

Twenty Years of PID Discussion: Are We There Yet?

September 2016

Views About PID Systems:
Workshop

Larry Lannom

Corporation for National Research Initiatives

<http://www.cnri.reston.va.us/>

<http://www.handle.net/>

20+ Years Later – Why Are We Still Talking About This?

- Seems easy enough – assign an ID and, if there is a corresponding resolution system, keep it current

BUT

- Assign it to what?
- Resolve it to what?
- Any required metadata and if so, where?
- If it resolves how can I understand what comes back?
- What if multiple pieces of data come back?
- Who is in charge?
- Who can I trust?
- How long does it have to keep working?
- Why will it keep working?

[Slide set from 1998]

History of the Handle System

-or-

How did we get to this point?

HS History

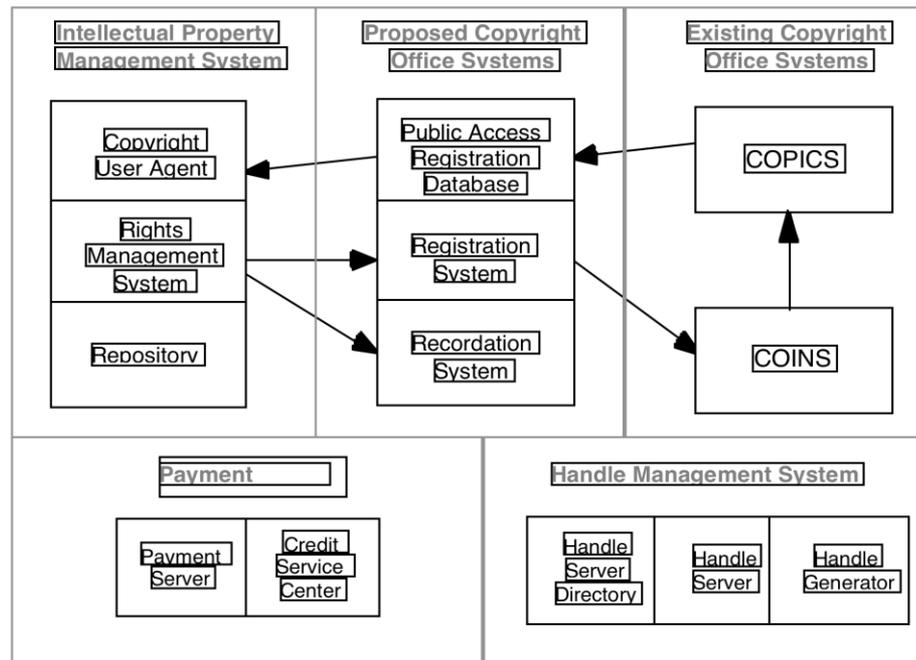
- In “Deposit, Registration and Recordation in an Electronic Copyright Management System”, Robert Kahn, August 1992:
 - When a document is deposited, a “unique numerical identifier” is assigned to the document.
 - Retrieval pointers to the Repositories are also mentioned.
 - Questions: Must the identifiers be unique - do they represent the object (or just one instance)?

HS History (2)

- An email message from Vint Cerf clarifies some of the issues:
 “...identifiers should never be reused...”
- First mention of the term “handle” in design meeting on 1/4/93. Handles may be associated with one or more Repositories and access methods.
- By March, 1993, two hs systems were identified: Handle Server(s), and the Handle Server Directory (which was used to identify the Handle Server(s)).

HS History (3)

- By June, 1993: initial Electronic Copyright Management System design was presented to the Library of Congress/ITS:



HS History (4)

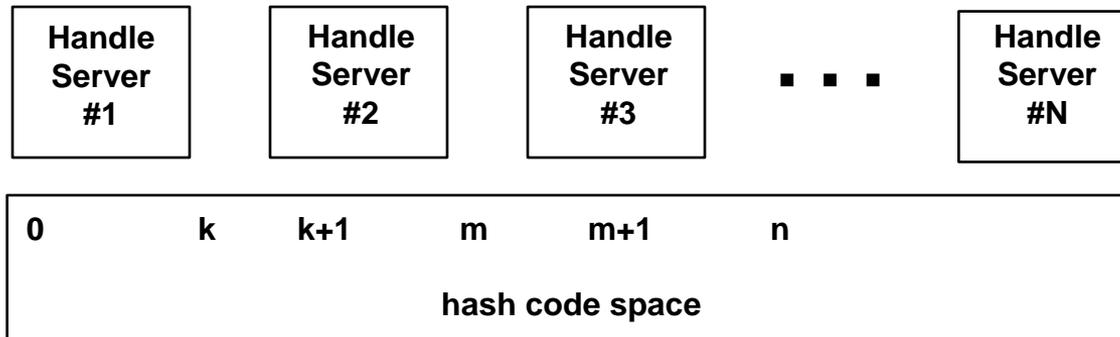
- Handles will be printable strings, using the following format:

authority name	time-sequece-stamp	descriptor
----------------	--------------------	------------

Authority names will be established, and each naming authority will be able to assign handles directly or create subzone authorities. The time/sequence stamp will be used to create a permanent unique identifier within a naming authority. The descriptor may be used to add semantic information to the handle. It does not need to be unique.

HS History (5)

- Using a hash to pick a server



HS History(6)

- The project slept for a year while the original CORDS system was developed.
- Charles started work in May, 1994 on the original HDM process, Tim Kendall had previously built a Handle Generator and the HQS. Judy Grass developed a modified MOSAIC (Unix only) that knew the HS protocol. Admin was done via email.
- August, 1994 - Presentation and Demo to CSTR group.
- Fall, 1994 - CMU Port of Mosaic w/ handle code.

History(7)

- October, 1994 - Review admin procedures with R. Kahn - start development of Group Handle and use of hierarchical naming authorities.
- November, 1994 - Selection of MD5 as hash algorithm.
- Tim completed caching server during this time as well.
- February, 1995 - Bill Arms arrives and decides to add local handle services and TCP support.
- Kendall develops first set of WWW pages (based on PERL)
- Two global servers shipped to ISI.

History (8)

- Summer of 1995: Hashing added to client library. Local HS coding changes (no secondary support yet). THQS coded.
- First set of install tools.
- Ed Overly develops first proxy.
- First Internet-Draft written.
- URN meeting @ University of Tennessee / first IETF mtg.

History (9)

- 1996:
 - Service Handle conceived and implemented.
 - Python API
 - Lots of testing - initial web pages, convert to Python
 - July, 1996 DOI proposal w/ Bowker submitted to AAP. (but where's the money!)
 - Two servers sent to ISI
 - September, 1996 - DOI proposal was accepted.
 - First AIX port sent to Library of Congress.

History (9)

- 1997:
 - Many more web page changes
 - Sam's handle system resolver
 - Secondary support.
 - DOI foundation
 - Caching Server modification.
 - Sam's Internet Draft submitted to IETF. Re-involvement with IETF.

Early DOI History

7. THE DIGITAL OBJECT IDENTIFIER (DOI)

In 1996 the American Association of Publishers (AAP) issued a Request for Proposal to develop a scheme for identification, originally called the Uniform File Identifier, and now referred to as the Digital Object Identifier . . . The AAP agreed the programme for the development of the DOI with a joint consortium of CNRI (Corporation for National Research Initiatives) and R.R. Bowker (a Reed Elsevier company). In essence the consortium brings together CNRI's highly developed Internet "Handle" technology, discussed later, and R.R. Bowker's experience in maintaining and administering an analogous identifier system as the U.S. ISBN Agency. The initial target was to develop and demonstrate a prototype of the DOI system for the AAP/PSP meeting in Washington, February 1997.

Norman Paskin. (1997). "Special Feature - Information Identifiers". *Learned Publishing* 10, 135-156. <http://doi.org/10.1087/09531519750147139>

Many Other Efforts

- URN (Uniform Resource Name)
 - RFC 1737: Functional Requirements for Uniform Resource Names (Sollins & Masinter, 1994)
 - RFC 2141: URN Syntax (Moats, 1997)
 - RFC 2168: Resolution of Uniform Resource Identifiers using the Domain Name System (Daniel & Mealling, 1997)
 - Many updates – new IETF group currently at work
- PURL (Persistent URL)
 - OCLC 1995
 - Close ties to W3C
- ARK (Archival Resource Key)
 - John Kunze, CDL
 - 2001?
- XRI (Extensible Resource Identifier)
 - OASIS
 - 2005?
- Recommend *Persistent identifiers – an overview* by Juha Hakala, 2010

DOI Discussion paper: From one to many
August 2000
Paskin & Lannom

“DOI is an actionable identifier: clicking on a DOI achieves some action. In the initial implementation, this is a single action: clicking on a DOI takes a user to one URL. We refer to this here as "one-to-one". "One-to-many" implies one DOI may be used to achieve many possible actions. This is the aim of the full DOI implementation. The purpose of this paper is to explore the mechanisms and consequences for the move from "one-to-one" to "one-to-many". It is put forward for information and to encourage discussion of the issues raised.”

Progressions

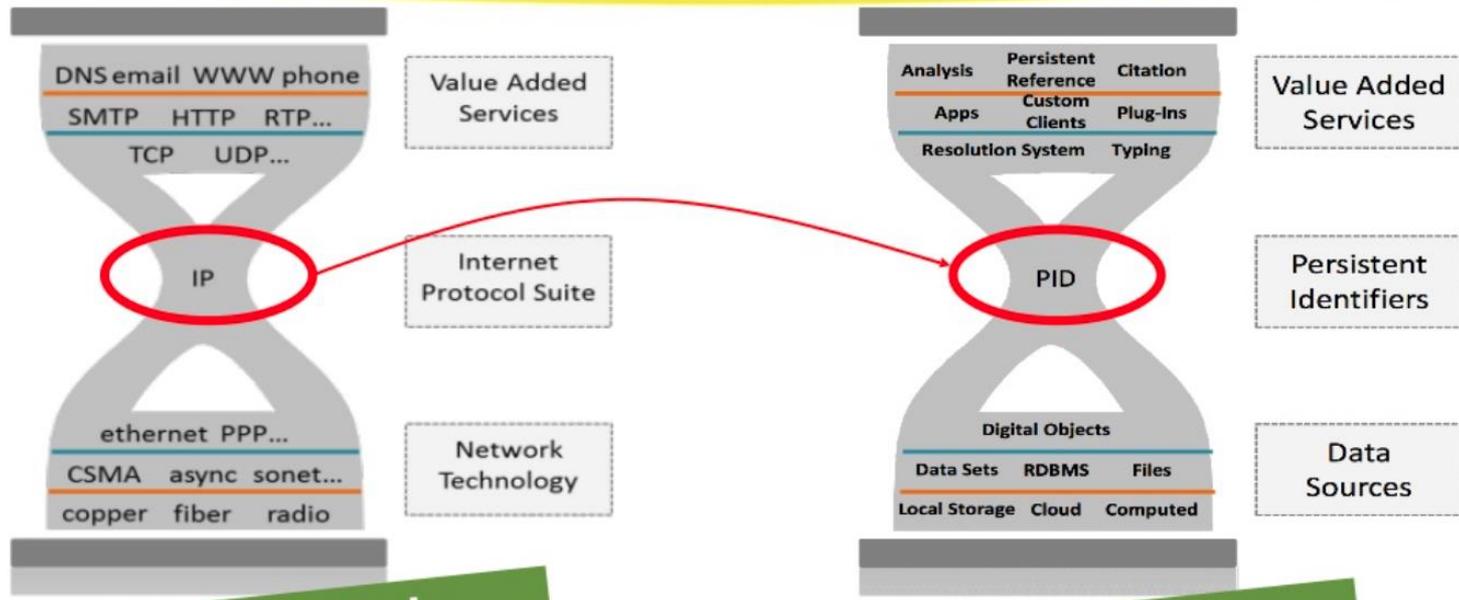
- Pre-Internet
 - Dedicated Systems – SSN, ISBN
 - Self-explanatory (affordance) – bib cite
- The Web hits
 - DNS + file name – terrific for quick start, disaster for long-term info mgmt
 - Permalink, Cool URIs, etc. – just be careful – its not enough for the important stuff over time
 - Our notion of PID
 - One level of indirection, e.g., 1 DOI = 1 URL: good start, but is that all?
 - One to Many – id resolves to current state data (original Handle goal)
 - Multiple copies, Services such as LoD
 - Typing of id values, typing of ids (DOI profiles did not take, RDA Data Type Registry, now going to ISO)
- Organizational Efforts
 - Research Orgs: CNRI, OCLC
 - Standards Bodies: IETF, W3C, RDA (small s standards), NISO, ISO
 - International Bodies: DONA, RDA again

What is Left?

- Can we get to the same level of common infrastructure for digital object ids as we have for physical addresses (IP) on the Internet?
- What would/should this include?
 - a common reliable resolution architecture (could be multiple, but all well known and obvious, possibly interchangeable)
 - an international governance body that was more dedicated to efficiency and openness than to profits
 - a platform on which multiple applications and businesses could be built – as unconstrained as possible
 - Enough accompanying information to allow the named entity to be understood and used
 - an abstraction/indirection layer for digital entities to take us to the next level – the ability to conjure a named entity into existence, given permissions, with no worries about where it is and how it is formed.

An Area of Convergence and Agreement

#



Internet Domain
 nodes with IP numbers
 packages being exchanged
 standardized protocols

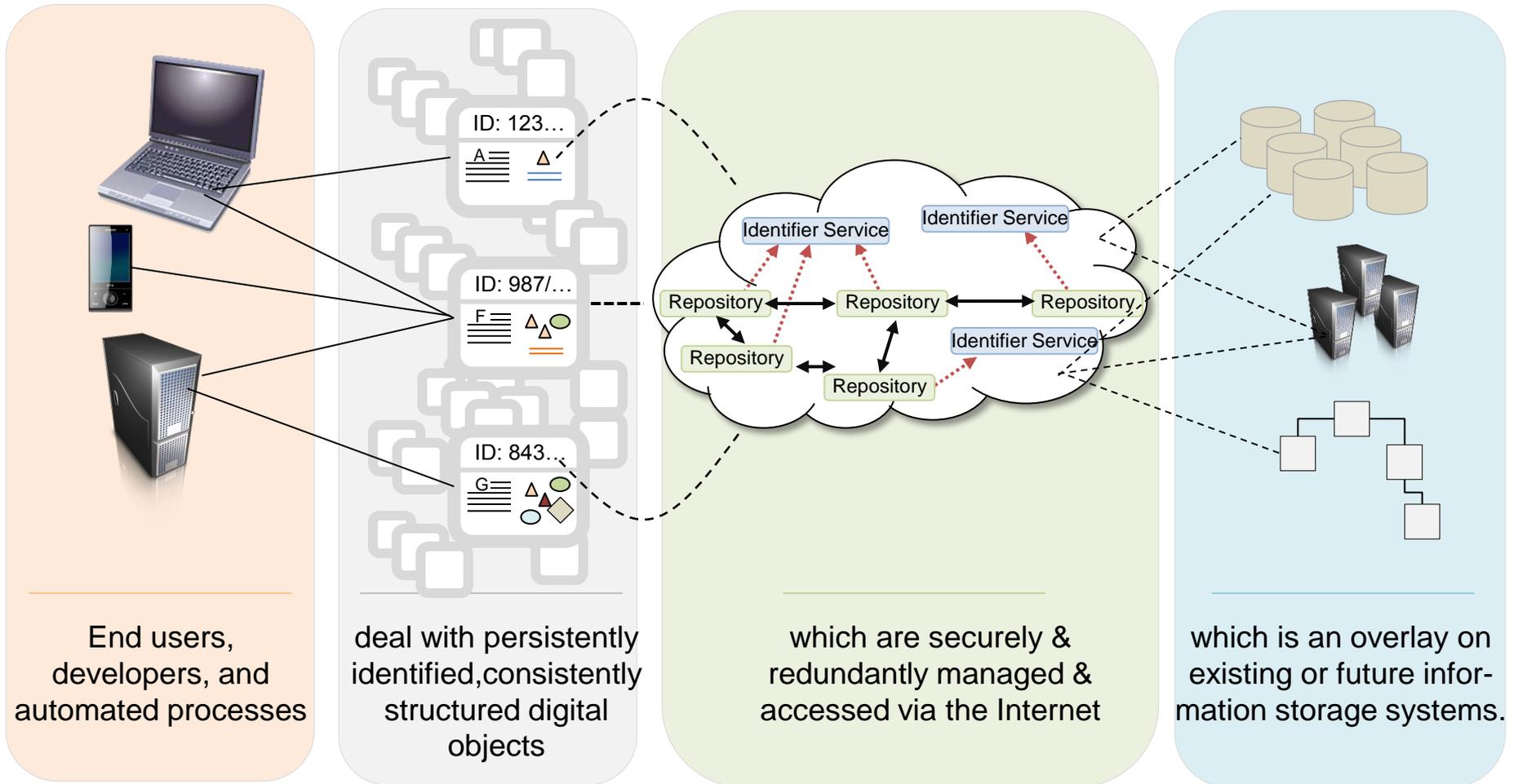
Data Domain
 objects with PID numbers
 objects being exchanged
 standardized protocols

research data sharing without barriers
 rd-alliance.org

Slide courtesy P. Wittenberg from L. Lannom



The DO Cloud





“It's tough to make predictions, especially about the future.”

— Yogi Berra