



DIGITAL UNIVERSE JOURNAL

ARTICLE 2006-1



The Future of Free Information

by Lawrence M. Sanger

A free online journal that fosters discussion concerning the dissemination of knowledge for the benefit of mankind, published by the Digital Universe Foundation.

The Future of Free Information¹

LAWRENCE M. SANGER

*The Digital Universe Foundation
100 Enterprise Way, Suite G-370
Scotts Valley, CA 95066*

Abstract

The ideal information resource would feature high quality of content (*i.e.*, be accurate and complete) as well as high accessibility (*i.e.*, excel in availability, ease of use, and interactivity). This very programmatic paper first describes these various features and their implications. Then it applies the set of features to some extant resources, arguing that the ideal information resource does not yet exist. The paper speculates that, in the future, there will be little debate whether a startlingly new and better information resource is possible, because that much will be taken for granted; the debate will concern what the resource's main features should be. Aiming to foster this debate, the paper concludes with a list of topics needed to be addressed to fully justify an answer to the question, "What would the ideal information resource look like?"

1. Introduction: The Ideal Information Resource

What would the ideal information resource look like?² Let me begin by clarifying what I mean by this question. By "information resource" I mean the sort of thing from which one can get a piece of human-communicated information that has some reasonable presumption of reliability.³ So consider some main examples of what various information resources there are: human beings such as librarians and experts; nonfiction books and all sorts of media, visual, auditory, and otherwise; and various *repositories* of these informative people, books, and media, such as libraries and the Internet.⁴

When I ask, then, "What would the ideal information resource look like?" I am asking you to imagine the finest feasible *single*⁵ information resource (surely no such resource now exists). It is not, of course, a single person, because a single person can reach only so many people, and knows only so much. Nor is it a traditional library, much less any single book, because only a limited number of people can access any single library.

Running over these few ideas, it seems clear enough that there are several conditions that tend to make an information resource better, whether or not they are jointly possible—although I see no reason to think they are not jointly possible. The more that an information resource is complete, accurate, available,

easy to use, and interactive, the better it is, or so I will argue.

These five characteristics of an ideal information resource can, I think, be usefully grouped under two more general heads, *quality* and *accessibility*, which are individually necessary and jointly sufficient to capture the notion of an ideal information resource that I have in mind.

High quality of content is essential because we want an information resource to provide us the means of gaining *knowledge*—which is what I take an information resource to be *for*. So I say that an information resource is of a higher quality the more conducive it is to the gaining of knowledge⁶ rather than mere uninformed opinion or, say, random unintegrated data and unassimilated facts.⁷

Accessibility is essential because an information resource can fulfill its function of facilitating knowledge only so far as people can and want to use it to get knowledge. We can say that while quality ensures that there is something to learn from, accessibility ensures that people actually learn it. So I am using "accessibility" in a very broad sense, to include even such features as coherent organization and engaging presentation. Mere *availability* is not sufficient.

While I am reasonably sure of these two broad heads, I am less sure about precisely what belongs under them. Next, then, I will elaborate on the above-listed five subheads, or characteristics of an ideal

information resource, with even more characteristics marshalled beneath the subheads. Afterwards I will evaluate a few current information resources in terms of how well they square with this ideal, and conclude with a list of some issues that need to be resolved in order to adjudicate the question of what an ideal information resource would look like.

2. Features of the Ideal Information Resource: High Quality

There are two features that determine the *quality* of an information resource: its accuracy and its completeness.

(1) *Accuracy* is perhaps the most obvious requirement, but what *is to be counted* as accurate is a thorny issue that would, no doubt, distinguish different conceptions of an ideal information resource. The free, user-created encyclopedia Wikipedia has defenders, for example, who claim that one can do little better than give all of humanity equal editorial access over the same body of information—thereby repudiating the notion that expert⁸ approval is necessary.

But it hardly needs argument (at least not in a paper of this size) to maintain that some sort of robust expert involvement and leadership will increase the accuracy of a resource. It would be more precise (and less controversial) to say, however, that expert involvement will increase a resource's *faithfulness to expert opinion*. How accurate expert opinion is as a body is a distinct question—and an issue, to be sure, that may inform the debate about what the ideal information resource looks like, as the case of Wikipedia makes all too clear. But, again, this paper is not the place to adjudicate that issue.

Still, all hands will admit that experts (under any useful description) can be mistaken, and this is one reason to think that a maximally accurate information resource should be *reviewed*. Indeed, the more widely reviewed it is—so long as there is a robust mechanism for the reviews to improve the end result—the greater the accuracy. We might say, then, that accuracy increases with *distributed editorship*.

Finally, even large bodies of experts reviewing a work can be systematically inaccurate, not only with respect to the facts themselves (which is obvious enough—consider any number of once-popular and now-debunked intellectual fads) but with respect to the full range of expert opinion.⁹ The ideal information resource, to represent expert opinion faithfully, would have to be *independent* both of any particular group in a given field and of corporate (including both

commercial and governmental) interests. In short, it would have to be *neutral* and *fair* to the broad range of opinion within a field.¹⁰ Otherwise, nonexperts could not trust the resource, even if it were faithful to the facts; from an outsider's point of view, there is just a puzzling disagreement among experts, and a resource that favors one group and omits the views of another will appear less valuable than a resource that treats the competing views more neutrally, fairly, or with equal sympathy.

(2) *Completeness* appears necessary for high quality. It seems one needs only add some missing piece of (“accurate”) information to a general information resource, and one increases its quality. Completeness is desirable in an information resource because a high degree of completeness is necessary for certain kinds of research. Sometimes only a truly exhaustive resource will do—as in the case of a patent database, where the decision to grant a patent depends on the conclusion that no *similar* prior art already exists. Without an exhaustive database, such a conclusion is not possible.

In discussions about Wikipedia online, some have taken the view that Wikipedia contains too much information, e.g., about obscure celebrities, video games, and science fiction universes. I find this view puzzling; so long as the information really is accurate, then there seems to be no advantage in excluding it. I suppose the view reflects an attitude that seems to be a holdover from the days in which the length or size of a resource was determined by publishing and other space concerns. But if the resource is digital, then of course there is no reason to truncate it artificially simply on grounds that a subject is rather more obscure than what can ordinarily be found in a general encyclopedia. Disk space is cheap.

There is some information, of course, that might be thought to be harmful to be included in an information resource, such as libel, pornography, speeches inciting violence, and instructions for making bombs or weapons of mass destruction. Another question is whether there ought to be a database listing every human being on the planet—if such a thing were possible. It seems that at some point, the interest of ordinary persons to their privacy might trump the interests of humanity in having a *truly* exhaustive database. For, even now, if the will of humanity and legal opportunity were there for it, we *could* have a public database listing every human being on the planet; there are no *technological* or *logistical* impediments. But, as no clear advantage would come of it, it is very

doubtful to me that there will *ever* be sufficient will for such a thing.

When I say that the ideal information resource should be complete, then, I mean complete with regard to *at least* responsible (not harmful) and general information. I take it that this might exclude whole types of information that it would be an infringement on privacy to include—such as a complete roster of living humanity—or that it would be impracticable to include.

There are two other things I would like to say about the aim or goal of completeness: the fulfillment of this aim requires both many people and many different *types* of information. The former first. Completeness in the above-described sense clearly requires maximum participation by both experts and the general public. Editors of specialized volumes and academic hiring committees know that frequently only a specific type of person will do, and very few of that type are available (sometimes it is a type that has one, or no, members). Furthermore, as Wikipedia well demonstrates, the general public far outstrips the expert community alone in terms of its available time, motivation, ability to keep much information up-to-date, and ability to write about, say, *Star Trek*. What we might say is that, until there is a quorum of expert and public participation, a general information resource simply cannot aspire to being nearly as complete as it could be, with vast participation from the entire spectrum of the educated population and from across the globe.

The ideal information resource would also be complete with respect to the range of information *types* available. It would be, of course, more than an encyclopedia; it would include books, website data, curricula (educational materials of various kinds), raw scientific data, photos, video, audio, software, and more. Libraries and the Internet are two possible examples of resources containing such diverse types of information. But, as I will explain, libraries lack accessibility, and the Internet at present lacks adequate quality.

3. Features of the Ideal Information Resource: Accessibility

There are at least three distinct features of an information resource that increase its “accessibility” (in the sense I mean): availability, ease of use, and an interactive community.

(3) *Availability*. The more widely available a high-quality resource is, the better it is in terms of the im-

pact it has. Availability is best achieved by making the resource both as *widespread* and as *inexpensive* as possible. If the resource is digital and accessible on the Internet, that will make it as nearly widespread as it can be (although, given the so-called digital divide, not as widespread as we would like it to be). Moreover, if it is open content—free “as in freedom,” as the phrase among open source advocates goes¹¹—then not only is it free of cost, it has the opportunity to be developed further and to become a self-perpetuating institution of free knowledge.¹² An open content license guarantees availability not only across income levels and physical space, but across time as well.

Outside of a few corporate talking heads and curmudgeons, there has been little opposition to open source and open content, probably because there is no good reason to be opposed to making freely-distributed information as widely available as possible—but also because the profits of proprietary projects have not yet seen much threat from these projects. It seems unlikely that *all* of the world’s information will be open content in the future; as long as authors, artists, and coders perceive no other viable model but traditional intellectual property to support their work, many of them will be opposed to simply “giving away” their work. But increasingly large segments of academe, government, and the general public, whose livelihoods do not depend on payment for specific pieces of work, have shown themselves to be perfectly willing to release their work under a license that makes it as widely available as possible. This trend is thriving.

Widespread availability imposes another sort of requirement, namely, that the ideal information resource be made available in the widest possible range of *languages* that are being read online. Wikipedia has, again, demonstrated one way how this might be done.

(4) *Ease of use*. No matter how high-quality and widely available it is, an information resource can easily slide into disuse or obscurity if it is hard to use. (For example, this is one reason that Google’s simple, effective search became more popular than Yahoo’s directory structure.) Consider: commitment to quality can create a valuable resource; making it available lowers the barrier between the resource and the user; but the ideal resource must take another step and as it were bring *the user* toward *the resource*.

One might well say that some relatively “complete” information resources—large university libraries and the Internet—are not *particularly* easy to use. But precisely the massiveness of an information

resource is at odds with its ease of use, and this surely places *some* constraints on how easy an information resource can be to use. Of course, design that improves based on public (*i.e.*, user) feedback can solve some of the more obvious problems. This is another reason to have an *open* project: it will give the public more incentive to help designers with usability testing.

A further difficulty with nearly all information resources is the fact that they lump together material for various levels of educational attainment. This is not the case, of course, with some encyclopedic and especially some educational resources, but it is certainly a difficulty with libraries and the Internet in general. Many a parent has wished for a reliable and vibrant “Internet for children,” but no such thing exists yet.¹³ In any event, a resource can be made more useful if it marks off what level(s) of background a user needs to have in order to appreciate its offerings fully.

“Hidden” information becomes easier to use as it becomes more *findable*. Google and other advanced search engines, as well as online library search tools, have solved this problem to a certain extent. Moreover, libraries use cataloging systems, such as the U.S. Library of Congress system, which allow readers to find related information serendipitously while browsing the stacks. But the Web directories of Yahoo! and the Open Directory Project (www.dmoz.org) notwithstanding, and with deep respect for their creators’ achievements, there simply is no *very* useful only taxonomic directory yet. Moreover, what does not yet exist is an information tool that places all different *types* of information about a single topic in a single “place” or “portal.” A new free, collaborative information project, the Digital Universe (www.digitaluniverse.net), will attempt to improve on previous resources by placing the most authoritative information *of all* sorts in topic-oriented portals which are, in turn, arranged taxonomically. Whether the project will succeed at this ambitious aim remains to be seen.

In addition, information becomes easier to use when its presentation is made more engaging or attractive. This can be facilitated by a simple yet well-designed interface, but what might really enhance the presentation of information is full integration with multimedia. The easier it is to see and hear what one would otherwise only be reading about, the better. Furthermore, the modeling of information in explorable 3D spaces—long anticipated, and now brought to a popular audience by NASA’s World Wind

and Google Earth—has the potential to completely change the way we interface with information.¹⁴

(5) *Interactivity*. Building communities around information resources can grease the wheels of interaction between users and the resource. Librarians play this role with respect to their libraries, and, if a school or university can be considered an information resource,¹⁵ instructors play this role with respect to their texts and other material. Sometimes, having more or less direct contact with an authority who as it were “stands behind” the resource is the best way to drive home a point. As Hubert Dreyfus (2001) well points out, direct human contact has a way of imprinting information and values that information imparted at a distance, over the Internet, cannot. Moreover, digital communities formed around an online resource can teach each other enormous amounts. But to ensure that the interaction is helpful, as befits an ideal information resource, there must be *safeguards against abuse*.¹⁶

Perhaps the most salutary aspect of interactivity is the fact that the resource itself can be improved through interaction, through what I have elsewhere (2005a) called “radical collaboration.” As the Web goes from “read-only” to “read-write,”¹⁷ it becomes more interactive. This redounds to the benefit not only of the participants doing the interaction, but also of the accuracy and completeness of the information in the resource—and thus benefits future users as well. Wikipedia is, at least in theory, probably the best example of this online; but universities, again if they can be considered information resources, are perhaps the best illustration of all, as interaction between faculty and students, and faculty with each other, leads to the addition to and improvement of the teaching resources created by the faculty.

4. The Ideal Information Resource, Summarized

An ideal information resource would, on this account, be an Internet-mediated project (including but not necessarily limited to a website) featuring maximum involvement by both experts and the general public, working together to create the widest range of information tools—an online library, web directory, encyclopedia, and so forth—enhanced with much multimedia content. The project would be free of commercial influence, and the content would be neutral, maximally free (open content), arranged into taxonomically-sorted portals, and available in multiple languages and education levels. Experts would be

leading the effort, but editorship would be widely distributed, and the public would have ample opportunity for both contribution and feedback. The experts themselves would be made available to the public, and everyone would be organized into a community of mutual aid, but with many safeguards against abuse.

I advance this just as a *suggestion* of an answer, by way of setting out what I *mean* by the question, “What would the ideal information resource look like?” A full elaboration and defense of my answer would require many more pages, and I cannot pretend to have done any more than simply sketch out and barely motivate one possibility. But I hope that, in having done at least this much, I have not only elaborated what the question means, I have gone some way toward explaining why it is important and worth thinking about.

5. Precedents and Prototypes

This paper’s considerations are not merely an impractical philosophical exercise. Having long since become aware of the potential of the Internet to work an information revolution, and, more recently, having become exposed to the power of asynchronous, distributed collaboration, various individuals and groups have started a wide variety of collaborative content creation projects. The media finally discovered this trend during the 2004 U.S. national elections, as blogs were touted as significant political tools, and as Wikipedia’s success became hard to ignore. As a result, more and more people are not just getting online, but getting together online, to create things. Increasingly I see evidence that people are asking themselves: “If we collaborate online, what is possible?” And more specifically, they ask,¹⁸ “How can we use online community tools, like wikis, to solve our problems or to achieve our goals?” This is a good question to ask, but it is not the best; there is a question that excites me much more.

The better question is the one addressed in this paper, *viz.*, what the ideal information system would look like. This is a *philosophical* question, and the best answer is apt to take the form of a set of principles. But in designing their projects, it seems that many people, rather than follow principles where they lead, prefer to take their cues from what they regard as models, or prototypes; then they apply those prototypes to their own problems or goals, with varying success. So what I want to do is to consider a few such prototypes and why they are less than ideal (according to the principles laid down earlier). The point

here is not merely to be critical, but to illustrate those principles further and justify why, perhaps, the model described in Section IV above is a natural development of all of these prototypes.

Let me begin with some more traditional information resources. Traditional proprietary encyclopedias, such as *Encyclopedia Britannica*, are passably accurate but very far from complete. Indeed, to compete with the amount of information available on the Internet—albeit, granted, often of dubious provenance—one would need to add the entire reference section of the library. Furthermore, subscriptions to digital encyclopedias require an investment that some households will find prohibitive, and hence such encyclopedias are not as readily available as one might like. Traditional multimedia libraries and archives are considerably larger (usually containing multiple encyclopedias) and are free of charge. But their primary failing is that they exist in particular places, and so they are not as widespread as one might like: they again fail the availability test. The Internet as a whole is enormous, mostly free of charge, and available wherever there are computers with Internet connections. Its main problems concern quality. Though it is so diverse that one dares not make generalizations, I will anyway: on average one finds relatively poor accuracy, and even weakness in terms of completeness, since (depending on the field) one must sometimes search very hard indeed to find much of the very specialized information one can find with ease in a good university library.

Commercial search engines and Web portals, like Google and Yahoo!, ameliorate these problems only a little. They do often place high-quality websites on the first page of a search. But Google’s algorithms—based on the insight that, if a website is linked more often, then it is better—are a measure of a kind of popularity, which must not be confused with authority or reliability. Yahoo! now has a similar search engine, but it built its reputation on its Web directory, and it is now becoming increasingly focused on community-building. To my knowledge, neither in its directory services nor in its community-building are many relevant subject area experts given any special role. I see little evidence of such roles being played, anyway, even if they are. Furthermore, even with the success of Google’s search algorithms, if one is looking for a very specific piece of information, search engines and Web portals may not be able to find it quickly, easily, or at all. After all, they can find only what exists online, and very far from all of the information we are looking for is online. The Internet

makes a rather poor showing compared to good university libraries (which, naturally, include terminals connected to the Internet, and thus as it were *contain* the Internet), simply because there is more specialized information in the library. For researchers and diligent students, living at the library is still the best way to put one in touch with the best information.

This brings us to Wikipedia, which is perhaps the most instructive example for us to study when thinking about the question what the ideal information resource would look like. Wikipedia is a free online encyclopedia and has recently added its one millionth article in English, and it boasts over three million in all languages combined. Its sheer size allows it to make a credible claim to be solving the specialization problem: a frequent observation in blogs and news reports about Wikipedia is that, for some topics, one can find specific information more quickly in it than through any other source, Google included.¹⁹ It is remarkable in that it is written by its users *via* a Web technology that allows anyone to edit any page on the spot—it is a so-called *wiki* website—and thus it has developed an interactive community. Wikipedia is also open content, which allows anyone to use and further develop Wikipedia's content free of charge, thus increasing its availability. Because so many people in the world want to teach each other, and because they are motivated to do so especially if their work does not profit any person or business, in this system in which contributing to an encyclopedia is so easy and quick, Wikipedia has grown its millions of entries in just five years.

Wikipedia's parent organization, the Wikimedia Foundation, also manages a number of other information projects, making use of wikis—such as collaborative book-writing projects and a collaborative dictionary. Thus, one might say, the Wikimedia projects are aiming collectively at something at least *like* the ideal information resource.

In an incident now well known, however, the distinguished journalist John Seigenthaler Sr. discovered that defamatory claims were made about him in his Wikipedia biography. He publicly denounced Wikipedia in a column (Seigenthaler, 2005) that sparked a firestorm of controversy. The much-debated question was whether a system so open to defamation and error could be trusted. In the aftermath, a news article from the leading scientific journal *Nature* (Giles, 2005) found that, when 42 articles on scientific topics from *Encyclopedia Britannica* and Wikipedia were compared, the *Britannica* articles

had an average of three errors, and Wikipedia articles, four.

This investigative news report—for it was not a peer-reviewed study—was then used by Wikipedia's defenders to rebut the increasingly shrill charges that the user-built encyclopedia was unreliable. *Nature* had confirmed that Wikipedia was almost as good as the *Britannica*, Wikipedia's defenders said; so all the talk of its unreliability was just elitist bigotry. The difficulty with drawing this conclusion, however, is that there was one simple metric in the study—"number of errors"—and, more importantly, the articles examined were all on scientific subjects. It is not the slightest bit surprising that Wikipedia's articles on scientific and technical subjects are relatively good, for the obvious reasons that more technically-minded people are apt to write for an Internet encyclopedia and that there are relatively well-agreed facts in science and technology. With virtually all other subject matters—the arts and humanities, for instance—one justifiably suspects that matters are quite different.

Wikipedia's entries are indeed, on the whole, quite a bit better than what one might expect from such a wide-open project. Whatever one might think of its reliability, it is clear that Wikipedia can serve as an example to the world of what can be achieved by a very open worldwide collaboration on a free information resource. Wikipedia's success demonstrates that much of the ideal outlined above is possible. Consider: Wikipedia features extensive involvement by the general public (and some experts are involved), and they are working together to create a wide range of information tools, and first and foremost, an encyclopedia. Editorship is widely distributed, and the public has ample opportunity for both contribution and feedback. The project is free of commercial influence, and the content is or at least aims to be neutral, maximally free (open content), and available in multiple languages. This is not the whole ideal; but it is much of it. So it is not at all surprising that Wikipedia should have such strong defenders.

Even excellent things can often be improved. To create the ideal information resource, the Wikipedia *model* (I do not say Wikipedia itself²⁰) would have to be extended in several ways. It would:

- 1) feature expert leadership and much-expanded expert participation;
- 2) extend participation to an even larger cross-section of the educated public than now feels comfortable participating;
- 3) extend the scope of the projects to the contents of traditional libraries and archives (with the

contents frozen in reliable versions, not left on wikis to be further edited);

- 4) more closely integrate the different information projects so that information is sorted by topic into taxonomically-arranged “portals” about the topics; and
- 5) feature content at and sorted usefully into multiple educational levels.

Some of these problems (particularly (1) and (3)) are addressed by a very exciting, relatively recent development: the Open Content Alliance. This appears to be the first serious effort by a consortium of major libraries, corporations, and other institutions and organizations to bring the contents of (largely copyright-free) library contents to a much broader audience. The current collection—found on www.archive.org—is wonderful in point of availability, but not (at present) so much in point of accuracy, ease of use, completeness: parts of works are missing, the collection is not quite as deeply searchable or well organized as one would like, and there are still many holes in the collection. But surely fixing these problems is only a matter of time. What might turn out to be a more serious (long term) problem with the OCA is its relative lack of interactivity—*i.e.*, it does not seem there are any plans to allow either experts or the public to interact with and improve upon the information in the database in a collaborative way. Although it is possible to contribute materials to it, it is for the most part still a read-only Web. It is, after all, an archive—something very useful indeed, but not the ideal described above.

Making up another sort of information resource are the many professional and academic Web projects that have sprung up in the last decade or so, such as PubMed Central in Medicine (www.ncbi.nlm.nih.gov), the Perseus project in Classics (www.perseus.tufts.edu), and the *Stanford Encyclopedia of Philosophy* (plato.stanford.edu). These are peer-reviewed resources that excel in representing expert opinion and research. Within their fields and missions, they are more or less complete (PubMed has a reputation for comprehensiveness, while the *Stanford Encyclopedia of Philosophy* still has many gaps), but virtually none is complete with regard to the types of information available in their disciplines. It seems unlikely that a compilation of all professional or academic Web project contents, taken all together,²¹ would be any more complete, and would probably be much less so, than Web content that is not part of any such academic project. Moreover, as any active researcher knows, such online resources simply are not a substitute for

library and archive research. Furthermore, while much academic information online is free of charge and likely to remain so, and thus adequately available, it is very frequently prepared by academics *for their peers*, not for the general public. Even further, the presentation of information on academic websites is frequently unexciting and not particularly easy to navigate; some of these websites even require registration for access. Hence the information is not particularly easy to use for most people.

Table 1 summarizes these observations of Part 5 so far. This table must not be taken too seriously; it is personal and impressionistic. Certain ratings are explained in the text above and would likely require significant discussion to explain and justify. For example, I rate university libraries and academic Web projects a “3” for accuracy when, of course, very many of the books found in libraries are full of inaccuracies. My point in giving these resources these high ratings is not to say they are sources of objective truth, but rather to say that they provide the best representation of expert opinion (some of which is, of course, wrong). A more informative table would actually break the five features listed into subcomponents.

Table 1. An assortment of information resources and their fit with the ideal information resource as defined above

	Accuracy	Completeness	Availability	Ease of use	Interactivity
Proprietary encyclopedias	2	1	1	3	1
University libraries	3	3	1	2	2
The Internet (as accessed by Google, Yahoo! and other search engines and portals)	1	2	3	3	3
Wikimedia projects	1	2	3	3	3
The Open Content Alliance	2	1	3	2	1
Academic Web projects (taken together)	3	2	3	1	2

Key: 3 = near ideal; 2 = good, but needs improvement; 1 = poor

As personal and impressionistic as this table is, it does illustrate a useful point: none of these information resources is in fact ideal across the board. What the world needs is a single information resource that has “straight 3s.” Moreover, I assert—without any attempt to defend this assertion—that an information resource with straight 3s is feasible. When and if the institutional will to create such a resource is found, it will come into existence. I am aware of no compelling technical, economic, or social reason to think that such a resource *could* not come into being. In fact, I think that it is very likely that such a resource *will* come into being as soon as the idea of the thing is propounded compellingly to those who can make it happen.

In fact, there is an Internet project that has a strong claim to being at least aimed at this ideal, though it is far from having reached it: the Digital Universe (www.digitaluniverse.net). It aims at accuracy in the sense described above, because it aims to amass large and diverse governing bodies from every field, led by genuine experts, including some academic luminaries. Its principals have committed publicly to being neutral and the project’s managing organizations are, or will be, nonprofit and noncommercial. It is also committed to building a large body of contributors, both expert and from the general public, and to host or link “deeply” to reliable information of every type that exists online. So the Digital Universe aims to be a very high-quality affair.

The Digital Universe also aims at high accessibility. Most of the information and supporting software will be free (open content and open source), although there will be some premium services to help pay for expert labor and for the cost of the platform. There will, in time, be versions in all major languages online, and at every level of educational attainment. Due to its heavy public involvement, there will be ample opportunity to perfect the user interface, to make it as easy to use as possible. Information will of course be searchable, but in addition, in portals devoted to particular topics, all *types* of information will be cataloged, with the topics, or portals, further arranged taxonomically. This Web experience will be married as much as possible to cutting-edge 3D and multimedia ways of navigating information.²² Finally, the Digital Universe aims to build an interactive community that both engages the public directly and brings experts and expert-vetted information to the public. The hope in general is that the Digital Universe will be very easy to use, and more generally, very accessible.

Whether the Digital Universe will be able to bring off such a high-minded and ambitious affair remains to be seen. What I think is likely, in any case, is that *some such* information resource will come into existence, as its broad outlines and its desirability become increasingly obvious to the leaders of academia, education, the Internet, and publishing.

6. Conclusion: The Coming Debate About the Future of Free Information

I conclude with some speculations about a coming debate over the future of free information.

As the world, increasingly interconnected, increasingly aware of the power of digitization, open content, and radical collaboration, awakens to the dumfounding possibilities before it, I believe the central question of this paper—“What would the ideal information resource look like?”—will come to the fore. Ultimately, the debate will not concern whether a startlingly new and better information resource is possible, because in time (perhaps very soon) that much will be taken for granted; the debate will concern what its main features should be. Furthermore, I think this debate will turn out to be both important and global, because humanity in concert, with its relative shortage of expertise in many specialized subjects, will be able to support only a small number of “super-projects” aiming at anything like the ideal information resource envisioned here. Ultimately the question will become: what sort of massively collaborative information projects deserve the support of the intellectuals of the world?

In the discussion above, I rather quickly passed over a number of extremely difficult issues—not, of course, because I thought they could be *settled* so quickly. Indeed, these issues need some careful, sober, well-informed philosophical thinking. If this essay has any lasting value, other than in identifying an interesting question—what the ideal information resource would look like—I hope it might be in the identification of the set of issues that need to be adjudicated in order to justify an answer fully. This is surely only a partial list:

- For purposes of developing an ideal information resource, should experts be given any special authority? And, if so, who counts as an expert? Should we simply try to get clear on what the conventional notion of what an expert is, or instead attempt to apply some new conception

of expertise? What sort of conception, if a new one?

- Is neutrality or fairness required for an information resource that aspires to be ideal? If so, in what does neutrality consist, and how can it be safeguarded? How widely should the net be cast in order to fulfill the requirements of neutrality? Should the full range of *expert* opinion be represented, or the full range of *all* opinion—or something more subtle than either of these?
- To be ideal, should an information resource be *absolutely* exhaustive or complete? Or should it exclude any information on grounds that it is harmful in some way? If so, harmful in what way(s)? Where, exactly, should the line be drawn? Also, is there some information that is simply too *trivial* or *poor quality* to be included?
- Are schools and universities to be considered as information resources for purposes of this discussion? That is, should our notion of the ideal information resource extend to education—*not* just educational materials, but education itself?
- Given both that the ideal information resource would be free and that the world’s “knowledge workers” need to be paid for their work, what financial model can be found for it? Should governments support it?
- Should there be a single global information resource for all languages managed by a single organization, or should different projects be started in different languages, with the best examples being “franchised” under independent management in other languages?
- What features in general make an information resource very easy to use?
- Would the ideal information resource include a “walled garden” of trustworthy, responsible information for children?
- How is information that differs by subject, type, educational level, and quality best sorted and found?
- Are 3D and multimedia “spaces” improvements on more traditional methods of information presentation? How important is it that we begin to navigate information in these ways?
- What sorts of safeguards against abuse are necessary to keep “healthy” a community that is organizing an ideal information resource? What concerns about user rights are relevant, and how are those concerns to be adjudicated?

References

- DiBona, Chris, Danese Cooper, and Mark Stone, eds. (2005). *Open Sources 2.0: The Continuing Evolution*. O’Reilly Media.
- Dreyfus, Hubert L. (2001). *On the Internet*. Routledge.
- Giles, Jim (2005). “Internet encyclopedias go head to head.” *Nature*. Published online Dec. 14, 2005. <<http://www.nature.com/news/2005/051212/full/438900a.html>>
- Gillmor, Dan (2004). *We the Media: Grassroots Journalism by the People, for the People*. O’Reilly Media. <<http://www.oreilly.com/catalog/wemedia/book/index.csp>>
- Jones, Pamela (2005). “Extending Open Source Principles Beyond Software Development.” In DiBona, et al., 273–80.
- Moore, Adam D. (2005). *Information Ethics: Privacy, Property, and Power*. University of Washington Press.
- Newman, John Henry (1873). *The Idea of a University*. Accessible at <<http://www.newmanreader.org/works/idea/>>
- Sanger, Lawrence M. (2001). “Neutral Point of View–Draft.” *meta.wikimedia.org*. Accessible at <http://meta.wikimedia.org/w/index.php?title=Neutral_point_of_view-draft&oldid=756>
- Sanger, Lawrence M. (2005a). “Why Collaborative Free Works Should Be Protected by the Law.” In Moore (2005), 191–206. Accessible at <http://www.geocities.com/blarney pilgrim/shopworks_and_law.html>
- Sanger, Lawrence M. (2005b). “The Early History of Nupedia and Wikipedia: A Memoir.” In DiBona, et al. (2005), 307–38. A slightly different version is accessible at <<http://features.slashdot.org/article.pl?sid=05/04/18/164213>>
- Seigenthaler, John (2005). “A false Wikipedia ‘biography’.” *USA Today*. Published online Nov. 29, 2005. Accessible at <http://www.usatoday.com/news/opinion/editorials/2005-11-29-wikipedia-edit_x.htm>

End Notes

¹ I would like to thank Eric Saudete, Tereza Sena, and their colleagues and the attendees of the 2005 Macau Ricci Institute Symposium: “History and Memory—Present Reflections on the Past to Build Our Future.” The theme of the day was “Preserving Memory and Teaching History.” Some ideas from an earlier version of this paper were presented in a panel discussion on “History and New Technology,” and this paper is forthcoming in *Chinese Cross Currents*. A very slightly updated version of the paper was presented at Purdue University on March 28, 2006. I would also like to thank my colleagues with the Digital Universe project (with which I am currently employed) and with Nupedia and Wikipedia for much insight over the years, which has found its way into this paper.

² This problem was relevant to the topic of the Macau Ricci Institute Symposium in a perhaps roundabout way. The problem of how best to preserve the memory of the human race to a large extent overlaps the problem of what sort of resource is best suited to organize and access information. For, after all, the information that humans produce in a sense constitutes the “memory” of the human race. The activity of cataloguing, archiving, chronicling, organizing, and otherwise aggregating human-generated information thus has the function (among others) of preserving human memory. If, then, one takes as a starting-point the question how best to preserve the memory of the human race, one might well begin by examining what the ideal information resource would look like.

³ Thus, thermostats are not information resources because their data is not communicated by humans, nor are (arguably) novels taken alone simply because they make no claim to accuracy. But a database containing the precise published texts of novels would be an information resource because there would be some presumption that it contained reliable versions of texts.

⁴ And perhaps schools and universities, but I will be discounting this possibility in what follows.

⁵ I do not comment here on how to individuate information resources, although perhaps I should. I could say that the limits of an

information resource are determined by where or how it is accessed. Note that I am willing to consider as “single” information resources both libraries and encyclopedias (contained in libraries), and both the entire Internet and specific Internet websites and projects. But I would not consider certain disjunctive sets of resources as “single” resources. For example, I would not say that the set containing the *Britannica* and *Wikipedia* is a “single” resource.

⁶ This is not to say that an ideal information resource would contain only objective truth. That would not be feasible, and in speaking of the ideal information resource, I do mean something feasible. Perhaps the highest practical goal at which an information resource can aim is the neutral, fair presentation of the entire spectrum of expert opinion. Note also that an information resource could be “conducive” to getting knowledge in various ways—not just in the straightforward way of, for example, reading and believing.

⁷ In drawing the distinction between a resource that is conducive to knowledge, or even “wisdom,” and one that is conducive more to unjustified opinion and unintegrated memorization of facts, I suppose I am saying that an ideal information resource has an aim similar in that respect to the aim of university education, according to John Henry Newman in *The Idea of a University* (Newman, 1873), Part I, Discourse 6 (and elsewhere). This is a distinction that makes a difference; it has various possible implications for the design of an information resource.

⁸ I do not here intend to offer an account of expertise, but I will say this. The notion of expertise I have in mind here is a very conventional one, and is measured by degrees, certifications, and other solid evidence of attainment—not necessarily in that order. In fact, more important for the old-fashioned notion of expertise I am working with is long, focused study, experience, professional-level conversations that *comes after* an in-depth but broad grounding in a discipline. It hardly needs saying that experts can be wrong. The more interesting question is *to what extent* expertise in this conventional sense improves the probability of a person’s testimony being true.

⁹ One might go further to say that an ideal information resource should represent the full range of *opinion*, period, about a subject, whether “expert” (on any conception of expertise) or not. This is yet another fascinating possibility that I do not have space to discuss here.

¹⁰ The sense of “neutral” I mean here is unobvious; it is the one I articulated on behalf of Wikipedia (see Sanger, 2001).

¹¹ Free software advocate Richard Stallman is particularly well-known for this locution.

¹² In Sanger, 2005a, I argue that “shopworks,” or free, radically collaborative works, form the basis for a new sort of institution that is so valuable that it should be protected by the law.

¹³ There are, of course, “walled gardens,” websites or services that specifically restrict access to other websites or services, as well as filters. What does not exist, however, is an entire network anything near to the size of the Internet, made specifically for children. Perhaps *that*, I

am saying, should exist. But I do not mean to claim that such a thing should *replace* the Internet.

¹⁴ This is another feature of the Digital Universe planned by co-founder Joseph P. Firmage.

¹⁵ The trouble with such a claim, however, is that it blurs a useful distinction between persons qua *persons* and persons qua *information resources*. Persons (qua persons) are essential to a school or university. A person is not information, although one can get information from a person. It is more helpful, and does not blur this distinction, to say that various *educational material* would be contained in the ideal information resource, such as lectures and discussions, or recordings thereof. The role of persons in an information resource *qua* information resource is to help bring people to the information. This might include some functions very like teaching, but insofar as, in a system, a person him- or herself is the focus rather than the guide of learning, the system includes an element of “schooling” beyond the mission of an information resource. Note that if a university *were* to be considered an information resource, then one might want to offer a much-expanded answer to the question, “What would the ideal information resource look like?” And then the question really goes beyond philosophy of information to philosophy of education.

¹⁶ This essay cannot hope to explore the issues that must be addressed to set up a really healthy community. Some pitfalls to avoid can be found in the experience with, *e.g.*, Wikipedia (see Sanger, 2005b) and Groklaw (see Jones, 2005).

¹⁷ See Gillmor, 2004, Chapter 2.

¹⁸ I have been approached many times over the past few years by people asking just this.

¹⁹ Although, speaking for myself, when I expect some information to be most quickly located on Wikipedia, I *use Google* to search Wikipedia. So in my case, the claim more precisely stated is that it is more difficult to find some specific piece of information using Google *without* Wikipedia than by using Google *to search* Wikipedia. A good search engine, in any case, is essential to finding the information.

²⁰ Bear in mind that I am not suggesting that Wikipedia itself actually change in these ways. I am not sure that, given the self-selecting nature of its community and its consequent strong commitment to a sort of intellectual egalitarianism, its community could agree to (1)–(3).

²¹ And it would be dubious to consider this a *single* information resource anyway: such a consortium of academic projects would have to be created to make a unified entry point.

²² As of this writing one must use a special Web browser to view the Digital Universe, and registration is required. These are admittedly barriers to access. Consequently, in the Spring of 2006 a “browser-neutral” version of the Digital Universe—*i.e.*, a specially-designed website that will work in any browser—will launch. 3D capabilities, which now require the browser, will probably be enabled in the future using browser “plugins” (software add-ons).