

# How to improve collection access using simple SEO strategies

SCOLMA Conference June 25, 2012

Dan Reboussin

## **Introduction: Blank screen**

Hello, I'm Dan Reboussin, African Studies Librarian at the University of Florida. Thank you all for making this meeting such an interesting and important event. I'm looking forward to continuing our conversations today and tomorrow. Please feel free to follow up with any questions you may have *during the question period* or contact me at the email address I'll provide on the final slide....

## **Slide 1: Title slide**

Today I'm presenting simple techniques you can use to dramatically improve access to your online collections. First I'll confess: None of this is original: what little I've created on my own was done without any strategic vision whatsoever. But I learned some important things on the way (thanks largely to my colleague, Dr. Laurie Taylor, Digital Humanities Librarian at the University of Florida), so I'd like to recount my experiences and hope these help us all to build better access to African-related primary collections. \*

My Search Engine Optimization voyage began in September, when Florida International University political science Professor John Clark suggested that I write a *Wikipedia* biography of Jean-Marie Derscheid after I told him of the digitized collection I had recently put online. “That is the first stop for many of us, these days,” he told me and “it’s a way to give back.”

While the Derscheid Collection is valuable, it’s somewhat esoteric as it concentrates on Rwandan pre-colonial history, colonial policies and contains primarily French language materials. The collector fascinated me as I came to know the details of his biography. He was a Belgian, from a wealthy family, who was imprisoned during WWI as was his adoptive cousin, Paul Panda Farnana (considered to be the first degreed Congolese and an early nationalist and Pan Africanist). During the 1920s, Derscheid was a very successful bird breeder, an active conservationist in Europe and a biologist with a top appointment at the Belgian natural history museum.

While I believed I had an interesting story, I was also concerned that the potential audience was rather small. \*

So, following Professor Clark's advice, I put together a biographical entry and submitted it to *Wikipedia* in October. It was really an epiphany for me to learn that *just one such online contribution* had a huge impact on this collection's position in Google search results (though I didn't think to capture results before and after my contribution).

## **Slide 2: SEO definition**

[Pause...allow audience to read definition].

My own experience illustrates a simple form of *Search Engine Optimization*: contribute relevant, rich information in the right online location to improve the placement of your collection in search engine results. This approach is satisfying in that it is immediate, transparent and so publicly visible—I myself benefit from it every time I Google this collection instead of remembering and typing out the entire URL.

### **Slide 3: Learning objectives**

My effort today will be successful if, after this talk, each of you can:

- Define what is Search Engine Optimization (SEO); and
- Name one or more actions that will make your collections more visible in search engine results.

### **Slide 4 [computer in study]**

Why is it important to make scholarly collections available to general web searches?

As Professor Clark indicated, most library research begins with general online searching. Most researchers engage in scholarly work outside of libraries (with great confidence in their skills and results). They are often removed from traditional library social spaces, so librarians have few opportunities to engage in any library research training.

## Slide 5: millennials

Aren't "Millennial" students experts at online searching? While the expectation (theirs and ours) is that millennial students are experts at searching online, recent behavioral research (Duke and Asher 2012; other references are available on the handout and online) shows that many *don't* know how to apply their general online search skills to the scholarly search environment. For example:

- They frequently search without authentication (they don't access library subscription resources);
- They are unaware of search engine algorithms for determining the relevance of results and are naive about ways to evaluate results independently for scholarly reliability and appropriateness; *and...*
- They regard librarians as "book experts" or guides to the physical building. Even when working in libraries, many or most may not request help with their library research.

## Slide 6: What can we do?

There are **3 approaches** to improving access to collections (I contribute in a varying extent to each of these in my own work).

The method I'm introducing today isn't meant to replace standard practices or proven approaches. The challenge I'm addressing is how to make our primary source materials (especially digital collections) more easily and fully discoverable, to the broadest potential population of users, with the goal of improving the quality of library research *as it is being conducted by the majority of our users*.

## Slide 7

One pathway toward improving access to collections is user education.

## Slide 8

### Library Instruction & Information Literacy

- Library instruction is valuable, but scalability may limit the scope of its impact. General, 50 minute or hour long sessions usually don't address manuscript and archival resources, unless these are specifically targeted as the focus of an instruction session.
- Courses for credit allow breadth of coverage and broader information literacy improvement, but even less scalability unless it is an institutional priority to establish broad enrollment.
- Creative ways to engage university students in information literacy include embedding librarians in dorms or departments, sponsoring gaming sessions with learning goals, etc. but all of these depend on reaching *known populations* and face limitations in their scope.

## Slide 9

### [card catalog]

A second approach to better collection access is to improve library tools (or creating better metadata), such as in finding aids, item level descriptions (associated with digital manuscripts and archives) and library catalog records (via OPACs). These all can have a positive impact for unknown and off site users.

Unfortunately, catalog records *are not easily discoverable* in general web searches. To the extent that manuscripts are described in the catalog (according to best practice standards) they are treated at the collection level in library OPACs.

*Promising federated or “Web scale discovery” services* may help to bring together primary source collections with relevant monographs and journal articles, but these products are very expensive to purchase and require complex set up.

## Slide 10

### [1950s technology]

Library discovery tools have been fundamentally based on bibliographic records in the catalog for a long time. I find this 1950s era image of electronic access humorous because of the apparent effort displayed to allow a single user to view each catalog card. It's hard to see what problem it's trying to solve...so it demonstrates that standards of practice (and expectations) change over time.

Likewise, collection level finding aids have been the standard practice in manuscripts and archival collections for some time, saving the expense of item level description. However, *item level metadata has become important again* because of the value of this information for online searches. Digitized collections are most accessible and useful when we provide as much rich, textual content as feasible.

Potential users benefit from related research content that a curator may bring together (on a landing page, for example) or that may be served to users via 'web scale' 'single search' or 'federated search' services.\*

SEO approaches are not standard (in libraries and archives) and are too new to be considered proven technologies for libraries...though they are critically important for online collections. I believe they won't look as silly in 50 years as this slide does to us now.

**To review, these are the two approaches to improving resource discovery presented so far:**

**Training users; and**

**Providing better library tools.**

## Slide 11

### SEO [Beach path].

The third approach I want to discuss is the set of surprisingly powerful SEO practices. These are capable of assisting *users who we might never know exist*, but whose general web searching can be improved by our offering a few simple, inexpensive, strategic online contributions. The users don't even have to know what they need or 'where they're going.'

Collection curators can provide these benefits with little or no support from others.

## Slide 12

### [Technical improvements]

While curators can pursue SEO strategies independently, their work is ideally supported by IT experts whose job it is to enable efficient search engine access to digital collection contents.

Search engine companies' computer software "*robot crawlers*" regularly and automatically check web servers for changes to their contents and \*

indexes. Complete user access depends on supporting these automated processes (for example, by using compliant **HTML and CSS code**), but importantly **doesn't depend on the knowledge and behavior of human search engine users.** **This is a key point:** Nobody has to read your blog, your newsletter announcement, or your *Wikipedia* biography of an obscure Belgian biologist in order for this strategy to benefit general web search users directly.

The **Sobek/CM** open source software package for managing digital collections (developed for use at my library) was designed with these goals in mind (and is freely available online for institutional adoption).

Curators themselves should also work to strategically create SEO content or supervise the content creation of staff, students, and others. Providing **rich textual content** in *Wikipedia* entries, blogs, newsletters and other online sources that **link to specific collection pages** helps search engines locate your digital collections when researchers key in related terms. These technical benefits are based upon good metadata, research context and rich content that explicitly connects the collection to related materials, content, and ideas on the open web.

## Slide 13

### Derscheid Collection

The *Wikipedia* article documents Jean-Marie Derscheid's biography, including so many diverse activities that at first I thought I was researching three different people:

- Most importantly, in relation to this collection, he was a critic of colonial agriculture policies (and their negative effect on wildlife conservation). This was his purpose for collecting the historical research on Eastern Congo and Rwanda during the 1920s and 1930s that forms the bulk of this collection;
- Beyond this work, he was involved in a large number of interesting (and ultimately tragic) activities. He joined the American celebrity taxidermist Carl Akeley's 1926 expedition to promote the conservation of Mountain Gorillas. Akeley fell gravely ill on the slopes of Mt. Mikeno, so Derscheid cared for him in his last days, continuing the objectives \*

of the expedition by surveying the Mountain Gorilla habitat, and completing the first census of the gorilla population;

- He was the first director of the *Parc National Albert*, the first national park in Africa (from its opening in 1930 until 1933);
- He taught at the Belgian Colonial University in the 1930s where he worked until the outbreak of WWII;
- Part of his collection of exotic sea birds were evacuated “in a cardboard box on one of the last boats to leave Belgium before the German invasion” and joined the London Zoo collections;
- In 1939 he returned to his WWI military unit as an army medic (but was demobilized with the Belgian capitulation in 1940);
- Derscheid continued his service with the Resistance, leading secret cells in the *escape lines* that rescued downed Allied pilots. He communicated with British forces using secret codes that he devised from African languages;
- He was captured by the Nazi GFP secret police in 1941, imprisoned and convicted as a spy (but continued to write on his work with \*

exotic birds, corresponding with members of the *Avicultural Society* of Great Britain).

- His execution was delayed by Masauji Hachisuka, a Japanese ornithologist with whom he worked in Congo and who had ties to Emperor Hirohito—himself a marine biologist—but Derscheid was finally guillotined at Brandenburg-Görden Prison in 1944.

In addition to contributing this biography to *Wikipedia*, I also:

- Created a **landing page** to bring related resources together and to provide a logical target for external links to connect to the collection;
- Uploaded a complete index of the Derscheid Collection contents to the University of Florida Digital Collections site as an overview for researchers. (In most cases this overview will be a finding aid);
- Uploaded a translated biographical dictionary article to my Institutional Repository;
- Secured the permission of Professor Emeritus René Lemarchand, a notable Rwanda scholar, to upload the full text of his 1970 monograph *Rwanda and Burundi* to the Digital Collections site in order to provide a research context for others; \*

- Contributed to **newsletter articles, blog entries**, and other promotional materials and solicited colleagues to link to the collection on their blogs and websites.

Again, this was not based on any grand plan I had in mind from the beginning. I'm simply relating a set of practical lessons that I learned in working through this project with my colleague Laurie Taylor, following my uploading of the biography in *Wikipedia* and experiencing for myself the dramatic difference in results when searching Google afterwards.

**Approx. 5 min. to go**

**Slide 14**

## **Landing page**

The landing page provides a target for researchers to easily discover. It should include rich textual content and specific, relevant links to related materials (such as online finding aids to describe the contents and their research context). \*

High quality item level metadata combined with links to research context and related research **alerts users** and **attracts general online researchers**, who may not be aware that the collection and its contents are relevant to their own research needs.

## Slide 15

### ***Wikipedia* results**

SEO requires a basic understanding of how general web search engines work—fundamentally, they rank pages in results dynamically, based on connections to content in trusted and popular web sites (for example in the .edu and .org domains, as are universities and *Wikipedia* sites).

But SEO strategies are available and accessible to curators and developers of digital online resources or collections. The creation of **rich textual resources** and **relevant research context** should be the responsibility of those who know the collection and its contents best (or they should supervise this process directly). \*

Because *Wikipedia* is a trusted and highly ranked source on the Internet, search engines use the contents of this online encyclopedia as one of the key sources for determining the rank of search engine results. Creating and editing *Wikipedia* articles is a public service that has a noticeable impact on the discoverability of scholarly research collections by general online searches.

## Slide 16

### CRL Focus

This strategic and integrated approach to making the Derscheid Collection materials more readily discoverable in general online searches was recognized in April 2012 with a **Primary Resource** Award for Access by the Center for Research Libraries (CRL), a respected organization specializing in providing support and primary resource access to member libraries worldwide.

**Approx. 2 min. to go**

**Slide 17**

**Measures of success**

The proximate goal of undertaking SEO strategies is to increase the discoverability and accessibility of library and archival resources through general, public web searches.

From this perspective, success may be measured **qualitatively**, through the improved placement of a collection in search engine results.

Similarly, access at the level of files viewed on the digital collections web server can be measured **quantitatively** through standard site statistics reports.

The ultimate goal is support for research needs and scholarly impact, measured in traditional terms by bibliographic citations and other accepted scholarly forms of acknowledgement and recognition.

## Slide 18

### Results ranking

The advantage of the kinds of SEO strategies that I'm focusing on today is that they can assist general (including naïve or uninitiated), unauthenticated users, off-site (or unknown to the curator) in the course of their normal work. **The key is that they don't have to know about the collection beforehand in order to gain access.**

This slide demonstrates the results of a Google search (which, ***as you can see in the search box***, doesn't identify the collection by name), with the resulting high placement of the Derscheid Collection due to my own SEO contributions.

## Slide 19

### File access

The impact of undertaking SEO strategies is demonstrable through file access statistics as reported monthly by the site server.

## Slide 20

I hope that now everyone here can:

- Define *Search Engine Optimization* (SEO); and
- Name one or more practical actions that will make your collections more visible in search engine results.

All of the slide photos are in the *Public Domain* or are available via [Creative Commons](#) license.

My presentation slides (with notes), the script and handout (with references) are available online at <http://ufdc.ufl.edu/AA00011385/>

Thank you all for your attention. I'll be happy to take any questions or comments now, or feel free to contact me at any time at the email address you see here.