2014

Pelikan's Antidisambiguation: "I didn't sign that. Wait. Did I?"

Michael P. Pelikan

Penn State, 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. (2014) "Pelikan's Antidisambiguation: "I didn't sign that. Wait. Did I?,"" *Against the Grain*: Vol. 26: Iss. 5, Article 59. DOI: <https://doi.org/10.7771/2380-176X.6893>

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.

Pelikan's Antidisambiguation — "I didn't sign that. Wait, did I?"

Column Editor: **Michael P. Pelikan** (Penn State) <mp10@psu.edu>

One often hears that the capacity to learn is a defining human characteristic, distinguishing us from other earthly cohabitants. The Wikipedia article on Learning explains that the process of learning entails three stages that must be active: encoding, storage, and retrieval.

This assertion stands up to reason.

Restated, *Encoding* is the process, enhanced by deliberation or undercut by habit, whereby information, facts, statements representing something that can be captured in speech or other forms of communication, can be represented in a form suitable for the relevant medium of transmission, reception, and storage.

Storage entails the reception, the "taking possession of" some gathered or received unit of encoded "content," and its presumably accurate re-representation (a kind of re-encoding) as retrievable information, most likely conforming to some systematic means of characterization that assists in the organization and retrieval of the millions of such things we try to stay on top of.

Retrieval, then, proves, verifies, validates the first two stages of the process. Retrieval involves a read-back of the re-encoded content. The learning process as a whole can be tested, therefore, by requesting such read-backs (representing content that has gone through the entire process) and comparing the retrieved results to the original content that was to be learned.

Such testing is important because, while we as humans are indeed "learning beings," we are also, by nature, "forgetting beings." Forgetting is as important to learning as remembering. We need to be able to unlearn anything that has made it through the learning process that is incorrect, non-useful, or counterproductive, regardless of the stage at which the errors were introduced. It's a little trickier than that, of course. We have to be able to remember that people once thought the world was flat, for example, even as we disabuse ourselves of the idea as a currently-held "fact."

It is probably well to review the fact that computers did not come into this world "remembering" anything. The early computer programmers of legend programmed ENIAC by connecting patch cords between sections of the complex machine. The configuration of the patch cords defined the data paths through the system, literally "hard wiring" the way the machine was configured to work for each problem it was set up to handle. When you pulled the patch cords the configuration disappeared. Some considerable time was to pass before anyone added non-volatile storage (or for that matter, even appreciable volatile storage, beyond those accumulators, etc., needed for calculations to function).

So the inherent state of the glass (eventually silicon) and steel machine was vastly simple compared to the human brain (what the early computer proto-nerds and MIT referred to as the "meat machine").

And yet, as the technology underwent evolution, the volatility of machine memory was seen as a technical challenge to overcome, rather than as a technical limitation to be accepted. And just to be completely explicit, this wasn't even machine "memory" per se, but rather, simply machine-based data storage, aligned largely with the second stage of the human learning process outlined above, preceded by encoding, and completed by retrieval. Over time, it came to be widely accepted that the default behavior of these machines ought to be to "remember," rather than to "forget."

When each machine was simply an entity in a room or on a desk this idea meant one thing: it became quite another with the introduction of networking. There were periods of evolution here too. Sometimes the network was envisioned as a grouping of more-or-less equal partners, that is, more-or-less fully capable machines that could exchange data with their peers on the network. Other times networks developed around specialized capabilities appearing on the network as resources to which otherwise more-or-less capable machines might connect when those specialized capabilities were required. This specialization led to dedicated printer servers, file servers, mail servers, etc. I'm selectively leaving a lot of the history of networking aside here.

The means of controlling machine behavior evolved as well. ENIAC's behavior was defined almost entirely by the state of the many patch cords that were employed to prepare the machine for a particular task. Individual desktop machines each had configuration settings at the system level, and in turn, each application exposed particular behaviors, some configurable, some expressed in code (and thereby not accessible to the common user).

If a number of such machines are joined in a network, the behavior which is in-common (that is, shared among the network users) represents the sum of the individual configurations, permissions, etc. If a system on the network is designed to serve out files, and if that machine is set up not to require any form of authentication or authorization to access those files, then barring other factors, those files will be accessible to anyone on the network.

The administration of permissions, authorization, and identities in these networked environments has proven to be a demanding, even preoccupying, business: a profoundly complicated conjunction of issues and technical challenges that meet up with regulatory, statutory, and policy influences to create a perfect storm.

Leaving aside the jurisdictional factors that purport to dictate the rules governing the behavior of systems on a network that crosses organizational, state, and national frontiers, we run straight into a stubborn fact of policy and technology: it is the technical configuration of a system that governs that system's behavior, regardless of how that system exists in or spans jurisdictional lines. If that behavior happens to align with governing and applicable policy, then all is well, I guess. But the fact remains that statute, regulation, or published policy do not govern system behavior, any more than posted speed limits govern the speed of your vehicle (or more accurately, of the vehicle driven by the idiot behind you during rush hour).

This brings us to Privacy, Google, and the EU (for it's in the news of late). We have spent decades building systems designed not to "forget." At the system level, that translates to default policies (expressed in configuration and system behaviors) to cache, to store in temp files, to mirror storage, to enhance information recovery in the event of system or network mishaps, etc. We've intentionally made it difficult for data to disappear. This has been designed-in as a technological or public Good Thing.

We need to draw a distinction, of course, between that information that might be regarded by the "reasonable person" (a legal construction) to be public information, and that which, say, a business can keep as part of its internal records. Even so, a "reasonable person" might well agree that a company has the right to keep records of what individual customers who visit their Websites look at, seem interested in, dwell upon, return to, and so forth.

This would seem as straightforward as the permissibility of a salesperson noticing a potential customer's interest in something on display and offered for sale, say, a pair of shoes. Management would want that salesperson to notice such things, not to mention if that same customer comes back several times and displays an interest in shoes, hosiery, suit coats, or shirts with French cuffs.

Things get a little more interesting when the store realizes that it can open up a sideline business by offering to sell to other stores its observations relating to customers visiting its own store. I'd guess that a customer's right to anonymity goes a little way here — no one would expect to have to hand over their name and contact information just to gain admission to the store.

And yet, millions of customers are willing to exchange identity and contact information and more, enticed by the prospect of a free cup of coffee, coupons matched to one's own spending patterns, or some small but measurable savings on purchases. This is the model that "loyalty cards" are built upon.

continued on page 107

Pelikan's Antidisambiguation from page 106

It's a reasonable guess that business owners have run the numbers to determine the sweet spot balancing between the costs of offering these savings on the one hand, and increased profits from increased return visits on the other. I'd guess businesses don't persist in the practice if they lose money on it. And truly, the customer has signed away the right to be surprised, appalled, or ashamed by these practices — what, you didn't read the Terms of Use?

If a company has secured your blanket permission, they needn't ask you any further for permission to gather, store, retrieve, rent, sell, or otherwise put-to-business-use anything about you at all they've gathered under the Terms of Service. You're free not to give that permission, just as you're free not to use Web search, online book or merchandise vendors, or the services of telecommunications companies. So what are people complaining about? We're perfectly free to don burlap sacks and live in the woods, too.

But since I'm in a guessing mood, I'd guess that most folks reckon they've come to realistic terms with what life in the twenty-first century is all about, even though maybe, just maybe, they'll make a note to "Review the blankety-blank Terms of Service this year, and for Real this time!" to their list of New Year's Resolutions. That way it'll be certain to happen, right? 🌳

The science of service.

At Eastern Book Company, we've spent more than half a century shaping our unique brand of service. The first step is fulfilling our customers' orders with unmatched speed and accuracy. Then we custom-fit our operations to our customers' needs, allowing libraries to streamline processes and maximize budgets. And finally, we cultivate next-generation technologies to help our customers build the libraries their users need.

Eastern Book Company

1-800-937-0331
www.ebc.com

Trust. Expertise. Service.

ONIX and More: EDItEUR's Standards in the Library Supply Chain

by **Tim Devenport** (Lead Consultant, Serials & Subscriptions Standards, EDItEUR, United House, North Road, London N7 9DP, UK) <tim@editeur.org> www.editeur.org

A Bit of Background

"Are you sitting comfortably? Then I'll begin!" So ran the introductory lines to a children's story time program, broadcast for many years by the **British Broadcasting Corporation (BBC)** and imprinted indelibly on the memories of a certain generation of listeners. The story of ONIX and **EDItEUR** is perhaps a little more prosaic than some of the gems broadcast by the **BBC**, but we hope still of interest to **ATG's** readers.

The international standards organization **EDItEUR**, with its unusual, French-sounding name, was born during the 1990s out of a European Commission (EC) sponsored research project into electronic data interchange (EDI) standards in the publishing and particularly books industry. Over the intervening years **EDItEUR** has grown and evolved into an organization with multinational membership and representation from all parts of the supply chain.

As we'll see later, **EDItEUR's** interests cover three main, and complementary, areas: standards and best practice, identifiers and their supporting metadata, and project/consultancy

work. Probably the most widely known of its outputs is the descriptive metadata standard ONIX for Books, now accompanied by several more formats (or should we say, ONIX for Other Things?!) designed to support a variety of metadata exchanges.

Beyond these descriptive exchanges, **EDItEUR** is also actively involved in standards supporting transactions like ordering, claiming, etc. These include continuing support for some EDIFACT messaging, more recent work on EDItX formats, and lots of involvement in the standards developed for the serials community by ICEDIS.

Stakeholders, Membership and Governance

EDItEUR is a nonprofit, based in London, UK and funded by a mixture of membership revenues and project fees. Its member organizations — now numbering nearly 110 and based in at least 23 countries around the world — are absolutely central to its mission. All of its activities are ultimately focused on listening to and responding to requirements articulated by the members. Based on those needs,

we endeavor to develop robust, efficient, and unambiguous ways of communicating information exchanges or transactions between business partners.

EDItEUR's members are also directly involved in shaping business directions and providing governance. Representatives of its charter members constitute the **EDItEUR** Board, whilst ONIX national and international groups (for books) and the ICEDIS Committee (for serials) are responsible for commissioning new standards and for signing off new or modified formats as fit for purpose.

Descriptive and Transactional Standards

The distinction between "descriptive" and "transactional" formats is mirrored in the naming of **EDItEUR** standards that have been developed since 2000. The ONIX standards are all descriptive metadata formats (the classic "information about information").

continued on page 108