December 2010

Pelikan's Antidisambiguation-"On the Establishment of Identity"

Michael P. Pelikan

Penn State University, mpp10@psu.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Pelikan, Michael P. (2010) "Pelikan's Antidisambiguation-"On the Establishment of Identity"," *Against the Grain*: Vol. 22: Iss. 6, Article 45.

DOI: <https://doi.org/10.7771/2380-176X.5723>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Standards Column from page 84

and-guidance). As the problems have grown in complexity, the number and scope of organizations investing time and energy in this space is increasing rapidly.

This growth in interest by organizations around the world makes the issue of coordination increasingly important. A favorite joke regarding standards is particularly relevant to the current situation regarding data distribution. **Connie Morella**, former congresswoman and ambassador to the **Organization of Economic Cooperation and Development**, said during an **ANSI's World Standards Day** gala, "Standards are like toothbrushes. Everybody wants one, but nobody wants to use anybody else's." This is especially true in the area of research data, which spans such a broad swath of the research community. What is taking place on one end of the earth in a particular discipline is often at odds with another project halfway around the globe or even next-door in a different discipline. While some of the challenges are domain-specific, many of the problems span all fields.

CODATA is one organization that is stepping up to the coordination question and some of the thornier questions of citation. **CODATA** is an interdisciplinary **Scientific Committee of the International Council for Science (ICSU)** that works to improve the quality, reliability, management, and accessibility of data of importance to all fields of science and technology. Last October during their biannual conference in South Africa, a **Task Group on Data Citation** was launched. This international group, organized jointly by several **CODATA** committees and the **International Council for Scientific and Technical Information (ICSTI)**, will explore the technical, scientific, socio-cultural, institutional, legal, and sustainability questions regarding data use and citation, including references to portions or subsets of data. They are also quite aware that citing a dataset has further implications regarding the ability to reliably identify, locate, access, interpret, and verify the version, integrity, and provenance of the dataset. The goal is to help coordinate activities in this area internationally and promote common practices and standards in the scientific community. The group hopes to organize a summit next fall to build awareness and to promote better cooperation among the various leading organizations at work on these topics.

The joint **NISO-NFAIS** project on **Supplemental Journal Article Materials** is another project that touches on this space. In scope, how-

ever, it is both larger and more tightly focused than the **CODATA** effort. It is larger from the perspective that it covers any type of supplemental material — not only research data, but also digital notebooks, textual supplementary data, software applications, audio, video, or any of the other supporting content that authors submit along with their articles for publication in scholarly publications. From the perspective of data, however, it is much more tightly focused on the publication-related questions, avoiding the more complex questions of provenance, copyright, security, data integration, packaging, and sharing. The project has begun with defining terms such as what content is supplemental, ancillary, and core to understanding. It is also looking at metadata questions and how to effectively link journal content and supplementary component elements. By working with the publishing community, the **Supplemental Journal Article Materials** project can help to codify and promote recognition of and use of these materials in the publication stream, as well as to ensure that libraries and researchers can effectively access and use them.

The **Science Commons** group, a sister organization aligned with the **Creative Commons**, is another organization with work underway. Their project, led by **John Wilbanks**, is looking at the legal structures necessary to share data among researchers. As is usually the case, copyright and legal protections regarding intellectual property are often among the most challenging issues for distribution of content. While U.S. Copyright Law doesn't protect factual items, there *are* protections for the organization and representation of data forms. Where the lines are drawn in scholarly data has not yet been determined by case law or regulation and will likely not be easily decided. In addition, different laws or regulations apply outside the U.S., where, in some cases, copyright in data can be asserted. If data is shared across international boundaries, a case can be made that the data that is returned will retain the more stringent legal strictures. **Science Commons** hopes to promote an open license solution to data sharing based on a similar structure to the **Creative Commons** licenses for publications and other creative works.

Existing work conducted by the **Open Archives Initiative on Object Reuse and Exchange (ORE)** (<http://www.openarchives.org/ore/>) could play a significant role in the packaging and distribution of datasets. The **OAI-ORE** specification presents a model for describing how elements within a compound digital object are identified, described, and related to one another. Although originally developed to deal with aggregations of Web resources, such as

Web pages or whole Websites, the specification has potential to be applied to scientific datasets. **ORE** has seen implementation in a few testing environments such as the **Chronicle of America's Historic American Newspapers** project (http://groups.google.com/group/oai-ore/browse_thread/thread/4a71d09b6b5a6feb?pli=1) and the **oreChem** project (<http://www.openarchives.org/oreChem/>). While **ORE** provide a semantic and logistical framework for packaging and distributing datasets, significant work remains before it can provide the needed tools for the scientific community.

One of the most critical success factors for the rapid adoption of the standards that are developed is making changes within the social and political environment. In the early- to mid-20th century, the publication of scholarly journal articles took off as tenure systems were developed that required the publication of research results for promotion consideration (the "publish or perish" mantra). The new government and non-government requirements for sharing of data, mentioned earlier in this article, are having a similar impact. However, these sharing mandates are only the beginning of what is needed to support a long-term infrastructure for data management. Along with legislation and policies, where the funding will come from for all of this data management is a major concern. The biggest inhibitor of adoption of data sharing is of course social, not technical or political. Some researchers are reluctant to share data and some of their organizations have created restrictions on sharing or developed incentives (like the promotion and tenure system) that could result in a mind-set of hoarding one's data. Both these organizational and individual tendencies to limit sharing will need to be overcome to succeed in large-scale data projects.

Each of these elements: legislation, organizational policy, individual behaviors, intellectual property, funding, technical infrastructure, technology, and information management standards will need to be addressed for the data sharing vision to be realized. These issues are large and interwoven and cannot be solved without significant collaboration between the affected parties and the many organizations that represent them. But the recognition of the value of research data seems to have become pervasive enough that now is the right time to facilitate this collaboration. The new government and non-government requirements for sharing of research data may just be the "tipping point" that is needed to ensure that standards are developed and adopted for the identification, citation, curation, and provenance of datasets. 🌸

Pelikan's Antidisambiguation from page 86

content would be what's popular to describe today as a great big "value add," for it would mean that **Amazon** would no longer be locked out of selling licensed content to owners of **Sony Readers** or **Barnes & Noble Nooks** — or the other way 'round, don'tcha see...

Then, the competition could be between makers of devices, based upon features, quality, snaziness, etc. There's room in the world for **Sears**, **Best Buy**, **The Sharper Image**, and **Hammacher**

Schlemmer. I mean, they all sell (or ought to sell) amazingly cunning nose hair trimmers. Why not content access devices?

And the content vendors could compete based upon the depth of their catalogs, the quality of their customer service, their ability to address the diverse interests of niche communities, and so forth.

Not so difficult from a technology point of view, really...

Well, ok, it is difficult. And you still have to empty the darned things (the nose trimmers, I mean...).

But so is every other worthwhile thing difficult

that we're all trying to accomplish in this increasingly complicated, inescapably interconnected world. And yet there must be at least fifty commercial concerns around the globe (my own wild guess, for which **Against The Grain** holds no responsibility) whose interests are focused upon perfecting the nose hair trimmer. (**Ishmael** said, "...we are all somehow dreadfully cracked about the head, and sadly need mending.") So there must be a market...

This means it's not really an argument about what can or cannot actually be done.

We just have to decide how we'd like this all to turn out. 🌸

Pelikan's Antidisambiguation — "On the Establishment of Identity"

by Michael P. Pelikan (Penn State) <mpp10@psu.edu>

In time for Black Friday, **Amazon** enabled purchasers to buy **Kindle** eBooks for other persons. "...**Kindle** books may be given or received by anyone with an email address." (note that they're no longer called "eBooks" — just simply "**Kindle** books.")

Next sentence from the FAQ: "**Kindle** books can be read either on a **Kindle** or on your **PC**, **Mac**, **iPhone**, **iPad**, **Blackberry**, or **Android** phone using one of our free reading apps." There you have it; they've covered everything.

Well, not entirely. Look, I really hate to be a mudgeon about this. I really, really like eBooks and eBook readers. I'm rapidly closing in on personally owning a half-dozen dedicated devices, not counting apps for various general purpose devices. This includes a 3rd generation **Kindle**, despite my oft-expressed qualms. Why? Because of **Amazon's** catalog, that's why. There are things I'd like to read that aren't available at any of the many sources of epub format eBooks, but which are available through **Amazon**, albeit in their proprietary format.

So I maintain multiple libraries and apps on multiple devices — but there are some titles that I can read only on the **Kindle** (or one of its apps). There are also many, many books that I've purchased from multiple vendors that I can read on the growing multiplicity of my other devices. Sure, I give up some advanced functionality, such as place-keeping as I move between devices, but that sort of feature discomforts me as much as it impresses.

So — in the user forums hosted on **Amazon** for the **Kindle** "community," there have been a number of responses to **Amazon's** "gifting" feature (my kid sister the Classics professor: "In English you can verb anything"). People are excited, have tried it, can't wait to use it, want to have delivery occur on a specified date in the future, want to be able to have their families give them books off their wish lists, want to give books to their friends living behind the barbed wire in some particular country, want to give books to their friends who use **Sony Readers**, want to be able to include **Barnes & Noble** books...

NOW HOLD IT RIGHT THERE!

The knowing rejoinders come back quickly, curtly, even "snarkily," to wit, "...why on earth would **Amazon** want to do that?" (direct quote)

So — let's take those last few in order.

According to one poster who claims to have read that it's so "...both people in the transaction must live in the same geographic location as far as availability is concerned...you can't get around the restrictions and give a person a book they couldn't buy themselves if they live in a place where that specific **Kindle** book isn't allowed yet."

Ho! That leaves kind of a cold & prickly, doesn't it? After all those warm & fuzzies?

Whose restrictions are we referring to anyway? And how could anything so harmless, green, and well, enlightened, as an eBook give offense? Oh dear... You mean that just because some repressive regime may have a handle on all the email addresses employed in their country (and maybe a few others) trying to lay hands on a forbidden eBook can get you a thump on the door in the middle of the night?

Well, as a responsible merchant, I oughtn't stray too far over any of those line, ought I?

At least with a pbook (like that phrase? Ran into it. Means "Paper Book" — y'know, what we used to call a Real Book) — anyway, at least with a pbook you stood a chance of maybe sneaking it into the country for your cousin, wrapped up with your other dirty laundry at the bottom of your steamer trunk. No more. All those switches and routers and unique addresses have characteristics that become consequential in the context of controlling what passes over your borders. Kind of why I still listen to shortwave radio.

(An aside: the most disastrously foolish thing the **BBC** ever did was to give up its long-established World Service frequencies, which upon abandonment, were instantly occupied by another country, which put extremely smooth sounding, (probably Oxford-trained) English speaking announcers into place, backed by huuuuuge transmitters, cementing (and demonstrating) a firm grasp of the means of world information shaping.)

So no smuggling that book into some country. No. Forget it. Nope.

As to why on Earth **Amazon** would want you to be able to give an eBook (ok, ok, a **Kindle** book) to the owner of a **Sony Reader** — well, why indeed? Let's see... I know! It would increase your **Kindle** book sales! No, wait, that's not the reason they won't do it...

Hmm. Oh! It would help break down the artificial barriers imposed by proprietary formats and make the world a little bit more interoperable! That's why we can't do it!

But why?

Sure, everyone making eBook readers doesn't want to lose money on the hardware. But if everybody could simply pick their hardware and then buy from any of the many quality merchants, wouldn't that be a Good Thing? Might even boost sales of hardware. I mean it's really silly. What's the difference between a dedicated reading device and a small tablet-styled device, say, a **Droid X** phone (pretty close to as useful in many things as a tablet **PC**) or an **iPad** — as long as you're the one selling 'em their content?

No — this is pure commercial spite. These companies simply want to stamp each other into oblivion or something.

So let's speak of Identity.

Right now, the content you buy is tightly coupled to an account that equates to a few specific actual pieces of hardware running specific software. There are globally unique identifiers for each, the piece of hardware or, if an app, the instance of the software. Globally Unique, see? I mean in the whole World.

The **DRM** is tied to those registered combinations. Aside from format, my **Amazon** eBook won't open on my **Sony Reader**, or vice versa, even though I'm the registered owner of both. I realize that there's nothing in this you don't already know. But consider. There's nothing inherent in **DRM** that makes this so. We could have a world in which instances of content were tied tightly to globally unique identifiers and still be able to move our content around.

How?

By having globally unique identifiers ourselves.

Let's imagine a deeply evolved Cloud environment. You have a bunch of content registered to you. Wherever you are, on whichever device you happen to pick up, whatever plane seat you happen to be sitting in, whatever taxicab you happen to climb into, well, there's your content!

The point is, your content would be understood to be related to you: not to your devices or your vendor-by-vendor individual accounts. Services would exist to manage and deliver your content — but your content would be yours. Please understand: I don't mean tearing down Intellectual Property rights here. "Buying" a book would mean that you've licensed your use of that content, not taken ownership of it. That license would simply permit greater degrees of freedom regarding expressions or renderings of that content. This increased mobility of the licensed

continued on page 85

ADVERTISERS' INDEX

25	AAAS	75	THE CHARLESTON ADVISOR	79	McFarland
19	ACM	20	THE CHARLESTON REPORT	88	MIDWEST LIBRARY SERVICE
15	ACS PUBLICATIONS	67	CHOICE MAGAZINE	55	MODERN LANGUAGE ASSOCIATION
31	AMERICAN INSTITUTE OF PHYSICS	23	COLD SPRING HARBOR LAB PRESS	13	NEW ENGLAND JOURNAL OF MEDICINE
35	ANNUAL REVIEWS	8 & 9	COUTTS INFORMATION SERVICES	17	OECD
33	ASBMB	47	EASTERN BOOK COMPANY	21	PALGRAVE MACMILLAN
37	ASME INTERNATIONAL	51	EMERY-PRATT	59	PROJECT MUSE
5	ATG	63	GIBSON LIBRARY CONNECTIONS, INC.	43	RITTENHOUSE
81	BASCH SUBSCRIPTIONS, INC.	11	IEEE	29	SPIE
41	BEGELL HOUSE PUBLISHERS	39	IEEE	2	SWETS
87	THE BOOK HOUSE INC.	7	IGI GLOBAL	71	TIMBERLINE INSTITUTE
83	BUSCA, INC.	27	IGI GLOBAL	3	YBP LIBRARY SERVICES

For Advertising Information Contact: Toni Nix, Ads Manager,
<justwrite@lowcountry.com>, Phone: 843-835-8604, Fax: 843-835-5892.