A Framework for Managing the Digital Object Architecture:

11th Wurzburg Workshop on IP:
“Visions of Future Generation Networks”

Robert E. Kahn
Corporation for National Research Initiatives
Reston, Virginia
August 1, 2011
Basic Internet Notions

- Global Connectivity
- Open Architecture
- Interoperability
- Definition of Internet
- End-End vs Not End-End
- Preplanned transactions vs. Manual
- Means of addressing
Motivation for the Digital Object Architecture

- Managing information vs. just communicating bits
- Leveraging existing Internet capabilities
- Trusting one’s information to the Internet
  - Structured information, platform independent
  - Controlled access & information verification
  - Assumes both short and long term requirements
- Providing mechanisms to stimulate dynamic new forms of expression and to manifest older forms
- While supporting privacy, security, intellectual property protection, open access and well-formed business practices
Management vs Governance

- Basic question
- Still being debated in the Internet context
- Has counterparts in other areas
  - Health Information Systems
  - Financial Systems
  - Archives (National, State, & International)
  - Entertainment
- Other key issues – IP protection, security
Digital Object Architecture

User/Client

Repositories / Collections

• Metadata Registries
  in lieu of traditional
  • Search Engines
  • Metadata Databases
  • Catalogues, Guides, etc.

An open architecture for Managing Information on the Net
Digital Objects & Identifiers

- Digital Objects (DOs) are interpretable by arbitrary computers, devices and other computational capabilities
- Each DO consists of multiple “elements” each of which is essentially a “type-value” pair
- Each DO is associated with a unique persistent identifier that identifies the data structure
  - That identifier is *always* one of the DO elements
Digital Object Characteristics

- Each element of a DO can be separately “managed”, if desired, by making it a separate DO with its own identifier.
- No reliance on traditional technology such as files, folders, etc.
  - Existing technology can be used, as appropriate.
- Long-term effectiveness of handles for “citations” is achievable by maintaining the associated handle records over time.
- Manifesting the information contained in an actual digital object, on a long-term basis, is a more complex issue and one that will benefit from continued research.
Digital Object Identifiers

- Called Handles; can be branded - e.g. DOI
- Structure of an identifier is prefix/suffix
  - the prefix is unique to an individual, organization or other entity
  - the suffix can be any string in any language represented as unicode
- From an initial allotted prefix, an organization can create its own derived prefixes
  - e.g CNRI was allotted 1895, we created 1895.22, etc.
- Each identifier resolves to a “Handle Record
  - entered by the authorized handle administrator
- Organizations can run their own local handle services or contract with others to do so
  - Changes to individual handle records are under a PKI system;
The Three Rs

- Repositories – store and access DOs
- Registries – to simplify finding DO identifiers for known materials
- Resolution System – to map IDs to relevant “state information” about the DO

- DOREPOSITORY.ORG
- DO REGISTRY. ORG
- HANDLE.NET
A Repository can utilize multiple storage systems.

Digital Object Management

Digital Object Protocol

Stores DOs and provides access to them with appropriate controls.

Storage System

REPOSITORY
Digital Object Management Software

- Takes inputs based on identifiers and returns digital objects, as authorized by owners of the objects
- Interfaces with existing and older legacy systems
  - but the technology is independent of them
  - achieves interoperability with other such systems for deposit and access
- Uses PKI for access control, where appropriate
- Can support additional application dependent functionality, if desired
Resolution System

- System is non-nodal
- Scalable & Distributed, fast & efficient
- With backup & mirroring for reliability and performance
- Supports local resolution
- Currently thousands of organizations using it
- Likely over a billion digital objects now identified
Metadata Registry

• Registers the existence and access conditions for Digital Objects
  – Enables Collections to be defined with appropriate access controls
• Provides a user interface to browse and search - plus an API for external programs to search the registry
• Incorporates existing technologies
  – Handle System for identification and access
  – Digital Object Repository for metadata object storage and access
• Can use XML for:
  – object description and submission
  – Specification of Metadata Schemas
Digital Object Protocol

<input sequence><H1> <H2> <Parameters> <output sequence>

Where H1 and H2 are identifiers for 1) an action to be taken and 2) the target of that action.

A typical protocol sequence follows:

1. Connect to the Repository
2. Validate the Repository {optional}
3. Specify an action {typically a deposit or access request}
4. Validate the User {at the discretion of the repository}
5. Perform the requested action, if permitted
6. Go to Step 2 or close the connection.
Repository Interface Extensibility

• New interface functions can be added by creating new digital objects
  – The interface itself need not change
• Repository interoperability with access control is a crucial byproduct
• Direct Access to the digital object is supported – for example, to specified portions of an hour long video or to specific entries in a patient’s medical record)
• Conditions, authorizations and other access controls for usage of DOs are typically invoked by the Repository
• Protocol design & implementation is efficient
  – Multiple Users at a given site can be multiplexed over a given connection to the repository
The Handle System is a collection of handle services, each of which consists of one or more replicated sites, each of which may have one or more servers.
Handle System Usage

- Library of Congress
- IDF (International DOI Foundation)
  - CrossRef (scholarly journal consortium)
  - MEDRA (Multilingual European DOI Registration Agency)
  - Office of Publications of the European Community (OPOCE)
  - TIB → DataCite
  - EIDR
  - Wanfang Data
- Various Countries (France, Australia, China, etc.)
- EPIC – pidconsortium.eu
- Numerous Digital Library Projects
Applications & Properties

- Identity Management
- System Interoperability
- Multi-level security
- Archiving
- Protected Access to Private Information
  - e.g. medical, financial, business, etc.
- Support for “type” resolution
- Interactions via mobile programs
Enable a Role for Third Party Contributions

• Software of Various kinds
  – e.g. User Interfaces
• Tailored Applications of the Infrastructure
• Ontology mappings
• Linkages to related systems
  • Insurance, banking, archives, research
• Support for dynamic data types
  • Signal processing, extrapolation, simulation
  • Aggregation of information
Achieving Interoperability

- At the semantic level
  - Knowing what information has arrived
    - And what to do with it
- Making room for the information independent of size
  - limited mainly by available storage
- Integrating the information with other information without direct human intervention
- Affording effective forensic techniques when things go wrong
Selected Attributes

- Ability to move relatively large data sets
- Full Authentication – content, servers, users
- Ability to manage group access
- Key management
- Maintaining Multiple Instances, ease of access/addition/deletion with security
- Audit Trails
- Versioning
- Searching with Privacy Protection
Global Handle Registry

- Used to register prefixes and to resolve them
- 0.NA/<prefix> → handle record
  - Information about the prefix resolution is maintained by the prefix Administrator
- Information about the GHR itself is available
- 0.NA/0.NA → handle record associated with the Global Handle Registry
- Resolution of an identifier requires two steps
  - resolve the prefix identifier to learn which LHS has the handle record, if any – or none, if none
  - direct the handle resolution request to the LHS that holds the handle record or stop (if none).
Mirroring the GHR

Administration

M M P M M

Contains System-wide Handle Records

user

user

user

Non System-wide Handle Records are in lots of Local Handle Services
Initial Testing of a Multi-Primary GHR

A limited number of primaries each Administered Separately

Plus Mirrors

\[\begin{array}{c}
P \\
\cdot \cdot \cdot \\
P \\
P \\
P \\
P \end{array}\]

Contains System-wide Handle Records

Non System-wide Handle Records are in lots of Local Handle Services

\[\begin{array}{c}
\text{user} \\
\text{user} \\
\text{user} \\
\text{user} \end{array}\]
Comments about Administration

• The Global Handle Registry (GHR) has been operated by CNRI with a goal of cost recovery for about a decade.
• During this period, CNRI has subsidized the operation of the GHR and recovers about half the costs of running it from small annual fees
  – These are charged for enabling prefix resolution in the GHR and for maintaining the overall GHR.
• An alternate approach for running GHR services, for oversight of GHR policies and procedures, and for handling cost recovery are being considered. An organization called DONA is being established to carry out this function and to provide a framework for management of the Digital Object Architecture.
Additional Comments

- Software components are available with open source licenses; no fees to download the software.
- CNRI has certain copyright, trademark and patent rights in the technology and software; and the conditions for use of this technology and software are spelled out in the applicable license.
- Currently, CNRI administers the Global Handle Registry. Once DONA is established, CNRI will be one of several administrators to be selected by DONA. DONA will receive rights in the underlying IP.
- To maintain the integrity of the handle system, GHR administrators will have to abide by a small but reasonable set of policies and procedures.
- Local handle service providers may select their administrator of choice and will be required to follow their policies and procedures including the provision of reasonably reliable services.
Proposed Framework
Digital Object Numbering Authority (DONA)

Mission
To provide a technical policy framework for the technical management of digital object identifiers and their use in achieving information access and in achieving information system interoperability.

Technical Oversight
Technical Policy formulation

Day-to-Day Operations
Framework Principles
Digital Object Numbering Authority (DONA)

DONA will be an independent authority, formed as a non-profit Organization, most likely under Swiss law.

DONA will seek certain protections typically afforded international organizations.

Initial policies and procedures will be developed for the management of DONA in close cooperation with appropriate parties so as to enable DONA to be acceptable on a global basis.

This would include discussion of how best to handle intellectual property matters, evolution of the technology, technical policy matters and the means for management of critical resources going forward.
Framework Principles
Digital Object Numbering Authority (DONA)

2 DONA will be managed by a Board with reasonable and balanced representation by individuals with relevant technical background and expertise. Such individuals may serve in a personal capacity or may represent organizations (including private sector, government, civil society, intergovernmental or international organizations) having an interest in promoting the technical work of DONA.

3 The DONA Board will have a rotating chairmanship.
Framework Principles
Digital Object Numbering Authority (DONA)

4 DONA will be responsible for technical oversight of the multi-primary system, including establishment of prefix allotments, as appropriate. DONA would also be responsible for formulating policy related to technical management of areas within its purview.

5 DONA will coordinate with other relevant technical bodies, as appropriate.

6 DONA will be responsible for formulating the framework and requirements under which MPAs would operate, as well as identifying and validating new MPA applicants.
Framework Principles
Digital Object Numbering Authority (DONA)

A body under DONA will provide a forum for MPAs to coordinate their day-to-day technical operations (MPA Coordination Group). Any dispute not resolved within this body could be escalated to the DONA Board, which maintains overall technical oversight.

In general, decisions by the DONA Board will be taken by consensus.