

**Erpanet Symposium on  
Persistent Identifiers**

**A framework for understanding  
Identifiers  
and “info” URIs**

Stuart Weibel

Senior Research Scientist

June 17, 2004

A globally unique ID for...

- Every concept (term)
- Every resource
- Every *objet d'art*
- Every agent
- Every cow, pair of socks, box of rocks and razor blade (RFID)



- We care about identifying resources
  - Physical
  - Virtual
  - Conceptual
  
- Knowing you have what you think you have
- Knowing we are talking about the same thing
- Reference linking
- Managing intellectual (or physical) property

# What do we want from Identifiers

- Authority
- Reliability
- Appropriate Functionality (resolution and other services)
- Persistence – throughout the life cycle of the information object
- What are the business models to support identifiers?
  - Not just a matter of money, but costs are part of the equation

# The Identifier Layer Cake

- Identifiers come in many sizes, flavours, and colours... what questions do we ask?



Social

Business

Policy

Technology

Functionality

The Web: <http>...[TCP/IP](http)...

# Functional Layer: Operational characteristics of Identifiers

- Is it globally unique? (easy)
- What is the means for matching persistence with the need?
- Can a given identifier be reassigned?
- Is it resolvable? To what?
- How does it 'behave'? What applications recognize it and act on it appropriately?
- Is the 'name' portion of the identifier opaque, or can it carry 'semantics'?
- Do humans need to read and transcribe them?
- Do identifiers need to be matched to the characteristics of the assets they identify?

# Information Assets

have life cycles with different characteristics

- Journal Articles
  - Created
  - Reviewed
  - Pre-published & Published
  - Versioned
  - Sold & Resold
  - Archived
  - Cited
  - Distributed in a variety of channels (appropriate copy problem)
- Concepts & Terms
  - Created and deprecated
  - Versions
  - Definitions
  - Abstract (concepts) and instantiated (terms)
  - Translations (for terms)
  - Position in a hierarchy (ontology)
    - Relations, linkages...

- What dependencies are assumed?
  - http... tcp/ip...(bar code|RFID) scanners...
- What is the nature of the systems (both software and social) that support assignment, maintenance, resolution of identifiers?
- Are servers centralized? federated? peer to peer?
- How is uniqueness assured?

- Who has the 'right' to assign or distribute Identifiers?
- Who has the 'right' to resolve them or offer services against them?
- What are appropriate assets for which identifiers can be assigned, and at what granularity?
- Can identifiers be recycled?
- Can ID-Asset bindings be changed?
- Is there supporting metadata, and if so, is it public, private, or indeterminate?
- Is there a governance model?

- Who pays the cost?
- How, and how much?
- Who decides?
  
- The problem with identifier business models...
  - Those who accrue the value are often not the same as those who bear the costs
  - You can't collect revenue on resolution

- At the end of the day, the only guarantee of the usefulness and persistence of identifier systems is the commitment of the organizations which assign, manage, and resolve identifiers
- Who do you trust?
  - Governments?
  - Cultural heritage institutions?
  - Commercial entities?
  - Non-profit consortia?

# The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces

- Internet Draft by Herbert Van de Sompel, Tony Hammond, Eammon Neylon, and Stuart L. Weibel
  - <http://info-uri.info>
- Separate resolution from identity
- An effort to provide a missing part of the naming architecture of the Web
- Basis for the naming architecture of Open URLs
- Possibly useful in many other areas (terminology identifiers)

# INFO URIs (continued)

- Substantial controversy about separating identity and resolution; IETF *pushback* is substantial
- Adoption and use will determine its future – will adopters find it provides sufficient value to offset cost of adoption?
- Early registrants:

Open URL	LCCN	DOI	OCLC
PubMed	OCLC	SRW Web Services	
Genbank	Fedora	SICI	
Astrophysics	Bibcodes	National Library of Australia	

# What does an “info” URI look like?

- `info:ddc/22/eng//004.678`
  - **Info:** specifies the “info” namespace
  - Namespace Token (`ddc/` in this case) is a registered namespace or brand
  - Everything that follows is at the discretion of the namespace authority that manages a given registered namespace
  - No implication of resolution, though clearly services (including resolution) can be expected to emerge.

# Identifiers for Concepts

- How do you use terminology in the Web World?
- The Semantic Web is about semantics: exchanging tokens of meaning between machines
- Identifiers are a fundamental part of this.

# Concepts can be expressed in language independent ways (even if imperfectly)

- Vietnamese War, 1961-1975  
DDC/22/eng//959.7043  
(English language version of DDC 22)
- American War, 1961-1975  
DDC/22/vie//959.7043  
(Vietnamese language version of DDC 22,)

# Boundary-Free Community Terminologies

- Controlled Vocabularies have been with us for a long time
- Hypothesis: there are specific functional requirements that terminologies should embody in order to be useful in the realization of the Semantic Web

- Global, persistent identifiers that reflect the functional characteristics of *webulated* controlled vocabularies can help us remove boundaries between and among communities and disciplines.
- Problem: Identify these functional requirements and tailor identifier systems to meet them.