

# 23<sup>rd</sup> IEEE Conference on Mass Storage Systems & Technologies

University of Maryland  
May 17, 2006



Presentation by  
Dr. Robert E. Kahn  
Corporation for National Research Initiatives

# Managing Digital Objects on the Net



- This Talk is about the Digital Object Architecture
- And about Managing Information on the Net
- And about one of its many applications with respect to the Internet – namely network storage & archiving

# The Motivation



- To reformulate the Internet architecture around the notion of managing information rather than just communicating bits
- Making use of its world-wide connectivity, but **not necessarily** its underlying transport mechanisms
- Enabling existing and new types of information to be reliably managed and accessed in the Internet environment over long periods of time
- Providing mechanisms to stimulate dynamic new forms of expression and to manifest older forms
- While supporting intellectual property protection, security, and well-formed business practices

# Digital Object Architecture



- **Technical Components**
  - Digital Objects (Dos)
  - Resolution of Unique Identifiers
  - Repositories from which DOs may be accessed
  - Metadata Registries
- **Community Applications of the Technology**
  - Provide Identifier and/or Resolution Services
  - Build a cohesive community of repository-based systems, initially around a core set of experimental projects
  - Demonstrate interoperability between heterogeneous repositories and repository systems
  - Involve government, business and academic interests

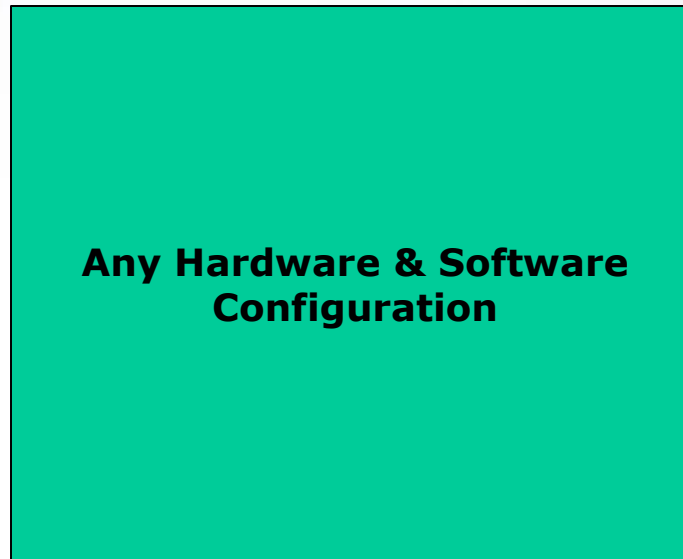
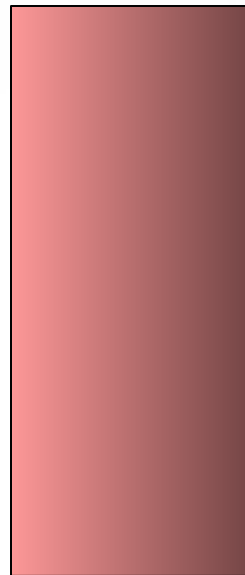
# Repository Notion



**Logical External Interface**



**RAP**



# *Nature of the Repository*



- Users can **“deposit”** & **“access”** digital objects
- Deposit produces a **“stored digital object”**
- Access results in a **“communications service”** that **disseminates** information in the form of a DO
- Repositories are themselves Digital Objects
- Digital Objects can be stationary or mobile

# Federated Repositories



- Key issue is **commonality of interests** in accessing information from multiple repositories.
- Financial Information is a prime application area
- Interoperability over time and across different underlying platforms with **security and trust**
- Metadata Registries allow for searching based on “user-supplied” inputs. The use of handles (however branded) to simplify access
- Use of local repositories, where appropriate, is an operationally desirable capability

# Open Architecture



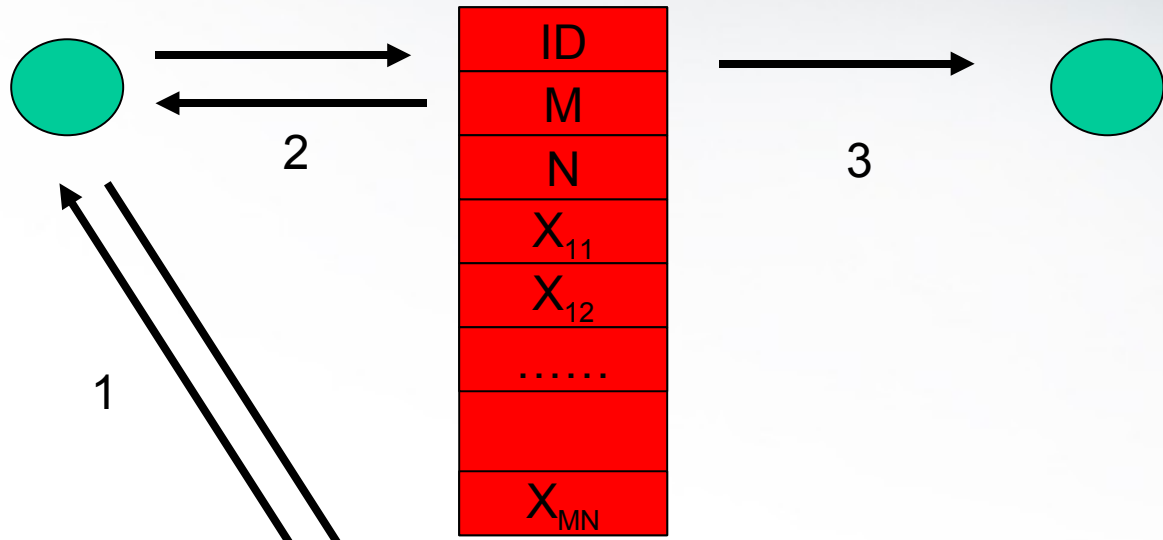
- Interoperability
  - Of Repositories
  - Of Identifier Systems
  - Of Metadata Registries
- Platform Independence
  - Over vendor choice and over time
  - System modularization
  - Protocols, Interfaces and Object Structures
- Multi-Stakeholder Solutions
  - Open to entrepreneurial development
  - Multiple participants, multiple options





Excel Spreadsheet

Another Matrix Handling Program



1

2

3

REPOSITORY

# The Handle System



- **Distributed Identifier Service on the Internet**
  - based on open interface specifications for a scalable, extendable, and efficient system (RFCs 3650, 3651, 3652)
- **First General Purpose Network Indirection system**
  - provides user-defined state information - optimized for speed & reliability on the Internet
- **Can be used to locate repositories that contain digital objects given their handles - and more!**
- **More generally, can be used to provide indirect references** - other rapid lookup information (e.g., PKI)
- The DNS was demonstrated to work on the Handle System and can co-exist with other resolution schemas within the Handle System

# Resolution Mechanism



Handle



Handle  
Record



Handle System

<[www.handle.net](http://www.handle.net)>

System is non-nodal

Scaleable & Distributed

Supports global (and local) resolution



# Handle System Features

- Full featured Identifier service
  - Supports ID resolution and administration
- Internationalized character sets
  - supports non-ASCII native characters
- Secured resolution service
  - Supports client/server authentication, service integrity, and confidentiality
- Persistent Identifier space
  - Separates identity of underlying digital objects from location

# MetaObjects & Metadata Registries



- MetaObjects provide a structural basis for indirection and for organizing information within the architecture
- MetaObjects are themselves DOs whose elements may reference other Dos
- Metadata is used to characterize digital objects, to access their identifiers and to assist in cross referencing
- Metadata may contain **terms and conditions** for use of Digital Objects
- Metadata Registries, when repository based, provide uniform access to metadata across multiple heterogeneous systems

# Managing Rights



- Terms and conditions for use may be contained within each DO
- They are intended to indicate clearly what one can and cannot do with a given DO, where such clarity is intended by the owner of the DO
- It is not an enforcement means, although it may be used by an enforcement system
- Mobile programs that are Digital Objects may apply such terms to themselves
- And on any digital objects that they contain

# Managing Transferable Records

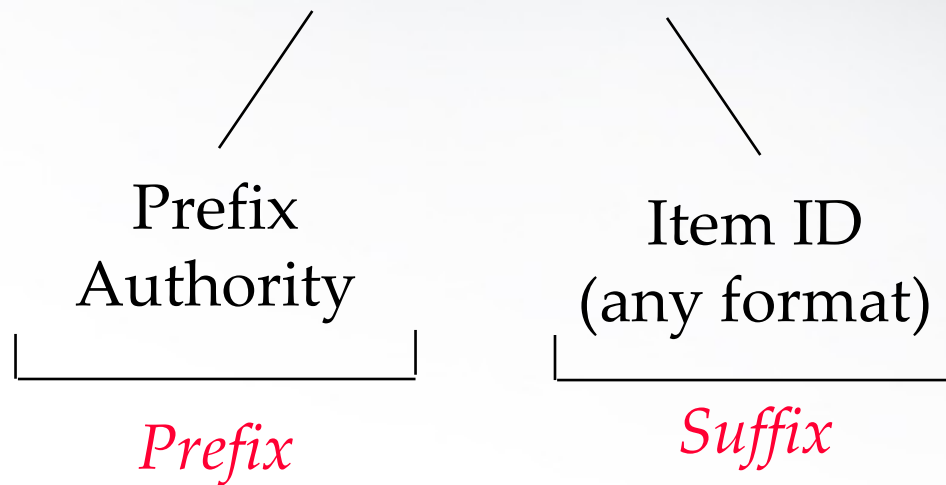


- **Relevant to many financial instruments - mortgages, deeds, bills of lading, bonds, etc.**
- **A bond is an incorporeal entity that has value; it is represented as a DO of type “Bond”**
- **No need for physical copies**
- **Bearer Instruments**
- **Full authentication**
- **Use of the Handle System supports both anonymous transfers & recorded transfers**



# Handle Format

2304.40/1234



**In use, a Handle is an opaque string.**



# Attributes of the Handle System



- The basic Architecture of the Handle System is flat, scaleable, and extensible
- Logically central, but physically decentralized
- Supports Local Handle Servers, when desired
- Handle resolutions return entire “Handle Records” or portions thereof
- Handle Records are also digital objects
- Handle Servers are certificated with the system
- Handle Records are signed by the servers

# The Digital Object Identifier (DOI®)



- **Used by the International DOI Foundation (IDF) to reference high-quality materials of publishers (and other owners of IP)**
- **Major Commercial User of the Handle System at present with approximately 20 Million handles**
- **Usage growing at about 4 Million per year**
- **DNS domain names, by comparison, are relatively flat with perhaps 30 - 40% churn per year.**

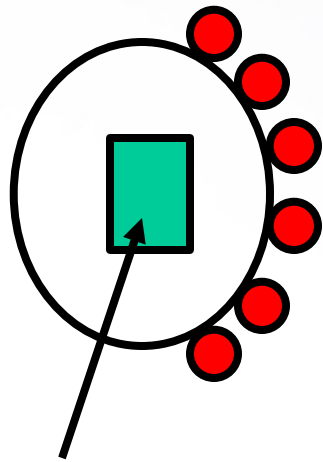
# Type Resolution



- Types are resolvable in the Handle System
- Types may be created dynamically
- Types may be locally named, mapped into bit strings without semantics
- Primary prefix zero “0” is used for system identifiers
- 0.type/<type> is the system handle for type
- Other handles may cross reference this handle (e.g. for international use)



# Object-Oriented Programming

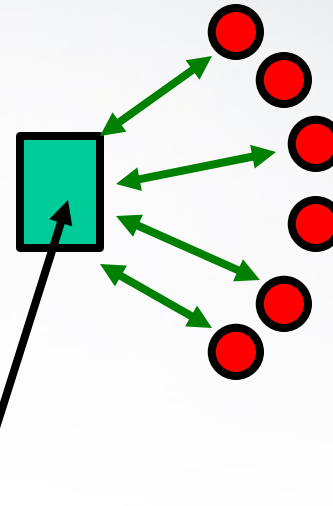


Methods

Internal Data Structure

The internal data structure is not directly accessible by the programmer

# Digital Object Model



Disseminators

Digital Object

Access to the object is subject to control by the owner. For example, a market in disseminators is possible.

# Setting up a *Local Handle Service*...



- Download the software from <http://www.handle.net>
- Follow the instructions in the installation script.
- Send your “site bundle”, containing the IP address of your server and your administrator information, to the Global Handle Registry<sup>®</sup> (GHR) administrator
- Site will shortly announce a formal software release under an open-source type of public license
- Users include DOI System, Open Archives, DTIC, Library of Congress, DSpace, Office of European Publications, German National Library of S&T

# Evolution of Policy for Global



- **Original Policy**

- Best efforts service; run in-house
- Cost paid by the Government
- Available to the research community for free

- **Current Policy (still in flux)**

- Best efforts global service; run 7x24 with backup
- Users pay for global prefix resolution via registration
  - Various Models of Usage
  - Single prefix → Delegation of Authority
- Handle System Advisory Committee oversees costs and evolution
- Provision for Interoperability with other identifier systems

# Cost of Global Services



- IPv4 → several million addresses; about 75M TLDs (excluding CCs)
- At say \$20 per year per TLD, the cost of global registration and resolution services is about \$1.5B per year – both costly and very profitable
- The handle system is comparable in size to the DNS (there are about 20M DOIs alone) and, at present, it costs about four orders of magnitude less per year to operate.
- The handle system can support the DNS, and also resolution to IPv4 and IPv6 addresses with or without the use of DNS

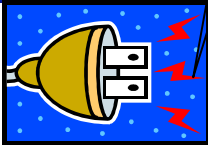
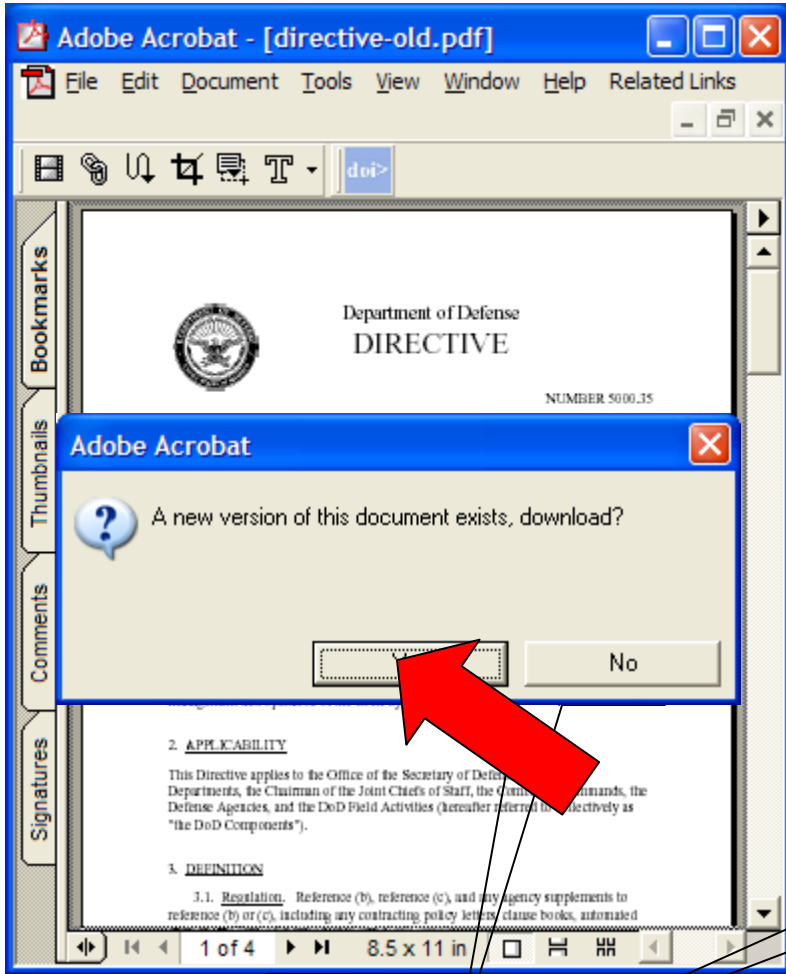
# Applications of the Technology



- Network Storage & Archiving
- Identity Management
- PKI Infrastructure
- Authentication of Information
- Personal Locator Information
- Steganography
- Digital Cash
- Managing Transactions
- Accessing up-to-date Publications
- Intrinsic Cataloguing and Indexing

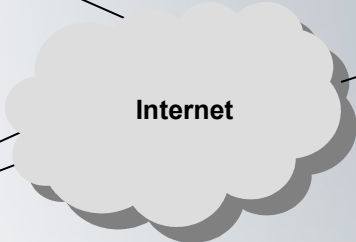


# Demo 1 – Version



Handle System	
DOI.MD	<pre>&lt;?xml version="1.0"?&gt; &lt;!DOCTYPE multires SYSTEM "multires.dtd"&gt; ... href="http://www.darpa.mil"/&gt; ... 2001 Jan 25, DA Regulatory System" "http://www.darpa.mil"/&gt; 5000.42, 200 Feb 3, Federal Acquisition Policy" "http://www.darpa.mil"/&gt; istr. 5229.5, 2002 Mar 2, Acquisition Policy Implementation" "http://www.darpa.mil"/&gt;</pre>
URL	<a href="http://kaluha/~jpetrone/directive.pdf">http://kaluha/~jpetrone/directive.pdf</a>
HS_MODIFIED	<b>D:20030107180731-05'00'</b>

Handle Record





# Network Storage & Archiving

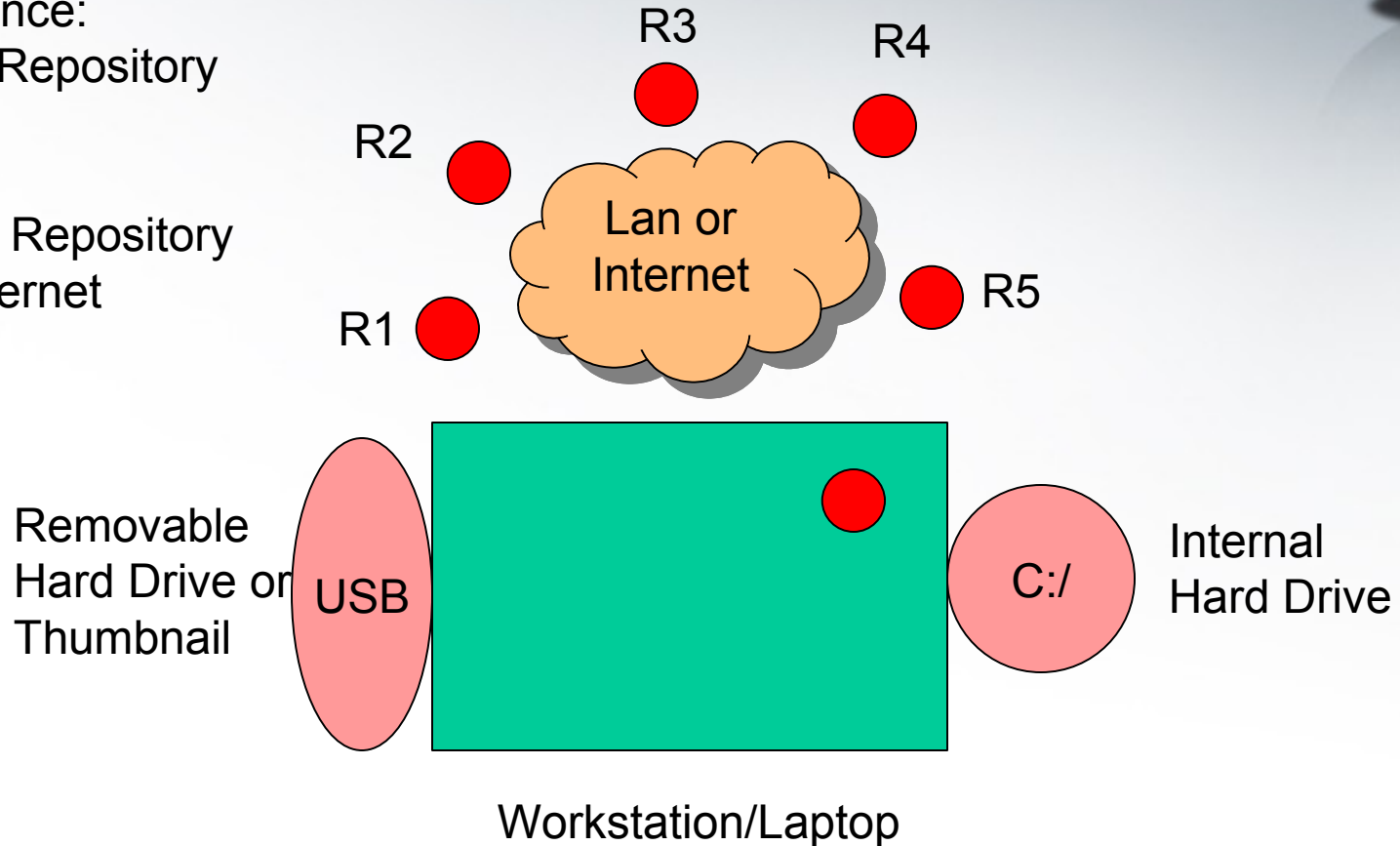
Precedence:

Internal Repository

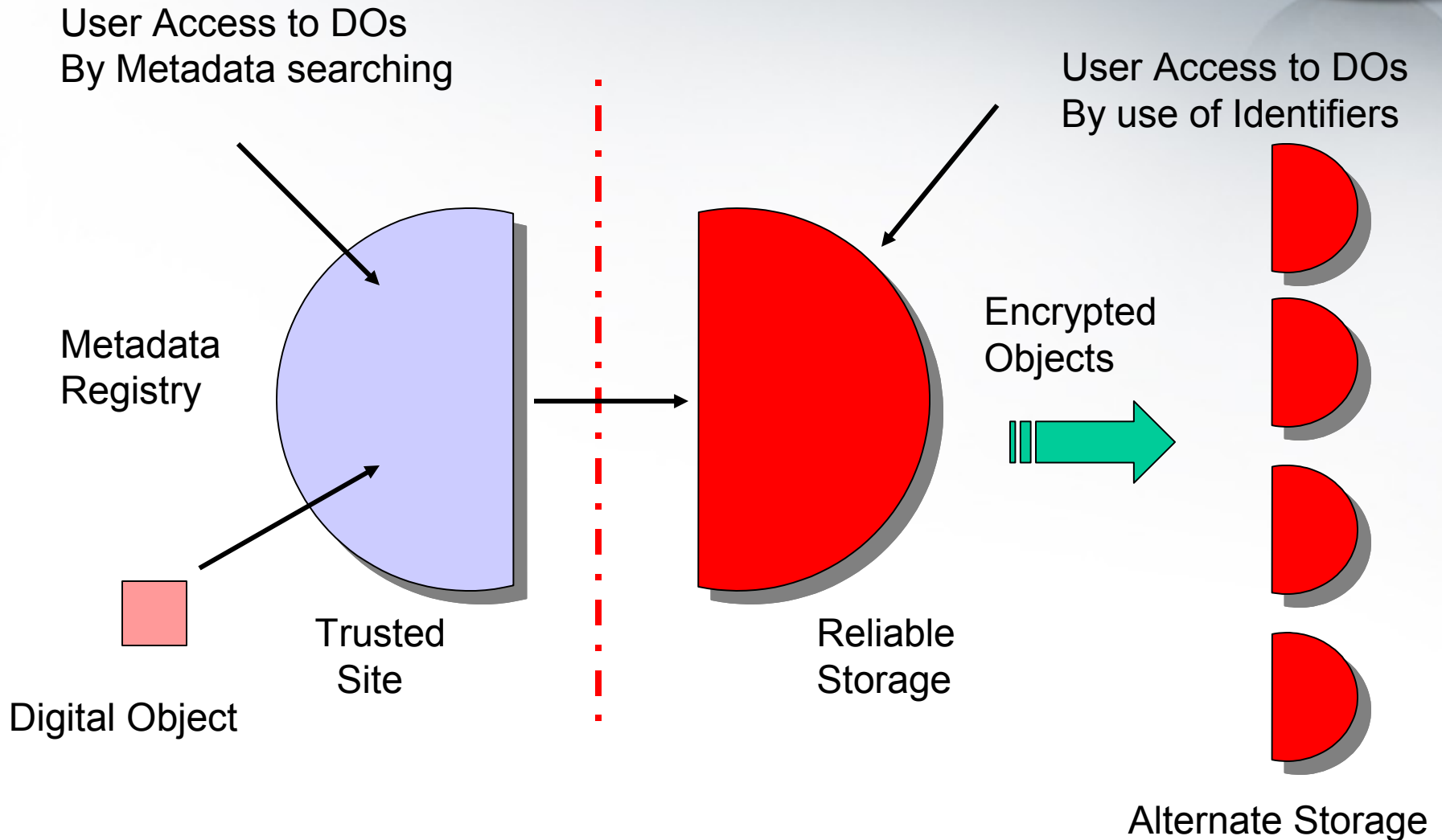
C:/  
USB

External Repository

Lan/Internet



# Metadata & Encrypted Objects



# Key Issues



- Versioning
- Firewalls
- Source tagging of Objects
- Identity Management
- Group Access Mechanisms
- Simplified Key Management
- Representation of Metadata Schema



# Conclusions

- The Digital Object Architecture is a core infrastructure component for managing information; especially over very long periods of time
- Technology Components are available from R&D
- Network Storage and Archiving - one key early application
- Other Applications (with user-friendly interfaces) need to be developed & deployed; Metadata registries need to be created and maintained; interoperability needs to be demonstrated and nurtured
- Third-party value-added providers will ultimately shape the long-term evolution as they did in the Internet
- Such an Infrastructure can fundamentally alter the net and how we use it

# And Finally, I expect



- The Internet architecture as we know it will evolve to a more flexible and dynamic plane
- The Infrastructure will expand to incorporate Digital Objects as basic information units
- Identifiers will be the linchpins of the System
- Network-based Storage & Archiving will become the Norm
- Repositories will allow for interoperability
- These concepts can and will also diffuse down to most aspects of network management