

June 2010

A Scholar Contemplates the Google Book Settlement

James O'Donnell
Georgetown University

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

O'Donnell, James (2010) "A Scholar Contemplates the Google Book Settlement," *Against the Grain*: Vol. 22: Iss. 3, Article 8.
DOI: <https://doi.org/10.7771/2380-176X.5547>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

A Scholar Contemplates the Google Book Settlement

by James O'Donnell (Professor of Classics & Provost, Georgetown University, Washington DC 20067)

There are three questions we must ask about a project for mass digitization:

First, should the work be done?

Second, should the results be made widely available?

Third, can the results be easily used?

I propose to answer these questions in order. The first I will answer gratefully, the second cheerfully, and the third grumpily. At a moment when everyone has an opinion about the **Google Book Settlement**, I may not seem grumpy enough to many, but my gratitude and my cheerfulness are real.

First, yes, the work should be done. Ten years ago, I was privileged to lead a **National Academy of Sciences** study group looking at the digital future of the **Library of Congress** (our results were published in book form by the **National Academies Press**, *LC21: A Digital Strategy for the Library of Congress* [2000], a volume that I think still holds up quite well). That extraordinarily talented team of scientists, scholars, and librarians was positively wistful about a possibility we imagined beyond our reach. "Congress should just appropriate a billion dollars and digitize the whole thing," one of our number liked to say, and we all shook our heads: this couldn't possibly happen, not in the real world of **Congress** and politics.

Well, a private company has gone and done it. Not all of it, not a billion dollars worth, but the quarter of a billion dollars or so that **Google** is said to have spent on this task is an extraordinary benefaction to the public weal. We should work hard to remember to say thank you. If they hadn't done it, nobody else was fighting to the fore to do it first, least of all **Congress**.

The second question is a trick, of course. Should the works be made widely available? Of course. But once that is said, the knives will come out. It should all be made widely and *freely* available. Information still wants to be free, access should be open, and rainwater should be beer. That is not quite happening, though it is worth noticing that as far as **Google** is concerned, the material over which they have complete say, the U.S. out-of-copyright material undoubtedly before 1923, *is* in fact being made freely available. The negotiated settlement with the publishers took longer than anyone

imagined, is complicated beyond the capacity of non-specialists to remember from one week to another, and does indeed put a price tag on the in-copyright material in the project.

This is not Google's fault. The intellectual property laws we have and the publishing industry to which they have given rise are real choices made in a free society and, indeed, we have all benefitted mightily from them. That there may be better ways to make laws and do business is obviously a topic for discussion and debate, and the wrangling proceeds. But for now, the **Google Settlement** as proposed reminds one of **Churchill's** crack about Democracy as the worst form of government — except all the others that have been tried. The **Google Settlement** will open vast ranges of published material to a broad public for easy use. It's a start. (And of particular note is that the **Settlement** will make it possible to have access to so-called "orphan works" — in copyright but owned by individuals unknown and not safely reprintable. This too is progress.)

So I am cheerful again, and not a little grateful. Impatient, as well: if the legal wrangling will end, I'll be able to get my hands on things now tantalizingly just beyond my reach.

Third question: can the material be easily used? This is another trick, of sorts. Making the material available free of charge might be nice, but there's much more to accessibility than price tag. Information has to be findable and, when found, usable. The reader needs what one of my colleagues on that **National Academy** study called "intellectual access". In a mass of information, the reader has to be able to find the right material at the right time and read it and think about it and even copy chunks of it.

Here's where **Google's** most notable failing strikes the eye. Numerous observers, perhaps most flamboyantly the linguist **Geoff Nunberg**, have called attention to the woeful defects of **Google's** metadata. **Google** trusts that search engines will gain all the access one needs and has accordingly shown astonishingly cavalier neglect towards one of the great achievements of human intellect: I mean that body of highly

structured and awesomely accurate metadata embodied in the catalog records of our libraries. Yes, those metadata have their limits. Yes, searches can find you much that the old MARC records cannot find. But those records remain astonishingly useful for many purposes. Each MARC record reduces a volume or cognate set of volumes to a consistent, structured, and easily intelligible representation in very small compass. When there are masses of information to be dealt with, those representations are far easier to use than a collection of search hits from the original material can ever be. But **Google** has thrown those data aside.

How does this affect a real working scholar? I will give just one example. If I want to find the seven volumes of **Tyrrell** and **Purser's** famous edition of **Cicero's** letters, published in the first decades of the last century, I have a very particular goal in mind. I want to find Volume One, then Volume Two, then Volume

Three, and so forth, until I have found Volume Seven. If I were in a physical library, I would expect to find them together on a shelf, or else in a record (keyed to the MARC record) of their present temporary location. In **Google Books**, I *think* I find several copies of Volumes One and Two, and no more. Furthermore, I find nothing to help me know whether I am looking at first editions or revised editions, and nothing to tell me where to look for Volumes

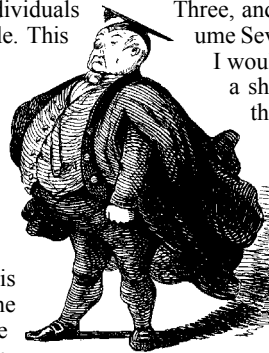
Three through Seven. Those books are still very useful to me: but not in **Google**.

Now **Tyrrell** and **Purser** cross the boundary into possibly-copyright, but there are not (when last I looked) any indications that Volumes Three through Seven have shown up on **Google's** scope at all. But I may simply be wrong.

This story repeats itself throughout the corpus of scholarly literature that I am curious about. I am squirreling away on my laptop huge PDF files of **Google** books versions of things I want to read — because I'm never sure I'll find them again, and because I'm frustrated that I can't find complete sets of things that were published even on the same day together.

One could make everything in **Google Books** absolutely free to me today and it wouldn't help this problem. The digital representations that **Google** has made available aren't yet a library — and indeed, in an important sense, they aren't even books yet, any more than a stack of de-accessioned volumes in a dumpster are books any longer. Books are books when they are alive and speaking, their contents known and knowable. Books do furnish a room, but care must be taken with the room and with the books. **Google** has a long way to go on that score.

And that's why, and where, and how I'm grumpy. 🐻



James J. O'Donnell is Provost at **Georgetown University**. He has published widely on the history and culture of the late antique Mediterranean world and is a recognized innovator in the application of networked information technology in higher education. In 1990, he co-founded **Bryn Mawr Classical Review**, the second on-line scholarly journal in the humanities ever created. In 1994, he taught an Internet-based seminar on the work of **Augustine of Hippo** that reached 500 students. He has served as a Director and as President of the **American Philological Association**; he has also served as a Councillor of the **Medieval Academy of America** and has been elected a Fellow of the **Medieval Academy**. He serves as Delegate of the APA to the **American Council of Learned Societies**, as Chair of the Executive Committee of the Delegates, and as a member of the Board of Trustees of the **ACLS**. 🐻