JAMES GRIMMELMANN*

Information Policy for the Library of Babel

The universe (which others call the Library) is composed of an indefinite, perhaps infinite number of hexagonal galleries.
—Jorge Luis Borges[1]

BORGES'S 1941 SHORT STORY *THE LIBRARY OF BABEL* DESCRIBES an unbelievably large library containing all possible books. Within the "total"[2] and "endless"[3] reaches of the Library,"[t]here [is] no personal problem, no world problem, whose eloquent solution [does] not exist—somewhere"[4] but also "[f]or every rational line or forthright statement there are leagues of senseless cacophony, verbal nonsense, and incoherency."[5] As Borges describes it, the Library is the greatest imaginable source of information: it contains "The Vindications—books of *apologiæ* and prophecies that would vindicate for all time the actions of every person in the universe and that held wondrous arcana for men's futures."[6] But the Library's vastness and disorganization also make it almost completely useless: "[T]he chance of a man's finding his own Vindication . . . can be calculated to be zero."[7] The image of the Library is haunting and suggestive. What would we do if we took it at face value? In this Essay, I propose to do just that: set out a few principles of sensible information policy for the Library of Babel.

I. THE LIBRARY

Suppose that we were advisors to the Federal Library Commission, an arm of the duly constituted government of the Library of Babel. In our capacity as lawyers and information policy experts, we have been asked to suggest ways to advance the

* Associate Professor of Law, New York Law School. As of January 1, 2009, this Article is available for reuse under the Creative Commons Attribution 3.0 United States license, http://creativecommons.org/licenses/by/3.0/us/.

1. JORGE LUIS BORGES, *The Library of Babel, in* COLLECTED FICTIONS 112, 112 (ANDREW HURLEY TRANS., PENGUIN BOOKS 1998) (1941).
2. *Id.* at 115.
3. *Id.* at 113.
4. *Id.* at 115.
5. *Id.* at 114.
6. *Id.* at 115.
7. *Id.*

public interest of the Library. What might we propose? The following propositions strike me as sound bases for the work of the Commission:

The public interest means readers' interest.

Intellectual property policy is accustomed to speaking of a trade-off between authors' interests and a public interest in access to information. In the Library of Babel, however, there are no authors. Borges states as an "axiom[ ]"[8] that "[*t*]*he Library has existed* ab æternitate. . . . [It] can only be the handiwork of a god."[9] All possible books already exist; no further incentive is required to bring them into being. Nor does any book contain any expression of an author's unique personality; every book in the Library is equally anonymous. "The certainty that everything has already been written annuls us, or renders us phantasmal."[10] In contrast, these books are of immense potential value to readers, some of it individual (the Vindications) and some collective ("The universe was justified; the universe suddenly became congruent with the unlimited width and breadth of humankind's hope.")[11] This "intact and secret treasure"[12] is a treasure because it can be appropriated by readers, not because it encourages authors or reflects the imprint of their personality.

Or, looked at another way, the Federal Library Commission must serve the inhabitants of the Library (or "librarians," as Borges calls them). There is no one else for it to serve. The inhabitants, however, encounter the Library first and foremost as readers. Indeed, their search for information in its stacks (or the repudiation of that search) is the principal act that gives their own lives meaning. They search for their Vindications, for "the books of the Crimson Hexagon—books smaller than natural books, books omnipotent, illustrated, and magical."[13] On the shelves somewhere are "the detailed history of the future, the autobiographies of the archangels . . . the treatise Bede could have written (but did not) on the mythology of the Saxon people,"[14] and other informational treasures beyond measure. We do our job well if we help our constituents find the true and beautiful books and steer them clear of the false and ugly ones.

Infrastructure matters.

Access to knowledge always depends on access to knowledge infrastructure. Borges details the arrangements of the books on shelves and the configuration of the hexagonal galleries of the Library because these facts matter. The numinous promise of

8. *Id.* at113.
9. *Id.*
10. *Id.* at 118.
11. *Id.* at 115.
12. *Id.*
13. *Id.* at 116.
14. *Id.* at 115.

the Library is that any book could actually be found within it; we could make the journey to the proper hexagon, look on the proper shelf, and hold it in our hands. The "spiral staircase, which winds upward and downward into the remotest distance,"[15] the vestibules linking the hexagons, and the shelves are infrastructure; they make it possible to reach and use the books. So too, more indirectly, are the sleeping compartments and toilets off the vestibules, and the "spherical fruits"[16] that illuminate the galleries; they are infrastructure that make it feasible for us to live in (and thus to consult) the Library.

We have made some strides in improving the Library's infrastructure: Borges refers in passing to "a hexagon in circuit 15-94,"[17] and we at the Commission might take credit for the numbering scheme. But on the whole, we're falling far short of what we could accomplish. The light from the bulbs is "insufficient, and unceasing";[18] some of our constituents "talk about a staircase that nearly killed them—some steps were missing . . . ."[19] Even more troublingly, basic education in the Library may be breaking down: "I know districts in which the young people . . . cannot read a letter."[20] Really helping citizens make full use of the Library requires provisioning the technologies they use to access it, and also helping them acquire the media literacies they use to make sense of it. We can help with both.

Censorship is usually irrelevant.

Some of the books in the Library are dangerous in themselves: "There is no combination of characters one can make—*dhcmrlchtdj*, for example—that the divine Library has not foreseen and that in one or more of its secret tongues does not hide a terrible significance."[21] Others are dangerous because they divert us from the books we seek: "thousands and thousands of false catalogs,"[22] the "proof of the falsity of the *true* catalog,"[23] "some perfidious version of his own [Vindication]."[24] In the face of these dangers, some "Purifiers" have turned to censorship:

> *"They would invade the hexagons, show credentials that were not always false, leaf disgustedly through a volume, and condemn entire walls of books. It is to their hygienic, ascetic rage that we lay the senseless loss of millions of volumes."*[25]

15. *Id.* at 112.
16. *Id.*
17. *Id.* at 113.
18. *Id.* at 112.
19. *Id.* at 116.
20. *Id.* at 118.
21. *Id.* at 117.
22. *Id.* at 115
23. *Id.*
24. *Id.*
25. *Id.* at 116.

In the abstract, because every book is meaningful to some possible reader, it might seem that purging a volume is an unpardonable crime. But the same considerations that make individual authorship moot also tend to make individual censorship moot. (Destroying a book is just the mirror image of creating one.) The Library endures far above our poor power to add—or detract. As Borges reminds us, in the vastness of the Library, "any reduction by human hands must be infinitesimal"[26] and for any book "there are always several hundred thousand imperfect facsimiles—books that differ by no more than a single letter, or a comma."[27] Censors who rip a book from our hands have harmed us, to be sure, as have those who burn down so much of the Library as to make appreciably harder the task of finding shelves with books to read. But on the long view, any one person is so insignificantly small when compared with the treasure-house that is the Library of Babel that a few depredations here and there do not much affect either the availability of any given information or the average librarian's search through the galleries. (Indeed, if Borges's final suspicion is correct, and the Library is "*unlimited but periodic*,"[28] it contains an infinite number of copies of each book, and censorship is infinitesimally irrelevant even within the Library's holdings of that precise title.)

The problem is access, not creation.

Here we come to the crux of the matter. There is no difficulty in ensuring that the Library contains a (near) copy of any book. The difficulty lays in our being able to make use of the books that the Library contains. It follows from the Library's encyclopedic collection and the disorder of its stacks that one browsing the shelves must face "the formless and chaotic nature of virtually all books."[29] The book that is intelligible to us is the rare exception. Borges speaks of a few: one that repeats the letters "M C V"[30] over and over; one that is gibberish except for the single phrase "O Time thy pyramids";[31] one "as jumbled as all the others, but containing almost two pages of homogenous lines."[32] The consequence of this rarity is that knowledge about intelligible books becomes a valuable, nearly mystical commodity. The book containing "O Time thy pyramids" is "much consulted"[33] the one with two whole pages of sense provided grist for linguists and philosophers for almost a "century."[34]

Similarly, Borges makes much of the physical difficulties of searching for books. I have alluded to the dangers of broken stairs; there are entire zones of the Library that are beset by "[e]pidemics, heretical discords, pilgrimages that inevitably degen-

26. *Id.*
27. *Id.*
28. *Id.* at 118.
29. *Id.* at 113.
30. *Id.*
31. *Id.* at 114.
32. *Id.*
33. *Id.* at 113–14.
34. *Id.* at 114.

erate into brigandage . . . ."[35] The narrator will die only "a few leagues from the hexagon where I was born."[36] The atomic useful informational artifact of the Library is not the book; it is the book in hand. We must know where to find the book, we must have some inkling of its contents, and we must be able to make the (potentially quite long) journey to it. Getting the information into the hands of those who need it is where all the hard work lies.

The Library is nearly but not completely useless.

And it is hard work indeed. The books are arranged in no order we can understand. Nor do we have any usable index. The letters on the outside of each book "neither indicate nor prefigure what the pages inside will say."[37] The "faithful catalog of the Library"[38] exists within it somewhere, but we do not know where, nor how to distinguish it from the "thousands and thousands of false catalogs."[39] Little wonder that "[i]nfidels claim that the rule in the Library is not 'sense' but 'nonsense'"[40] and that "[o]ne blasphemous sect proposed that the searches be discontinued and that all men shuffle letters and symbols . . . ."[41] The Library of Babel seems like a pure and perfect example of information overload.

But it is not. Borges swears that the Library "includes not a single absolute piece of nonsense."[42] Nor is it completely disordered. The narrator has seen a book titled "*The Plaster Cramp*"[43] and one titled "*Combed Thunder*,"[44] along with two Vindications, "which refer to persons in the future, persons perhaps not imaginary."[45] Such finds (along with the previously mentioned book of "M C V"s and a book with two pages of "a Samoyed-Lithuanian dialect of Guaraní, with inflections from classical Arabic"[46]) are significant instances of order. Once found, they stay found. Our useful knowledge of the Library's contents is gradually increasing. It may take eons, but even random explorations will slowly build a sub-Library of useful books we actually have seen. The more librarians we enlist in the search, finding and tagging books and sharing what they have found, the faster this sub-Library will grow.

And we should perhaps be optimistic about the search itself. Some of the books Borges mentions are extreme statistical improbabilities. One who spent his whole life flipping through books in a truly random library would be profoundly unlikely

35. *Id.* at 118.
36. *Id.* at 112.
37. *Id.* at 113.
38. *Id.* at 115.
39. *Id.*
40. *Id.* at 117.
41. *Id.* at 116.
42. *Id.* at 117.
43. *Id.*
44. *Id.*
45. *Id.* at 115.
46. *Id.* at 114.

to find books displaying that much structure. A book containing nothing but endless repetitions of "M C V" may contain less information than a book of English text, but it is also far rarer. Unless Borges, through some kind of narrative anthropic principle, is inordinately lucky among librarians, the most natural inference is that the Library is in fact ever so slightly non-random. Its shelves show just enough structure that a "faithful catalog" might just barely propound some system by which books are slotted into their assigned places. If only we could lay our hands on that catalog.

II. THE BOOK-MAN

One idea in the story is particularly intriguing: the Book-Man:[47] "On some shelf in some hexagon, it was argued, there must exist a book that is the cipher and perfect compendium *of all other books*, and some librarian must have examined that book; this librarian is analogous to a god."[48]

How would our job working for the Federal Library Commission change if we had a Book-Man to contend with? Why would the other librarians consider him "analogous to a god?"

The Book-Man makes the Library useful.

With only a little bit of reading between Borges's lines, it's possible to believe that a Book-Man would utterly transform our informational lives. The Book-Man is the one who has seen the "cipher and perfect compendium." If we identify this "total book"[49] with the "faithful catalog"[50] mentioned previously, the nature of its perfection becomes clear. There is no way we could compress the contents of all possible books into a single book, but a single book could lay bare the Library's hidden immanent order. Asking the Book-Man questions about what he has learned from the faithful catalog is the next best thing to consulting it ourselves. He would thus not be particularly knowledgeable about the actual contents of the books in the Library, but would instead be a guide to the Library's shelving patterns.

"Where can I find the lost books of Tacitus?" we might ask him, or "Where can I find a book that explains the significance of 'O Time thy pyramids?'" Given the dense complexity of a total book, the imperfections of human memory, and the vagueness of our questions, I suspect that the Book-Man's knowledge would be notably partial. Pinpointing a specific book seems unlikely; instead, he would probably direct us to a shelf, or a hexagon, with the claim that it was likely to contain a book relevant to our question. Often it would; sometimes it would not. But "often" would still be a radical improvement on what we could do without his help. Borges

47. *Id.* at 116.
48. *Id.*
49. *Id.* at 117.
50. *Id.* at 115.

reminds us that "for a book to exist, it is sufficient that it be *possible*."[51] In the Library of Babel, any question that has an answer will therefore have an answer somewhere on its shelves, and the Book-Man can usually direct us to that answer. In short, the Book-Man could solve our access problem.

An impostor could not pretend to be the Book-Man.

The Book-Man's task is insanely difficult. Any librarian could make recommendations, only to have them shown up as useless after his disgruntled questioners returned from their searches with their hands empty. A librarian who had chanced into seeing a few noteworthy volumes might be able to put on a good exhibition. But answering questions about the locations of books on arbitrary topics is a hard problem; there is no way to do it reliably unless one really does have information about all books. Simply by asking for help finding books on sufficiently many random topics, we could with arbitrarily high probability expose as a fraud any pretender to the title of Book-Man. The impossibility of a false Book-Man, perhaps, justifies the "sect that worshiped that distant librarian"[52] and the desperate searches the narrator and others have made for "the idolized secret hexagon that sheltered Him."[53] Out of all the religious and delirious schemes the librarians pursue, this is one of the least mystical—it really is possible to distinguish success from failure.

The Book-Man could keep secrets from us and we'd never know.

A putative Book-Man cannot pretend to be better at his art than he really is. The reverse is emphatically not true. A librarian, looking through the faithful catalog, might decide that the burden of being a Book-Man was too heavy to bear. He places the book back on the shelf, slowly backs away, ascends to the next hexagon, and never says a word to anyone else. Another librarian, suspecting some secret knowledge, asks him about the lost books of Tacitus. Our renunciant names a hexagon utterly at random. Short of unspeakably barbaric acts, there is no way we can convince an unwilling Book-Man to assume the mantle, and thus no way we can spot one unless he wishes to be spotted.

A disturbing corollary follows. We have noted that even a true Book-Man will give vague and sometimes unhelpful directions. There is nothing to stop him from underplaying his powers a little. We ask him about the lost books of Tacitus, and he claims not to know (when of course he does). Unless we should happen to stumble across the books ourselves (an extraordinarily unlikely event in the Library), we might never suspect anything is amiss. We ask him for a Vindication, but he points us to a perfidious version of it, instead. We discover the falsehood just in time, but when we confront the miscreant who pointed us to it, he pleads the uncertainties of

51. *Id.* at 117 n.3.
52. *Id.* at 116.
53. *Id.* at 117.

his art, the briefness of his glance at the catalog. Who are we to question his story? He has, after all, been wrong on plenty of other occasions. He might even have the catalog secreted in a nearby hexagon, on a shelf otherwise occupied only by gibberish; all those mistakes are just an act; we'd never know.

The problem is that the Book-Man's quasi-mystical knowledge is based on a source inaccessible to us, surrounded with inherent uncertainties, and subject to his personal discretion. Any pattern we think we perceive in his answers could be sandbagging, or it could be an artifact of an imperfect human attempt to process the faithful catalog's trans-Byzantine organizational system, or it could be an unavoidable glitch introduced by the question-asking process. The Book-Man's system is a knowledge expander: he starts from a short fragment of information (e.g. "the lost books of Tacitus") and makes a guess as to which much larger body of information it best refers. He could be wrong. As Borges notes, the Library is a place of linguistic diversity: "[A] few miles to the right, our language devolves into dialect and . . . ninety floors above, it becomes incomprehensible."[54] Nor are words themselves fixed: "[I]n some [languages], the *symbol* 'library' possesses the correct definition 'everlasting, ubiquitous system of hexagonal galleries,' while a library—the thing—is a loaf of bread or a pyramid or something else, and the six words that define it themselves have other definitions."[55] The speaker of such a language who asks the Book-Man for what she thinks is a book about baking will not be happy to be directed to a book on the architecture of the Library. Further, the very nature of our search is that we ourselves don't entirely know what's within the covers of the book we're looking for when we ask the question, so that the question could plausibly refer to any of trillions of trillions of possible books. The Book-Man must guess at all these ambiguities. ("You who read me—are you certain you understand my language?"[56]) Within the space we must allow him to make those guesses, he could hide any number of bodies.

The Book-Man can play favorites.

Perhaps most troublingly of all, the Book-Man might reserve his trickiest, most misleading advice for his secret enemies. He's helpful often enough that his enemies trust him utterly, but he's just saving up for the one great bum steer that will end with them falling over a railing; "[their] tomb will be the unfathomable air"[57] between the galleries; their bodies "will sink for ages, and will decay and dissolve in the wind engendered by [their] fall, which shall be infinite."[58] The rest of us will never know. Even if they recognize the trap and warn us of a dire fate narrowly

54. *Id.* at 114.
55. *Id.* at 118.
56. *Id.*
57. *Id.* at 112.
58. *Id.*

averted, their claims will be anecdotal, sporadic, hard to recognize as a pattern, easily explained away.

He doesn't even have to give different answers to the same question to different people (if he did, we might be able to lay bare the betrayal). All he needs to do is to select particular topics on which his directions will be misleading. People naturally ask him about such a wide assortment of topics that he can target people he doesn't like by giving bad suggestions on the topics that matter to them but not to other people. Thanks to his dislike of the cut of your jib, you'll never find your Vindication; it is nowhere near the hexagon he recommended. If I ask him about your Vindication, I get the same recommendation. But when either of us asks about *my* Vindication, the Book-Man can point us not just to the right hexagon, but to the right shelf. He's given the two of us exactly the same answers, while still using his unconstrained discretion to favor me over you.

The more Book-Men, the better.

One Book-Man may be forgetful. But suppose we had two? We could put each question to the both of them. By combining forces, the Book-Men could find any book that either one of them alone could find. Given the variety of reading styles, it seems likely that each of them would have understood the total book in a different way, have absorbed different ideas. Once we have multiple Book-Men, the variability of their knowledge of the Library becomes not a weakness, but a strength. The more divergent their thinking, the less likely they are to make exactly the same mistakes—and thus the more likely we are to find the books we seek.

If we separate the Book-Men, and have them answer our questions simultaneously from different hexagons, we can even fix some of the trust problems. This protocol makes it hard for them to collude; any rigged answers would need to be agreed upon in advance. With a sufficient diversity of questions and a sufficient intensity of follow-ups, we can be statistically quite confident that any given common answer is not the result of some diabolical scheme on the part of the Book-Men. By comparing their answers against each other, and against the contents of the shelves, we can also make bad advice more visible. If one Book-Man directs us to a pack of lies, nothing may seem amiss. But if one Book-Man directs us to a book that proclaims the exact opposite of the book recommended by the other, our suspicions should be aroused. The advice of the second Book-Man, by giving us another picture of what is out there, helps us check-up on the first—and vice versa.

Not only is it harder for one Book-Man to steer us awry, but it is harder for him to slack. If over time we find one Book-Man more reliable than the other, we can start checking the shelves he selects first, demoting the other a few notches from his god-like status. Thus, all in all, the competition among the two Book-Men improves the overall quality of our searches, makes it harder for them to mislead us (and thus makes us more trusting of them), and creates an incentive for them to work hard at giving good advice. This argument, of course, generalizes to the case

of more than two Book-Men. It's not a complete solution to the problems of the Library, but it's a step in the right direction.

III. THE INTERNET

As it announces in its very first sentence, *The Library of Babel* is an allegory for the universe. This essay has also treated it as an allegory—and an anachronistic and transparent one at that. For "Library of Babel," read "Internet." For "book," read "Web site." And for "Book-Man," read "search engine." It's almost a cliché to assert that the Internet is like a vast library, that it causes problems of information overload, or that it contains both treasures and junk in vast quantities. Looking at it through the lens of Borges's Library amplifies these themes to their utter limit, and thus makes them fresh again. The ten principles set forth above are completely serious. Here they are again, using the proper terminology:

1. The public interest means readers' interest. This may be the most surprising claim about the Internet, but it is also the most important. In the Library, it is trivially obvious. A decade or two from now, it will seem trivially obvious on the Internet too. In an environment of extreme informational abundance, the principal moral imperative is to get that information into the hands of the people who want and need it. If the information production problem has been solved, then the information-consuming public is properly the sole beneficiary of information policy.

2. Infrastructure matters. Not all information policy has to do directly with information. Keeping the electronic infrastructure of the Internet up and running and enabling it to grow is the single best thing that government has done for information policy in the last two decades. Making sure that citizens have the literacies they need to learn from and to evaluate critically what they find online is similarly a basic task of information policy.

3. Censorship is usually irrelevant. When information is overwhelmingly plentiful, the deck is stacked against would-be censors. As many governments and media companies have learned, getting a particular file offline and keeping it offline is like playing a constantly accelerating game of Whac-a-Mole. As offensive as some individual censorship efforts have been, even substantial filtering systems have not (thus far) crippled the Internet as a whole.

4. The problem is access, not creation. The divide between informational haves and informational have-nots is wide. So is the gap of a different sort between those who make information and those who need it. What we as the reading public most need are reasonable, fair, and effective ways to get our hands on the vast treasure-houses of knowledge that already exist.

5. The Internet is nearly, but not completely useless. There are billions of Web pages. No one can possibly read them all, or even any significant fraction of them. It is possible to be hopelessly lost in your own inbox. You can find interesting things by browsing hyperlinks or by taking recommendations from friends of inter-

esting things they have seen, but the natural infrastructure of the Internet provides almost no useful structure for finding the information you need.

6. Search engines make the Internet useful. Information overload demands good filters. There are many technologies that pare down a vast universe of possibilities into a smaller higher-quality set, but search technologies are special because they are responsive to individual users' informational requests. Good search across a sufficiently large knowledge base promotes autonomy by letting each person find and use the information she herself knows that she needs. Search transforms the Internet from a confusing mess to the most useful informational resource of all-time.

7. An impostor could not pretend to have a good search engine. Search is hard. You can't provide good search unless you really have crawled large swathes of the Internet and done something intelligent with what you've seen. It's also somewhat verifiable. If a search engine gives junk results, people will recognize that the results are junk. It's easy to test a search engine out on the topic most important to you. Thus, there are substantial meritocratic components in people's use of search engines, and we should be reasonably confident that a successful search engine really does offer recommendations that people find useful much of the time.

8. Search engines could keep secrets from us and we would never know. Precisely because search is hard, it's easy to play games with a search algorithm. Merely by inspecting results, there's no way to prove that a search engine has demoted a site in its rankings for illegitimate reasons. There are just too many other reasons why the site might appear where it does. We can probably count on search engines not to be deliberately incompetent, but for any given ranking, distinguishing incompetent from malicious from opinionated from truth-telling is a nearly impossible task for anyone except the search engine's own programmer. This is not a problem that can be completely solved through pure market forces.

9. Search engines can play favorites. This point is really just a special case of the last. So many different considerations go into a search engine algorithm that some degree of favoritism is inevitable. Whether consciously or unconsciously, the search engine will be more useful to some users than to others. Equal access to search is ultimately as important as equal access to other informational resources, and good information policy will make sure that a reasonable baseline of search is available to everyone.

10. The more search engines the better. These may seem like dire fears. But many of them are much less worrisome if there is effective competition among search engines. The engines themselves can learn from what the others do right; users can mix and match to optimize their searches. Comparisons can point out general areas of trouble and police against particular abuses. A diversity of search engines will favor individual users' ability to get the particular search services—and thus the information—they need.

* * *

The Library of Babel provides an exhilarating and frightening metaphor for the Internet. Exhilarating because it reminds us that we are all now "the possessors of an intact and secret treasure"[59] of knowledge beyond compare. Frightening because it reminds us that that knowledge is shut away in a "feverish [place], whose random volumes constantly threaten to transmogrify into others, so that they affirm all things, deny all things, and confound and confuse all things, like some mad and hallucinating deity."[60] Only the god-like Book-Man, whose knowledge of the Library is an "honor and wisdom and joy,"[61] can make sense of it for us. In the Library of Babel, the Book-Man is but a "superstition,"[62] but on the Internet, his name is Google.

59. *Id.* at 115.
60. *Id.* at 117.
61. *Id.* at 117.
62. *Id.* at 116.