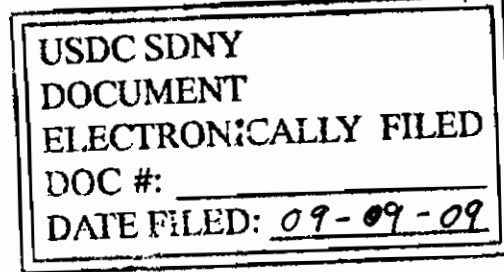




Erez Lieberman-Aiden
Jean-Baptiste Michel

Program for Evolutionary Dynamics
1, Brattle Square, suite 6
Cambridge, MA 02138, USA

The Honorable Denny Chin
United States District Judge
Daniel Patrick Moynihan Courthouse
500 Pearl Street
New York, NY 10007



September 3, 2009

Re: The Authors Guild et al. v. Google, Inc. Case No. 1:05 cv 8136 (S.D.N.Y.)

Dear Judge Chin:

With the court's permission, we would like to submit this letter as amici curiae in support of the settlement in the above-referenced case. We are scientists who feel strongly that the 'nonconsumptive' research corpus described in section 7.2(d) of the settlement will dramatically advance research in the sciences and in the humanities, and will provide important benefits to the public at large.

Scholarly work in the humanities today consists almost entirely of the careful analysis and dissection of individual passages and exemplary instances drawn from the canons of language, literature, and the arts. This method has enabled a deep comprehension of the historical and cultural trends that have shaped our civilization.

The Author's Guild et al. v. Google Inc.

Doc. 303

Nevertheless, this method is not the only one of which we might conceive. In 2007, we described a method in the journal *Nature* which required us to spend two years sifting through a large number of texts and collecting a large amount of data on the history of the English language. We used this data to quantify historical trends and to derive dynamical equations that accurately describe these trends, and then used these equations to make specific, novel predictions about the future of the English language. This work, which appeared on the cover of *Nature* and was widely discussed in the popular press, illustrated the vast potential of computational and theoretical methods in the humanities.

However, the availability of massive amounts of relevant text, in digital form, is crucial to this method. By contrast to traditional ('consumptive') methods, we do not directly read the texts; instead, via statistical (i.e., 'nonconsumptive') means, we are able to extract the underlying dynamics. The trends we have found are often impossible to observe by a scholar reading the books one at a time. Instead, these powerful trends become apparent only when vast amounts of data are analyzed at once.

For 'nonconsumptive' studies, the ideal resource would be a digital collection of centuries of the cultural output of human civilization, including, for example, all works of art, books, and newspapers. Currently, scholars lack access to any digital resource that even approaches this ideal. The digitization of nearly all available books, undertaken by Google, can transform this state of affairs: by recording all books into a unified format and data structure, it becomes possible to query a small but significant slice of the cultural output of human civilization. The terms of the settlement would give the world's scholars access to a 'nonconsumptive' corpus of an extraordinarily vast scale, with the potential to profoundly reshape the humanities.

Just one example of the possibilities of such a corpus is our recent work in connection with public health. Using the Google digital corpus, we searched for evidence of historical epidemics of such diseases as cholera and the flu. We used the simplest possible approach, gauging public concern by the frequency with which the words 'cholera' and 'influenza' occurred in books published in a given year. The results – which took only a fraction of a second to compute – provided us with a year-by-year historical outline of epidemics of these diseases, showing the rise, the ebb, and the frequency of such events. The recent furor surrounding the emergence of swine flu highlights our need for better ways of anticipating epidemic disease, and the value of objective, historical data for predicting the contours of future pandemics cannot be understated. From the standpoint of public health alone, the Google data is of phenomenal value.

For the reasons described above, we respectfully request that the court approve the settlement in order to make these resources available to the community of scientists and scholars and the world at large.

Respectfully submitted,



Erez Lieberman-Aiden
*Harvard School of Engineering and Applied Sciences
Harvard-MIT Division of Health Science and Technology*



Jean-Baptiste Michel
Department of Systems Biology, Harvard Medical School

Cambridge, Massachusetts
September 3, 2009