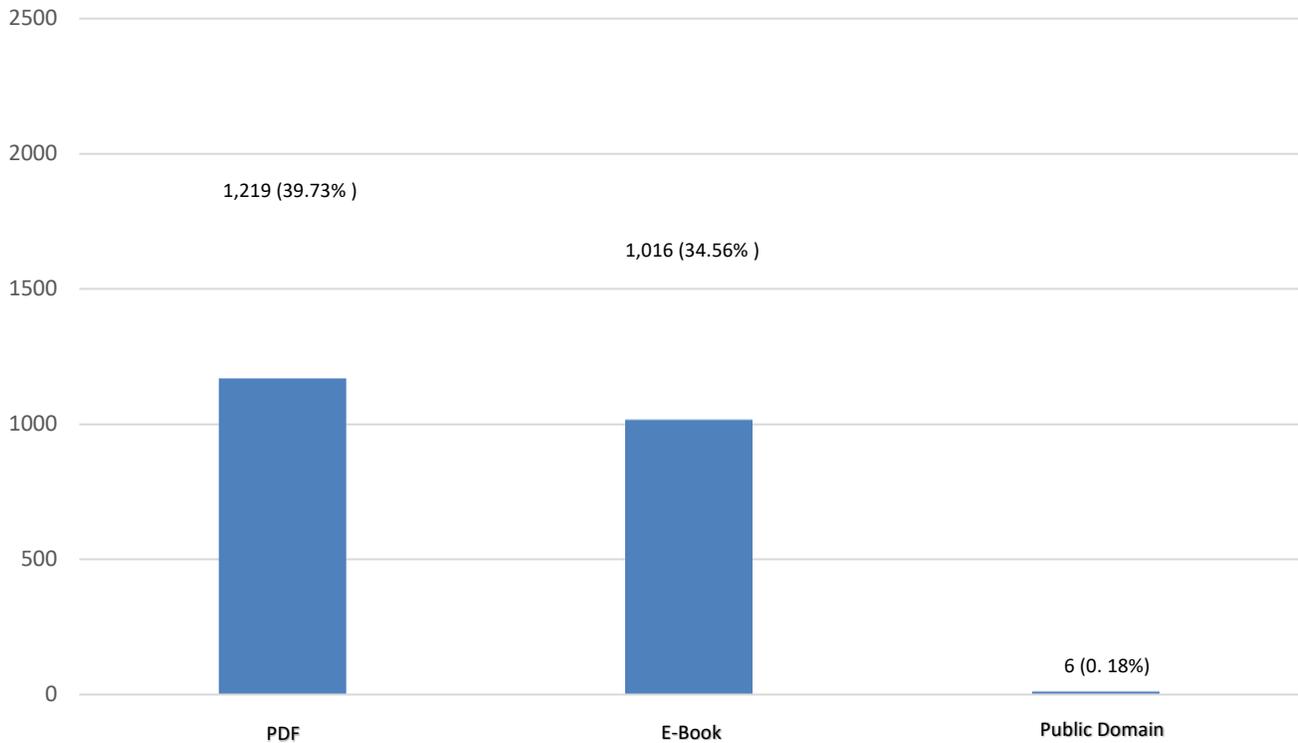


Figure 4. Availability of 2,940 books in digital formats and public domain license.

As per figure 4, only 0.18 percent (six) of the books had a version available in the public domain according to Google Books.

Cost Results

The Google Books API only returned prices for 596 (20 percent) of the books that we searched for. Within that sample, the cost ranged from \$0.99 to over \$452, as illustrated in figure 5. The median price of a book was \$40, and the mean price was \$56.67.

As there are on average 3.96 books per course, this implies an average cost to students of \$224.41 per course taken. As students take an average of 8.05 courses per year, this further implies a cost per year of \$1,806.50 per student if they were to buy new versions of all the books.

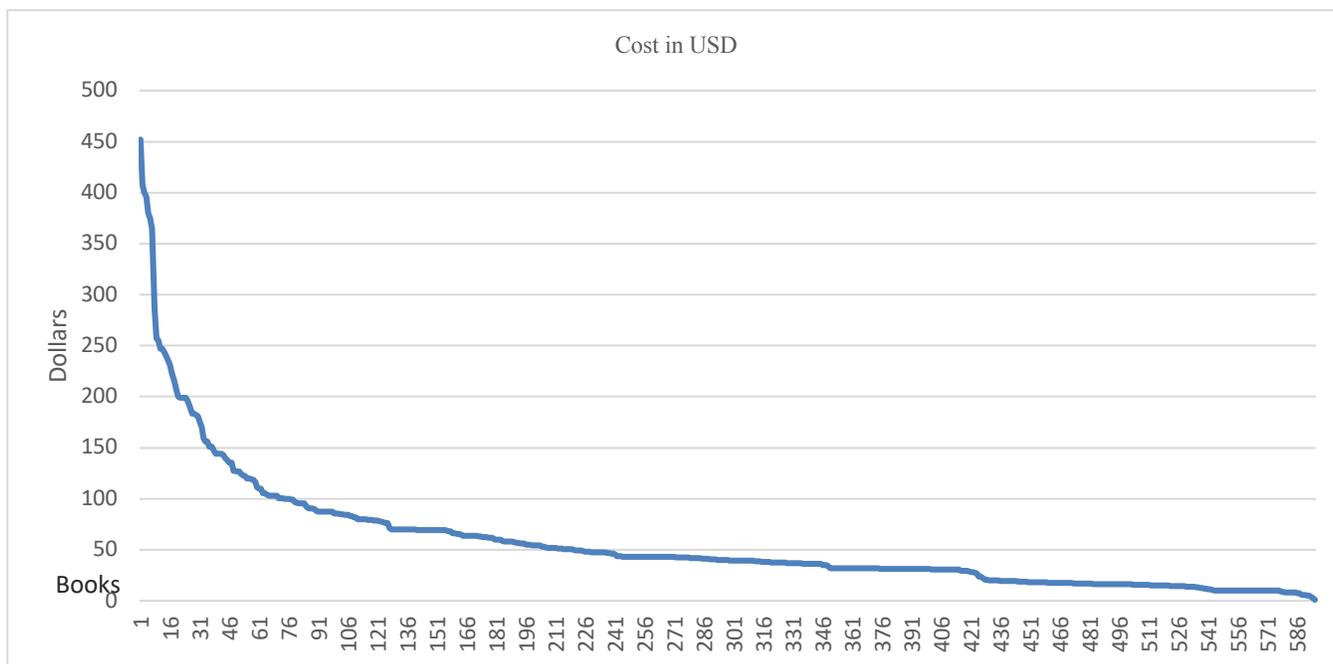


Figure 5. Summary of Book Prices (n = 596).

DISCUSSION AND CONCLUSION

We have demonstrated that it is possible to programmatically search and determine the prices of large numbers of books. We used this information to attempt to estimate the full economic cost of books to students on average in an Irish HEI. We are still actively developing this tool and encourage others to use and even contribute to the code which we have published with the dataset. This proof of concept tool may allow stakeholders with an interest in book costs for students to quickly get real data on large numbers of books. Ultimately, we hope that this will help highlight the costs of many textbooks. Our findings also highlight relatively low levels of digital book availability.

Very few books were found to be in the public domain. A limitation of this research is that there are issues around the coverage of Google Books and its index policies or algorithms. In a literature review of research articles about Google Books in 2017, Fagan pointed out that the coverage of Google Books is “hit and miss.”¹³ In 2017, Google Books included about thirty million books, though Google did not release specific details on its database, as emphasized by Fagan. It is known that content includes digitized collections from over forty libraries, and that US and English-language books are overrepresented.¹⁴ Furthermore, Google Books is only returning results for books that are in the public domain and cannot tell us if books are made available through open licenses such as Creative Commons. Accepting such caveats, however, we have found the Google Books API to be a very useful tool for answering questions about large numbers of books in a systematic way and hope that our findings can help others.

The prices that we derived in this study were for new books only. However, the new book prices provide a baseline for all other prices, e.g. a used book or a loan book price will be relative to a new book price and library budgets will need to take account of new book prices.¹⁵ Further study is required to determine a more realistic figure for the cost of textbooks and the next phase of our



wider open textbook research projects involves interviews and focus groups with students to better understand the lived reality of their relationship with textbooks.¹⁶

REFERENCES

- ¹ Stephen L. Baglione and Kevin Sullivan, "Technology and Textbooks: The Future," *American Journal of Distance Education* 30, no. 3 (Aug. 2016): 145-55, <https://doi.org/10.1080/08923647.2016.1186466>.
- ² Etan Senack and Robert Donoghue, "Covering the Cost: Why We Can No Longer Afford to Ignore High Textbook Prices," Report, The Student PIRGs (Feb. 2016), www.studentpirgs.org/textbooks.
- ³ Elaine Allen and Jeff Seaman, "Opening the Textbook: Educational Resources in U.S. Higher Education, 2015-16," Report, BABSON Survey Research Group (July 2016), <https://www.onlinelearningsurvey.com/reports/openingthetextbook2016.pdf>.
- ⁴ William and Flora Hewlett Foundation (2019), <http://www.hewlett.org/programs/education-program/open-educational-resources>.
- ⁵ 2012 Paris OER Declaration, http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/WPFD2009/English_Declaration.htm.
- ⁶ Mary Burgess, "The BC Open Textbook Project," in *Open: The Philosophy and Practices That Are Revolutionizing Education and Science*, Rajiv S. Jhangiani and Robert Biswas-Diener (eds.). (London: Ubiquity Pr., 2017): 227–36.
- ⁷ Nicole Allen, "Congress Funds \$5 Million Open Textbook Grant Program in 2018 Spending Bill," *SPARC Open* (Mar. 20, 2018), <https://sparcopen.org/news/2018/open-textbooks-fy18/>.
- ⁸ Jennifer Ma et al., "Trends in College Pricing," Report, The College Board (Oct. 2017), https://trends.collegeboard.org/sites/default/files/2017-trends-in-college-pricing_0.pdf.
- ⁹ Kaitlyn Vitez, "Open 101: An Action Plan for Affordable Textbooks," Report, Student PIRGs (Jan. 2018), <https://studentpirgs.org/campaigns/sp/make-textbooks-affordable>.
- ¹⁰ Mark Brown, Eamon Costello, and Mairéad Nic Giolla Mhichíl, "From Books to MOOCs and Back Again: An Irish Case Study of Open Digital Textbooks," in *Exploring the Micro, Meso and Macro. Proceedings of the European Distance and E-Learning Network 2018 Annual Conference, Genova, 17-20 June, 2018* (Budapest: The European Distance and E-Learning Network): 206-14.
- ¹¹ Google Books API (2018), <https://developers.google.com/books/docs/v1/reference/volumes>.
- ¹² Eamon Costello and Richard Bolger, "Textbooks Authors, Publishers, Formats and Costs in Higher Education," *BMC Research Notes* 12, no. 1 (Jan. 2019): 12-56, <https://doi.org/10.1186/s13104-019-4099-1>.

¹³ Jody Condit Fagan, "An Evidence-Based Review of Academic Web Search Engines, 2014-2016: Implications for Librarians' Practice and Research Agenda," *Information Technology and Libraries* 36, no. 2 (Mar. 2017): 7-47, <https://doi.org/10.6017/ital.v36i2.9718>.

¹⁴ Ibid.

¹⁵ Anne Christie, John H. Pollitz, and Cheryl Middleton, "Student Strategies for Coping with Textbook Costs and the Role of Library Course Reserves," *portal: Libraries and the Academy* 9, no. 4 (Oct. 2009): 491-510, <http://digital.library.wisc.edu/1793/38662>.

¹⁶ Eamon Costello et al., "Textbook Costs and Accessibility: Could Open Textbooks Play a Role?" *Proceedings of the 17th European Conference on eLearning (ECEL)*, vol. 17 (Athens, Greece: 2018): 99-106.

