ARTICLE
EXCERPTED
FROM:

ISQ INFORMATION STANDARDS QUARTERLY

SUMMER 2011 | VOL 23 | ISSUE 3 | ISSN 1041-0031

TOPIC: ORGANIZATION AND PEOPLE IDENTIFIERS

INTERNATIONAL STANDARD NAME IDENTIFIER (ISNI)

THE OPEN RESEARCHER & CONTRIBUTOR ID (ORCID)

THE NAMES PROJECT

STANDARD ADDRESS NUMBER (SAN)

INSTITUTIONAL IDENTIFIERS (I^2) AND ISNI

NISO
How the information world
CONNECTS

OP [OPINION]

A judgement formed about something;
a personal view, attitude, or appraisal

Geoffrey Bilder

GEOFFREY BILDER

Identify This! Identifiers and Trust

I am a self-confessed identifier-dweeb. And I am acutely aware that talking about identifiers is just about the best way to shut down conversation at all but the geekiest of social gatherings. Identifiers are boring. Talk about identifiers and people think about inventory control, supply chain management, rollup reports, and government bureaucracy.

This is because most people are only familiar with the more prosaic applications of identifiers. Yes, ORCIDs will make manuscript tracking and researcher evaluation systems more efficient. Yes, Institutional Identifiers will allow publishers and librarians to manage subscriptions more reliably, and yes, CrossRef DOIs enable publishers to avoid having to negotiate complex bilateral linking agreements—but these quotidian applications of identifiers often obscure their more profound and long-term importance. Identifiers are the foundation upon which we will increasingly rely in order to publish trustworthy electronic content. In short, identifiers are about "trust."

Certainly anybody having to go through the daily slog of weeding out spam, avoiding phishing attacks, or running anti-virus software is aware of the trust problem on the internet. And this is to say nothing of the relatively new problems associated with detecting astroturfing, link-farming, sock-puppets, or any of the myriad of increasingly sophisticated techniques that unscrupulous people are using to promote their content and agendas.

When Tim Berners Lee and Nigel Shadbolt recently launched an initiative to create an academic discipline called "Web Science," they summarized the issue of trust and the web as follows:

> *"How can we determine whether we can trust the material emanating from a site? The Web was originally conceived as a tool for researchers who trusted one another implicitly; strong models of security were not built in. We have been living with the consequences ever since. As a result, substantial research should be devoted to engineering layers of trust and provenance into Web interactions."*

Indeed, but what do we mean by "trust?" Phil Windley in his book, Digital Identity, defines trust as:

> *"...a firm belief in the veracity, good faith, and honesty of another party, with respect to a transaction that involves some risk."*

One of the techniques that we normally use for evaluating trustworthiness is to assess the provenance of the entity we are being asked to trust. "Do I know this person?" "Am I familiar with this institution?" In "meatspace" we have countless cues and heuristic tools that we automatically use for evaluating trustworthiness. "Do I recognize this person's face and/or voice?" "Have I seen other branches of this store before?"

Similarly, with physical media, we could use heuristics as an aid in judging the trustworthiness of the content therein. The binding of the content, the weight of the paper and the quality of the printing, the presence/absence of scholarly apparatus (footnotes, indexes, bibliographies, graphs, equations, etc.)—all gave us clues as to the reliability and authority of the content in question.

How do we judge *trustworthiness* on the Internet

— a world where content is protean, provenance is vague, and identity is cheap?

But how do we judge trustworthiness on the Internet—a world where content is protean, provenance is vague, and identity is cheap? "Do I believe this e-mail is who it says it's from?" "Is this web login page really from my bank?" "Are the blog reviews of this restaurant authentic or were they written by shills?" We don't really have good tools for answering these questions. The cues and the heuristics that we can use for such evaluations on the Internet are negligible compared to the counterparts that we use every day in our physical interactions.

In my particular industry, scholarly publication and communication, the issue of trustworthiness is paramount and Windley's definition of trust is useful because it touches on many salient aspects of the scholarly publication process. In the case of a researcher, the "transaction" we are talking about is that of consuming and acting on formally or informally published information. The "risk" associated with this transaction is, minimally, that the researcher wastes time reading or acting on information that is somehow flawed. But often the risk can be far higher; it can, for instance, damage one's reputation or do serious harm to third parties.

So how do researchers mitigate this risk? Ultimately, of course, researchers use their discipline expertise to assess the content before they act on it. However, before assessing the content (a time-consuming process), the researcher often employs a useful heuristic shortcut, that is they look at whatever "brands" are associated with said content. Recognition of the brand will often tell the researcher something about the risk they are taking in consuming and using the information. Of course, the degree to which the brand can serve as a shortcut varies greatly from brand to brand and it is this level where publishers fiercely compete to earn the researcher's recognition and, one hopes, trust. Similarly, the researcher understands that the use of brand as a shortcut is a heuristic and like all heuristics, it is fallible. The best journals occasionally publish rubbish. Unknown journals occasionally publish gems.

There are also times when researchers cannot use brand as a shortcut. They may be unfamiliar with the brand because they are not experts in the field (e.g., a novice researcher, a cross-disciplinary researcher, a journalist, or a government functionary), but it might be because the brand has not yet been established. Even the most powerful publishing brands were, at some point in their history, entirely unrecognized.

When researchers cannot use brand as a shortcut, their next step at attempting to identify reliable information is to establish the provenance of said content. To do this, the researcher gathers and confirms evidence as to the time and place the content was created, evidence relating to the parties responsible for the creation and

CONTINUED »

Content is increasingly dynamic, increasingly copyable, and increasingly modifiable.

production of said content, and evidence of the procedures that were used to ensure the content's integrity. Researchers expect the scholarly record to aid them in assessing the provenance of content. And let us remember, this scholarly record may go back years, decades, or even centuries.

It is here where identifiers have become increasingly important in establishing the provenance and trustworthiness of electronic content. People's names "collide." People change their names and sometimes just record their names differently according to mood or situation. Organizations—not just companies, but universities, government departments, and entire countries—mutate, merge, split, and sometimes disappear. Content is increasingly dynamic, increasingly copyable, and increasingly modifiable.

All of which, combined with the general internet trust issues discussed above, means that the apparently simple act of accurately citing and crediting scholarly work is becoming more fraught. A researcher has a reasonable expectation that when he or she cites something today, that another researcher in twenty years time who follows that citation will see exactly what was cited, not some new or modified version of what was cited. And similarly, researchers expect that their work will be credited to them properly and not to somebody who shares their name.

For this to work accurately and to scale, we will increasingly have to rely on unique identifiers for people, organizations, and content. So you see, even though the value of identifiers in the short term might be to make our operations more efficient, in the end identifiers will become the foundation of a new epistemic infrastructure for reliable and trustworthy computer-mediated communication. Identifiers are not boring. Identifiers are about trust. | OP | doi: 10.3789/isqv23n3.2011.05

GEOFFREY BILDER <gbilder@crossref.org> is Director of Strategic Initiatives at Crossref.

RELEVANT LINKS

astroturfing
en.wikipedia.org/wiki/Astroturfing

CrossRef
www.crossref.org/

Government body changes: Quango list shows 192 to be axed
www.bbc.co.uk/news/uk-politics-11538534

Institutional Identifiers (I^2)
www.niso.org/workrooms/i2

link-farming
en.wikipedia.org/wiki/Link_farm

List of university and college name changes in the United States
en.wikipedia.org/wiki/List_of_university_and_college_name_changes_in_the_United_States

meatspace [definition]
www.urbandictionary.com/define.php?term=meatspace

Name Changes Since 1990: Countries, Cities, and More
www.mapping.com/changes.shtml

ORCID
www.orcid.org/

Shadbolt, Nigel and Tim Berners-Lee. *Web Science: Studying the Internet to Protect Our Future.* Scientific American, September 15, 2008.
www.scientificamerican.com/article.cfm?id=web-science

sockpuppets (Internet)
en.wikipedia.org/wiki/Sockpuppet_(Internet)

Windley, Phil. *Digital Identity.* O'Reilly Media, 2005. ISBN 978-0596008789
oreilly.com/catalog/9780596008789/