

Why the I-900 Plan to Consolidate Multiple Reading  
Rooms  
Should Not be Implemented  
In Light of  
Best Practices for Reference Service  
At the Library of Congress

A Series of Papers

**Paper #1**  
**Background Reference Philosophy and Readers’  
Expectations**

Prepared for AFSCME 2910  
The Library of Congress Professional Guild  
Representing 1,350 professional employees  
[www.guild2910.org](http://www.guild2910.org)

Thomas Mann

April 16, 2013

No copyright is claimed for these papers.  
They are open source, and may be  
freely reproduced, reprinted, and republished.

# Paper #1

## Background Reference Philosophy and Readers' Expectations

### **Are assertions of “researchers’ expectations” of one-stop shopping, federated searching, and elimination of “silos” an adequate grounding for best practices in reference service at the Library of Congress?**

Library Services has been sponsoring a series of speakers under the new rubric of “Silos to Synergy” as an overall philosophy of service—i.e., that the walls of separate “silos” of information must be broken down in order to form merged resources for unified, consolidated, one-stop transdisciplinary service. We are told, further, that a major goal of the I-900 reconfiguration is that of “foster[ing] transdisciplinary knowledge sharing and best practices.”<sup>1</sup>

There is indeed a strong “one-stop shopping” implication to the I-900 proposal specifically. This philosophy, admittedly prevalent in much of the library profession, nevertheless needs very serious reality checks if it is now going to be applied uncritically to LC operations. These reality checks must be considered on two levels:

- 1) on the issue of eliminating silos for cross-disciplinary online searching of either websites or subscription databases (a prevalent assumption in the library profession as a whole); and
- 2) on the issue of eliminating the silos of separate reading rooms and their separate reference collections at LC in particular as a way of promoting transdisciplinary service.

It is only the second issue that is the present concern of AFSCME 2910, and it will be discussed in detail in the subsequent papers; but since the philosophy behind #2 is grounded in the prevalent philosophy of #1, we must first examine whether elimination of silos on issue #1 is in fact as desirable as it is apparently assumed to be within Library Services.

### **Eliminating “silos” in the online environment: searching open websites in contrast to searching restricted subscription databases**

Regarding the domain of freely-available Internet resources it is taken for granted that search engines (Google, Bing, Yahoo) should search “everything” simultaneously, with results relevance-ranked by various algorithmic weightings of one’s specified search-term keywords. In a sense, no one can object to this—AFSCME 2910 included—for the simple reason that there is literally no alternative method available or even conceivable for searching and sorting billions of websites.

---

<sup>1</sup> *Minding Matters*, December 20, 2012.

Swept under the rug, however, in readers' (and many librarians') expectations for "one stop" transdisciplinary searches across the entire Internet are two very different problems that inevitably show up in the actual results of such searches. The first problem is the retrieval of tens of thousands of "junk" hits that happen to have the right search-term keywords in the wrong contexts; the second is the *lack* of retrieval of thousands of other sources that are indeed conceptually relevant, but which get overlooked because they use different keywords for the same subject.

### **The difficulties of seeing "the shape of the elephant" of relevant literature via one-stop Internet searching**

It is impossible to see "the shape of the elephant" of the literature on one's topic when the only search technique employed is that of guessing which keywords might work best, no matter how extensive the content-coverage of the websites or databases may be. Researchers will always get *something*, but they never know what they are missing—just like the Six Blind Men of India in the fable who were asked to describe an elephant. None could see how many other parts of the animal lay beyond the "something" that they each grasped initially, nor could any grasp how the many diverse parts fit together.

While it is entirely admirable that Internet search engines effectively break through the walls of so many subject-specific silos in the extent of their coverage, it is simultaneously not admirable that their silo-busting reach simultaneously causes researcher to miss as many relevant sites as they retrieve—and also that relevance-ranking algorithms are not adequate to set *conceptual boundaries* to any topic, resulting in too many retrievals of irrelevant hits in irrelevant contexts. In many cases the boundaries created by subject or disciplinary silos would be much more of help than a hindrance.

### **The inadequacies of keyword searching within the open Internet**

Keyword searching achieved by silo-busting or transdisciplinary Internet searching—even when that is what "readers expect"—is not the best way to provide good service.<sup>2</sup> Let me offer an example of a Google search. This example is by no means intended as an overall criticism of Google because that service does so many things amazingly well—for both readers and reference librarians. We use it ourselves routinely for the things that it is good at providing. But no Internet search engine can be "expected" to provide "everything" in all situations. Using the Net for scholarly purposes frequently entails substantial *trade-offs* that students, in particular, are dangerously unaware of.

For example, a Google search of open Internet sites produces these results:

---

<sup>2</sup> It is not the best way if the researcher's desire is to see "the shape of the elephant" of all relevant literature; it is always very good for finding "something"—if that is all that is desired. Scholars at LC desire much more.

- “Huron Indians” (in quotation marks): 50,300 hits
- “Wyandot Indians” (in quotation marks, the designation preferred by the tribe itself): 24,200 hits
- “Huron Indians” minus (-) “Wyandot Indians”: 48,000
- “Wyandot Indians” minus (-) “Huron Indians”: 21,900
- “Huron Indians” “Wyandot Indians”: 24,600<sup>3</sup>

The fact that the overlap between sets varies under different search specifications is due to the mysteries of Google’s black box algorithm software; we cannot know from the outside what the software is doing because Google will not release that proprietary information.

The bottom line, however, is this: In transcending disciplinary silos in Internet searching you can never see “the shape of the elephant” of relevant sources. (The problem obtains in *all* Internet search engines.) You never know what you are not getting, that would have appeared if you had searched via even slightly different or unanticipated keywords.

Relevance-ranking software is of no use in solving *this* problem. It can only rank the terms that are typed in; it cannot find other relevant words that have not been typed in. (Library search mechanisms, in contrast to those of the Internet, *can* find those missing terms much more efficiently. [Examples below])

Again, however, in the context of the open Internet there is literally no alternative to relevance-ranked keyword searching. The other methods of searching more systematically cannot be scaled up to cover billions of records. Students who are accustomed to Internet searching, then, come to “expect” similar access mechanisms in libraries—in large part because they have never been shown any alternatives.

### **The inadequacies of keyword inquiries in federated searches of library-supplied subscription databases**

When we move to online resources within a research library context, however, there are indeed powerful alternatives to searching via relevance-ranked keywords; and these alternatives are much more effective in showing researchers “the shape of the elephant” of the literature relevant to their topics. “Online resources in a research library context,” here, refers to the hundreds of subscription databases that libraries can offer (ProQuest, EBSCOhost, Gale, Thomson Reuters, etc.) that are not freely available on the open Internet. An exception is our own online public access catalog (OPAC), which, being confined in its coverage to manageable onsite holdings (with no need to “scale up” to billions of websites) can indeed provide access via vocabulary-controlled search terms—which are much more efficient in showing “what the library has” within the collections that receive cataloging treatment. (Our controlled LC subject heading **Capital punishment**, for example, rounds up titles such as *Philosophy of Punishment*, *Death Penalty*, *Legal Homicide*. *Legal Executions in the Western Territories 1847-1911*,

---

<sup>3</sup> These searches were done on 3/18/13. Google results can change by the minute.

*Death Sentences in Missouri 1803-2005, Lethal Injection, Dokumentation über die Todesstrafe.* and *Contre ou pour le peine du mort*—i.e., it creates a conceptual category rounding up all the relevant works no matter what keywords their authors have used, and also no matter what languages their authors have used.)

Federated searching of all of a library's databases—i.e., searching both its OPAC and its subscription databases all at once—is not the same thing as federated searches of all free Internet websites at the same time. This distinction remains important even though it is frequently swept under the rug by many university libraries that are now promoting federated, transdisciplinary keyword searching of “all” of their databases as the best solution to their clients' research needs. (EBSCO Discovery Service, Primo Central, and Summon Article Finder are three of the commercially available services that enable federated “one stop” searching of all of a library's databases simultaneously.)

### **The serious problems with consolidated “one-stop” searches of “everything”**

In fact, this kind of “one stop”/consolidated/federated searching is often highly misleading to researchers and detrimental to quality scholarship because it leaves researchers with a false impression that they have indeed found “everything” that is available. What they do not perceive is that a federated search that *covers* a wide swathe of databases does not automatically *find* everything relevant within them because of widespread terminology variations. Holding up such a misleading prospect (“covering = finding”) is not a goal that should be advocated or modeled by the Library of Congress. Even if reference librarians elsewhere in the profession are failing to note this distinction, the reference staff at LC itself is sufficiently aware of it to the point of having rejected the EBSCO, Primo Central, and Summon services—in spite of extravagant claims by their salespeople that the “one stop” shopping covering “everything” that they provide is indeed what readers “expect.”

At the Library of Congress our reference staff wishes to provide better service—service that will actually show “what the library has” rather than merely provide “something” that is relevant regardless of how much other relevant material is being overlooked.

We believe that “best practices” entail showing researchers “the shape of the elephant” rather than just a disconnected and incomplete array of separate and unintegrated parts.

### **A specific example: Indians of North America, Native Americans, Indians of North America**

Here is a concrete example of the problem with the “one stop shopping” mindset, specifically at LC. These are the results of federated, transdisciplinary searches of all 67 disparate databases in our subscription to EBSCOhost research databases:

- A search of **Indians of North America** as a Subject Term produces

- 76,285 hits.
- A search of **Native Americans** as a Subject Term produces 26,023 hits.
- A search of **American Indians** as a Subject Term produces 13,420 hits.

The noteworthy point is that *the three different sets have only 26 hits in common*. The results are even more misleading and confusing when the same terms are searched in the default keyword mode (rather than as Subject Terms):

- Indians of North America [in “Select a Field (optional)”]: 83,031 hits
- Native Americans [in “Select a Field (optional)”]: 141,098 hits
- American Indians [in “Select a Field (optional)”]: 159,657 hits

The overlap of these three sets is only 27,527.

In other words, “one stop” transdisciplinary searching of all of our EBSCOhost databases under any *one* of these terms will cause researchers to unknowingly miss tens of thousands of relevant articles *because the different databases use very different terms for the same subject*. The variant keywords—no matter whether they are searched in controlled or uncontrolled fields—produce *vastly* different retrievals.

This fact is not at all apparent to researchers, who are seriously misled into thinking they have indeed found “everything” on their topic through any “one stop”/transdisciplinary/consolidated search, especially if that single federated search is misrepresented to them *by librarians* as covering “everything” that needs to be searched.<sup>4</sup> Simple aggregation of databases for “unified” searching does nothing to standardize their vocabulary terms. All of the computerized relevance-ranking of the keywords “Native Americans”—no matter how many databases are searched across disciplines—will not round up the disparate terms above, “Indians of North America” or “American Indians” that are simultaneously being ignored in the same databases.

### **An additional problem with one-stop searching: loss of initial focus on core literature**

The problems with “one stop” searching across multiple “silos” are not confined to the choice of search terms. An additional, equally serious problem arises in distinguishing the *core literature* within any discipline from the tangential literature that surrounds it.

Whether it is noticed in current library literature or not, readers do in fact usually prefer to have the most important sources brought to their attention first, before being led into outlying areas. This is something that we see in practice, in actually working with them, that gets overlooked by theorists (and salespeople) who assert that readers “expect” everything all at once from a consolidated transdisciplinary search—even though federated searching fails to deliver any such result even under the best of conditions, and

---

<sup>4</sup> Comparable non-overlapping results appear when doing federated searches of **Death penalty, Capital punishment, and Execution**; or **UFOs, Flying saucers, and Unidentified flying objects**.

also frequently buries the core literature of any subject area within mountains of chaff having the right keywords in only tangential contexts.

### **The *importance of silos***

It thus does not occur to many theorists in the profession who call for “transdisciplinary” service that “silos” were created in the first place in order to *solve* the serious problem of identifying, indexing, and abstracting the most important sources *within their disciplines*, so that retrieval results would not be clogged by thousand of irrelevant sources, outside the “core” silo subject area, that happen to have the right keywords in irrelevant or merely tangential contexts. This is a persistent and serious problem with Internet searching in general, but equally a problem with federated searches of library subscription databases.

What this means, again, is that the Internet model of “breaking through the walls of silos” does not apply to the many searches that are required—and that are far superior—within a library context.

*Separating the most relevant sources into disciplinary silos to begin with is in fact a highly desirable thing to do; naïve calls to “eliminate silos” and merge all sources together in one search across all disciplines produce very misleading results. The federated “solution” to covering a range of databases also entails the creation of severe problems in two other areas: 1) the even more serious problem of choice of search terms, and 2) the retrieval of too many irrelevant or tangential hits outside the core literature of any subject. And these problems remain even when readers “expect” a unified search mechanism to provide them with “everything.”*

### **An additional problem with one-stop federated searching: its elimination of search mechanisms more powerful than keyword searching**

A third and equally serious problem with federated searching of subscription databases is that it routinely *eliminates the best mechanisms for solving* the persistent problem of which search terms to use.

For example, Thomson Reuters, EBSCOhost, and ProQuest now offer many databases that allow researchers to do not just vocabulary-controlled searching but also citation searching and related record searching.

Citation searches tell the researcher which articles cite any good source that has already been identified—and very often the citing sources, while being very much “in the same ballpark,” nonetheless use entirely different and unanticipated keywords.

Related record searches enable the inquirer to find articles having *shared footnotes* with any good starting-point article—and articles having shared references, too, while being “in the same intellectual ballpark,” may also use entirely different keywords. For example:

**Starting point article:** “Economics of Antiquities Looting” [1995]

### **Citation Search Results**

[subsequent articles that cite starting point article]

- “Protecting newly discovered antiquities”
- “Cultural security: the evolving role of art in international security”
- “Heritage for sale? A case study from Israel”
- “Intellectual property crimes”
- “Occupiers’ title to cultural property: Nineteenth-century removal of Egyptian artifacts”
- “Spoil of war? A solution to the Hermitage trove debate”

### **Related Record Search Results**

[articles having shared footnotes with starting point article]

- “Evaluating the effectiveness of foreign laws on national ownership of cultural property in United States courts”
- “Reaffirming McClain: The National Stolen Property Act and the abiding trade in looted cultural objects”
- “Who owns the past in US museums? An economic analysis of cultural patrimony ownership”
- “A Tale of 2 innocents – Creating equitable balance between the rights of former owners and good faith purchasers of stolen art

In other words, subscription databases frequently offer search mechanisms other than controlled vocabularies (descriptors or subject headings) that directly address the problem that is both created and exacerbated by federated searching of keywords: *these extra search methods show readers directly-relevant articles whose keywords they could never specify in advance.*

But—and this is the crucial point—such citation search and related record search capabilities *disappear entirely* in federated/consolidated/“one stop”/transdisciplinary searches of multiple databases at the same time.

The reason is that “one stop” federated searching eliminates all search features that are not shared across *all* of the databases being searched.

In other words, federated searching radically dumbs down retrieval to only keyword searching because *that* is the only search technique that can be used across *all* different databases. Even controlled subject headings lose their “control” in federated searches because the subject descriptors from one database do not cross-search the others (e.g., Indians of North America, above). Controlled subject terms themselves become only uncontrolled keywords when they are searched outside the unique silo databases they are intended for.<sup>5</sup>

---

<sup>5</sup> Still other crucial features that are needed to cut down huge keyword retrievals vanish entirely in

## The serious problem with naïve understanding and promotion of transdisciplinarity

The bottom line is this: fostering transdisciplinarity by eliminating database “silos” through “one stop” federated searching usually results in a serious diminution of the quality of the scholarship that results. Any unquestioned assumption that such “transdisciplinary” searching is an unqualified good, or that it should be promoted at the Library of Congress at the expense of subject expertise within individual disciplines—subject “silos”—will only result in our diminishing the quality of research that gets done here.

Of course there is an important place for transdisciplinarity, *but not to start with*: it gets in the way of seeing the shape of the core literature. Psychology students, for example, are quite happy to be able to do their initial searches in *PsycINFO*—or History students in *Historical Abstracts* or *America: History & Life*—without being burdened by identical keywords that may pop up in *African Newspapers*, *American Film Institute Catalog*, *CQ Historic Documents*, *Making of Modern Law*, *Early English Books Online*, *GeoRef*, *Environmental Sciences & Pollution Management*, *Nursing and Allied Health*, *Philosopher’s Index*, *Oxford Music Online*, *North American Women’s Letters and Diaries*, and *Zoological Index*.<sup>6</sup>

In terms of a reference philosophy of “best practices,” then, the reality of high-quality reference work demands that we do *not* assume “readers’ expectations” of federation or consolidation or “one-stop shopping,” *in whatever guise*, to be a desirable goal—even if surveys indicate that readers “expect” it.

If *we* do not know any more than readers do about the serious problems created by these naïve expectations then we have no business setting ourselves up as their professional guides.

It is our job to educate rather than to reinforce ignorance.

We cannot “exceed the expectations of our users” *if we ourselves cater to those very expectations* by jumping uncritically on the federated/consolidated/one-stop-shopping/transdisciplinary bandwagon that is so prevalent in the library profession at large. (See **Paper #6** regarding the FRD’s bibliography on this.)

While such “one stop” transdisciplinary panaceas do indeed sound wonderful in theory, the realities met with *in practice* are substantially different. And librarians of the Library of Congress must deal with the realities that are actually experienced when “one stop” shopping produces such miserable results. (The surveys that say readers “expect”

---

federated searches, e.g., limiting features such as document types (obituaries, literature review articles, editorials) or dates of subject coverage (as in *Historical Abstracts* and *America: History & Life*).

<sup>6</sup> One of LC’s Kluge Scholars is now writing on “how the self-image of academics is affected when they become parents.” Federated searching produces overwhelming lists of irrelevant hits; much better results come from the use of focused “silos” such as *PsycINFO* and *Sociological Abstracts*.

“one stop” searches of everything don’t have to deal with those readers when their searches so often produce unsatisfactory results; the reference librarians here do have to show them, *routinely*, how to get beyond what they initially “expect.”)

### **Best practices at LC**

It is not at all adequate, then, to specify “best practices” as a need to “transition from passive referral to active engagement in regard to our reference transactions,” “foster transdisciplinary knowledge sharing,” and “foster new ways of problem-solving, innovation, and creativity,” all with a view to providing the “types of services our readers will expect”<sup>7</sup>—especially if these practices and expectations are uncritically assumed to entail an overall transition “from silos to synergy.” We need both, not one at the expense of the other--which is unfortunately what is entailed in the philosophy of the I-900 proposal.

We need to do *better* than readers’ expectations rather than dumb down service to cater to those expectations.

When readers who search only keywords in any of the Internet search engines *or* federated subscription databases<sup>8</sup> are routinely inundated with tens of thousands of hits—most being quite irrelevant to their interests *in spite of* “*relevance ranking*” of the results—they also then *expect* little more than *Wikipedia* to be available to provide some overall orientation to their topic. This is the source they most often turn to in the hope that it will give them some initial perspective on “the shape of the elephant” of their subjects. They “expect” little or nothing better because they have never been *introduced* to anything better.

At this very point, however, when they are most confused about how to see “the shape of the elephant” of either the basic facts about their topic or the most relevant literature on it, *the peculiar strengths of LC’s reference collections—not our websites or online databases—can provide the very solutions they most need.* And our reference collections can do it efficiently *in ways that none of the readers expect.* (See **Paper #2.**)

---

<sup>7</sup> Roberta Shaffer, *Minding Matters*, *ibid.*

<sup>8</sup> Researchers at LC can do limited federated searching of multiple subscription databases at present—but only within limited groups of databases offered by the same vendor, not across all databases provided by all vendors.