

# Founding email for CUL PITF

Persistent Identifier Task Force

Dean B. Kraftt <dean.kraftt@cornell.edu>		Tue, Sep 9, 2008 at 12:24 PM	
To: Jon Corson-Rikert <jon.corsonrikert@gmail.com>, Martin Kurth <mk168@cornell.edu>, Simeon Warner <simeon@cs.cornell.edu>, adam smith <cjs17@cornell.edu>, Brian Caruso <bdc34@cornell.edu>, Bill Kehoe <wrk1@cornell.edu>, Enrico Silferra <es287@cornell.edu> Cc: Oya Yildirim Rieger <oyr1@cornell.edu>, Tiffany Howe <tlh39@cornell.edu>			
		<p>As most of you know, I'd like to put together a short-term task force to develop a policy and implementation recommendation for persistent identifiers (URI/URLs) for the Cornell Library. I'm hoping to make this a pretty quick project (we'll see) with a tentative goal of getting the recommendation together by early November. If you are on the To: line above, then I'm asking you to be a member of this task force. I've talked directly to most of you about this - if I haven't, then you were "volunteered" by your manager.</p> <p>I've included below a few of the initial thoughts that Jon and I had on this issue, and I've attached some background documents to give us a starting point. I'd also like to get a wiki space set up where we can share documents and ideas, and collaborate on a draft recommendation. Is there an existing space where this would naturally sit, or should I get a new one set up?</p> <p>I'll ask Tiffany to set up a meeting sometime in the next week. The first order of business will be to write our own charge and try to make sure that we properly scope what we're trying to accomplish.</p> <p>Thanks in advance for your willingness to join in this effort. I look forward to meeting with you soon.</p> <p>-- Dean</p> <p>----- There's a good article on persistent identifiers in the latest issue of Ariadne:  <a href="http://www.ariadne.ac.uk/issue56/tonkin/">http://www.ariadne.ac.uk/issue56/tonkin/</a></p>	2 attac hmen ts

My original email:

One thing that's come up in several of my meetings with people at the Library has been the need for uniform, persistent URI/URLs to Cornell's digital resources - and potentially to provide digital names for non-digital resources as well. eCommons uses the Handle System, with the default hdl.handle.net domain. In many cases there seem to be no standard URLs, or only URLs specific to a particular delivery system.

I see at least two pieces of technology that are going to rapidly drive us toward wanting to have uniform, persistent URLs: OAI-ORE and RDF. Vivo is already making use of RDF, and if we want it to talk in a uniform way about Library resources, then we need to have an agreed-upon standard for URI naming of the Library stuff that it's going to talk about. OAI-ORE has the potential to expose clear structuring and relationships for web resources. It will support statements of URI/URL equivalence, but again, it would be really great to have a standard.

There are some definite challenges here: do we go with opaque identifiers or do we try to give them some human-interpretable meaning? What do we do about major systems (e.g. arXiv) that already have standard URLs?

I will suggest a starting point, at least for things that don't have clear identifiers already - using URLs that are interpretable as Handle System handles, but are in a Cornell domain. That brands the item as Cornell's, increases our flexibility to support, for instance, OAI-ORE Resource Maps for the URLs, and allows us to guarantee that we can still support them even if handle.net goes away. An example might be:

<http://handle.library.cornell.edu/1813/6298>

We could also go with a mixed and/or partially transparent scheme, which could support cases where there are already existing unique ids - perhaps something like:

<http://resource.library.cornell.edu/ecommons/1813/6298>

<http://resource.library.cornell.edu/arxiv/0803.1500>

<http://resource.library.cornell.edu/euclid.acs/1176346809>

<http://resource.library.cornell.edu/bbid/3603344>

I agree that this is an important question to raise and to do it sooner rather than later, given the Web Vision project and the likelihood of new collections based on LSDI content. We are also having to figure out a way to do cleaner URLs for Vivo for human readability, to facilitate linking from other websites, and to reduce the risk of being ignored by search engines that might interpret our current URLs with embedded URLs as redirections.

\* I think Cornell branding in either of the formats you suggest would be an improvement over non-branded handles, although we would have to address redundancy so that URLs could always be resolved.

- \* I'd like to think through the distinction, if any, between the URLs any application or collection generates as users browse or search one collection (and which will be picked up by search engines) and URLs designed to be persistent and in a library-wide namespace

- \* It's worth looking carefully at how virtual hosting and URL rewriting interact with any system we adopt -- we want a solution that will be as easy to implement and document as possible.

<b>persistent_identifiers.pdf</b> 96K			
<b>Persistent Identifiers Proposal.pdf</b> 33K			

<b>Enrico Silterra &lt;es287@cornell.edu&gt;</b>	Tue, Sep 9, 2008 at 12:30 PM
To: "Dean B. Krafft" <dean.krafft@cornell.edu>	
	<p>Dean,</p> <p>I am happy that you are pursuing PIDS -- I hope that you will acknowledge the work that has already been done on this, to which Bill Kehoe and others (like, me) have contributed.</p> <p><a href="https://confluence.cornell.edu/display/PIDS/Report+of+the+CUL+Working+Group">https://confluence.cornell.edu/display/PIDS/Report+of+the+CUL+Working+Group</a></p> <p>Rick [Quoted text hidden] --</p>

Dean B. Krafft <dean.krafft@cornell.edu>	Tue, Sep 9, 2008 at 12:58 PM
To: Enrico Silterra <es287@cornell.edu>	

Enrico Silterra <es287@cornell.edu> Tue, Sep 9, 2008 at 1:32 PM

To: "Dean B. Krafft" <dean.krafft@cornell.edu>  
Cc: Jon Corson-Rikert <jon.corsonrikert@gmail.com>, Martin Kurth <mk168@cornell.edu>, Simeon Warner <simeon@cs.cornell.edu>, adam smith <ajs17@cornell.edu>, Brian Caruso <bdc34@cornell.edu>, Bill Kehoe <wrk1@cornell.edu>, Oya Yildirim Rieger <oyr1@cornell.edu>, Tiffany Howe <tih39@cornell.edu>

The internal report that Dean attached to the previous email is also available at:

<https://confluence.cornell.edu/display/PIDS/Report+of+the+CUL+Working+Group\\>

I guess we should have assigned the report a persistent identifier!

Rick Silterra

On Tue, Sep 9, 2008 at 12:24 PM, Dean B. Krafft <dean.krafft@cornell.edu> wrote:  
[Quoted text hidden]  
[Quoted text hidden]

Simeon Warner <simeon@cs.cornell.edu> Tue, Sep 9, 2008 at 1:47 PM

----