

Reflections on the Digital Object Architecture



**by
Robert E. Kahn, CNRI**

**A presentation at a Symposium on Trusted
Repositories in Rome, Italy on November 17, 2003**



The Motivation

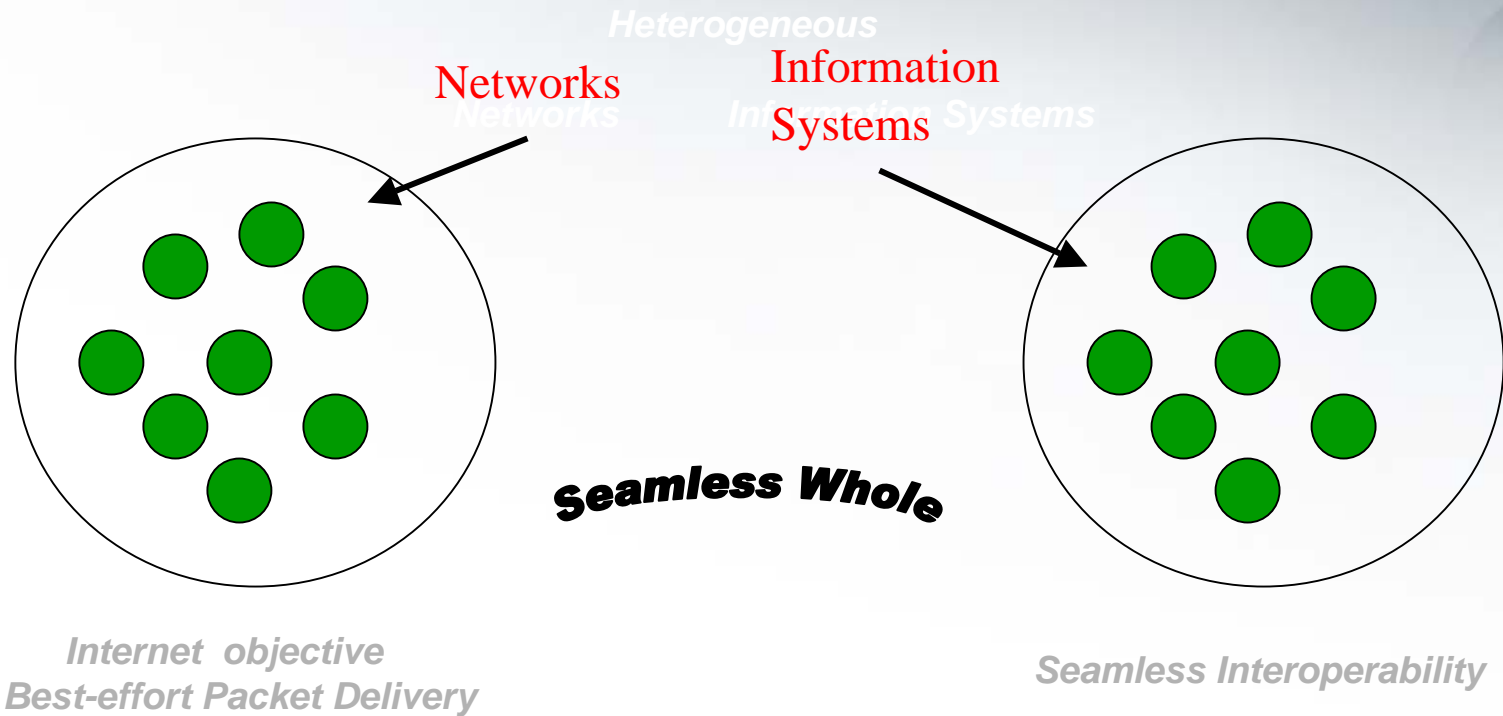
- To reformulate the Internet architecture around the notion of uniquely identifiable data structures
- Making use of its world-wide connectivity
- But **not necessarily** its underlying transport mechanisms
- Enabling existing and new types of information to be reliably managed and accessed in the Internet 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 and well-formed business practices



The Background

- Started with the original **Knowbot** work at CNRI in the 1980s on Digital Libraries
- Which was then split into two categories
 - Digital Objects including Mobile Programs
 - Repository Systems
- A split that was largely illusory since
 - Repositories can be Mobile Programs and in motion on the Internet
 - Repositories and mobile programs are themselves DOs
 - Mobile Programs need not necessarily move
- Engaged in a community development effort under the DARPA supported Computer Science Technical Reports (CSTR) project in the 1990s

Objective of the Framework



Federating Heterogeneous Systems

Digital Object Architecture

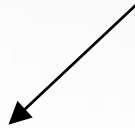


- **Technical Components**
 - **Digital Objects (Dos)**
 - **Resolution of Unique Identifiers**
 - **Repositories from which DOs may be accessed**
 - **Metadata Registries**
- **Community Applications of the Technology**
 - **Build a cohesive community of repository-based systems, initially around a core set of projects at universities, non-profit organizations, and government**
 - **Demonstrate interoperability between heterogeneous repositories and repository systems**
 - **Involve business interests such as the publishing industry**

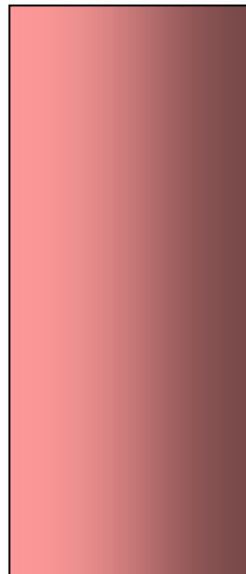
Repository Notion



Logical External Interface



RAP

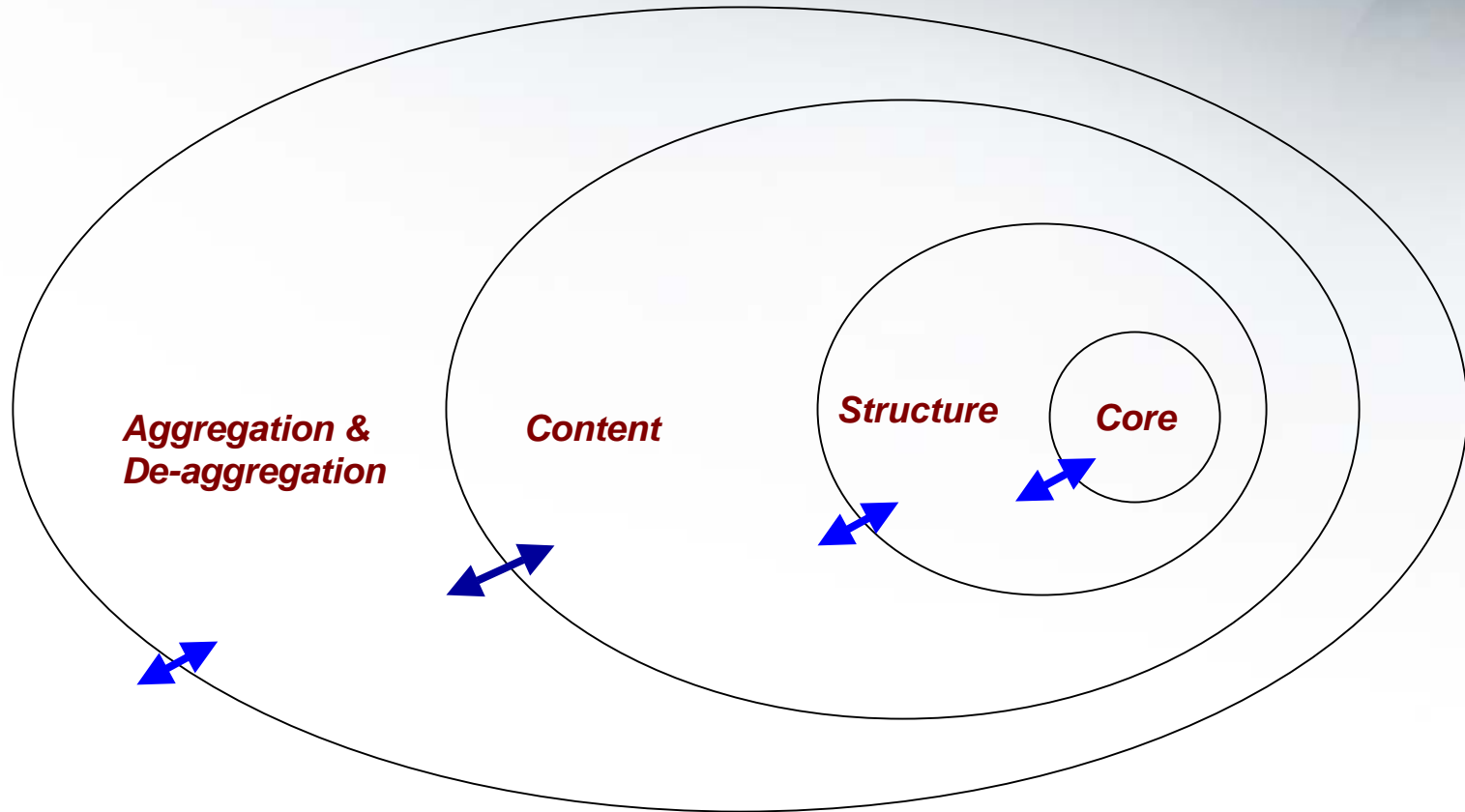


Nature of the Repository



- **Not** like a bookshelf or a pantry
- More like a service-oriented restaurant
- One 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
- Like restaurant ordering results in a culinary service which results in an eating experience

Nesting of Repository Functionality



*Core Interface must be present at each level
Other levels could be separately defined later*

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

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

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

Communicating Digital Objects



- Generation or Retrieval of Digital Objects for Dissemination
- Transporting Digital Objects
- Making requests of Digital Objects
 - Sending email to a book
 - Interactions between DOs
- Switching Digital Objects
 - Mapping Handles into IP Addresses at the source
 - Or using IP as a substrate mechanism enroute
- Managing Disseminations
- Observing Relevant Terms and Conditions

Managing Rights



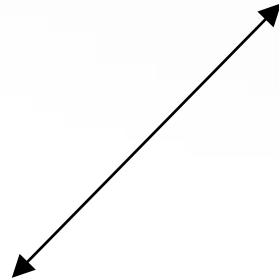
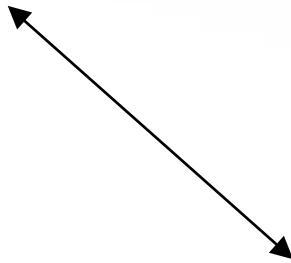
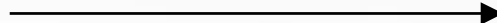
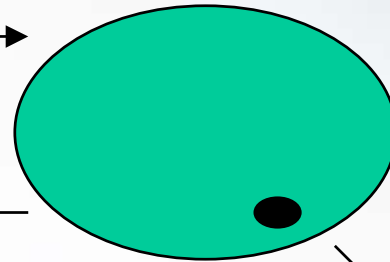
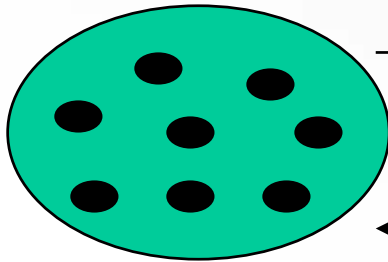
- Terms and conditions for use may be contained within each DO
- They are intended to indicate clearly what one can and cannot do with a given DO, where such clarity is intended by the owner of the DO
- It is not an enforcement means, although it may be used by an enforcement system
- Mobile programs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

Interactions between Repositories



Repository A

Repository B



**Stored
Digital Object**

- For Backup
- To Communicate
- For Distributed Tasks
- For Replication, Mirroring



User's Computer

Managing Transferable Records

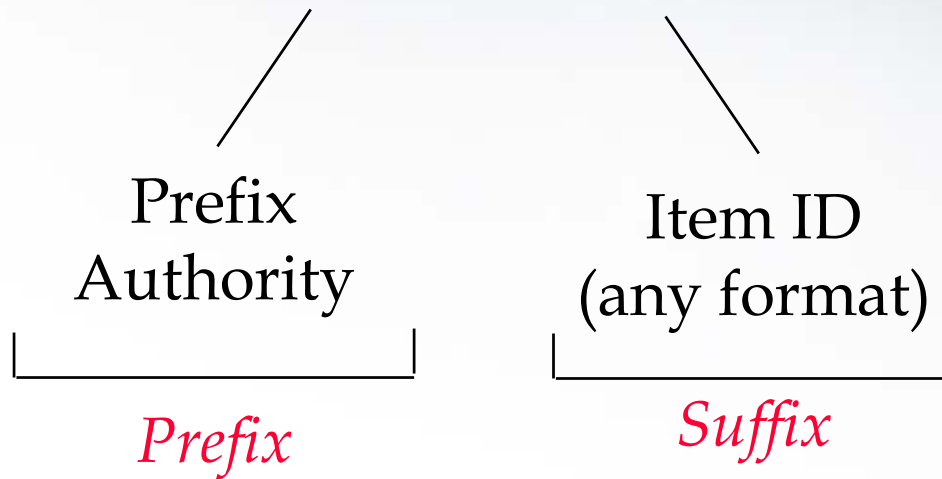


- Relevant to many financial instruments - mortgages, deeds, bills of lading, bonds, etc.
- A bond is an incorporeal entity that has value; it is represented as a DO of type “Bond”
- No need for physical copies
- Bearer Instruments
- Full authentication
- Use of the Handle System supports both anonymous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



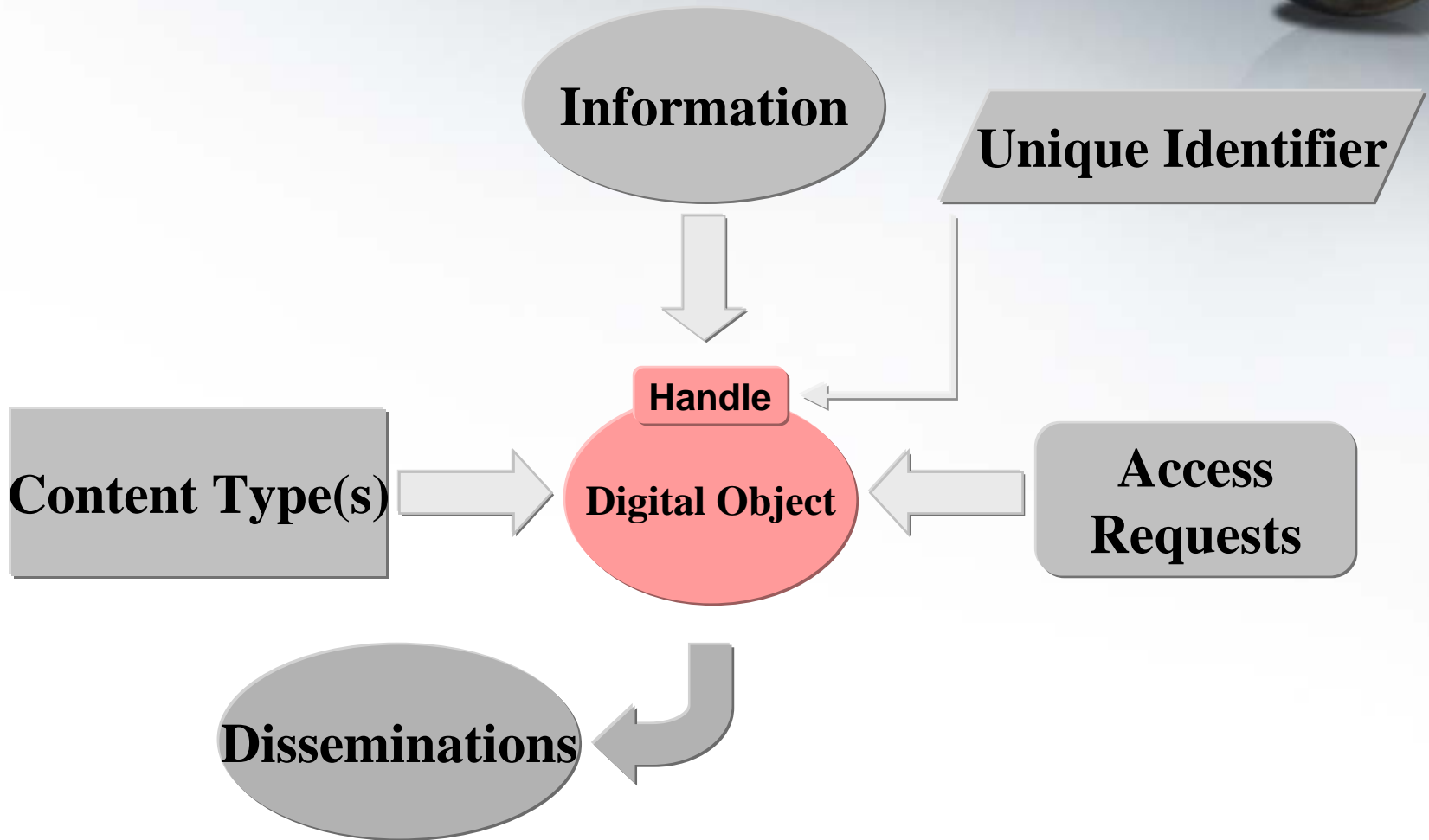
- **Used by the International DOI Foundation (IDF) to reference high-quality materials of publishers (and other owners of IP)**
- **DOIs are handles whose primary prefix is “10”**
- **Initially, DOIs resolved to a single URL, now moving to multiple resolution**
- **Policies and Procedures for use of DOIs**
- **“Qualified” Registration Agencies**
- **Central DOI Directory for backup and reliability**
- **Enhanced browsers for direct handle access**
- **Use of Proxy servers for unenhanced browsers**



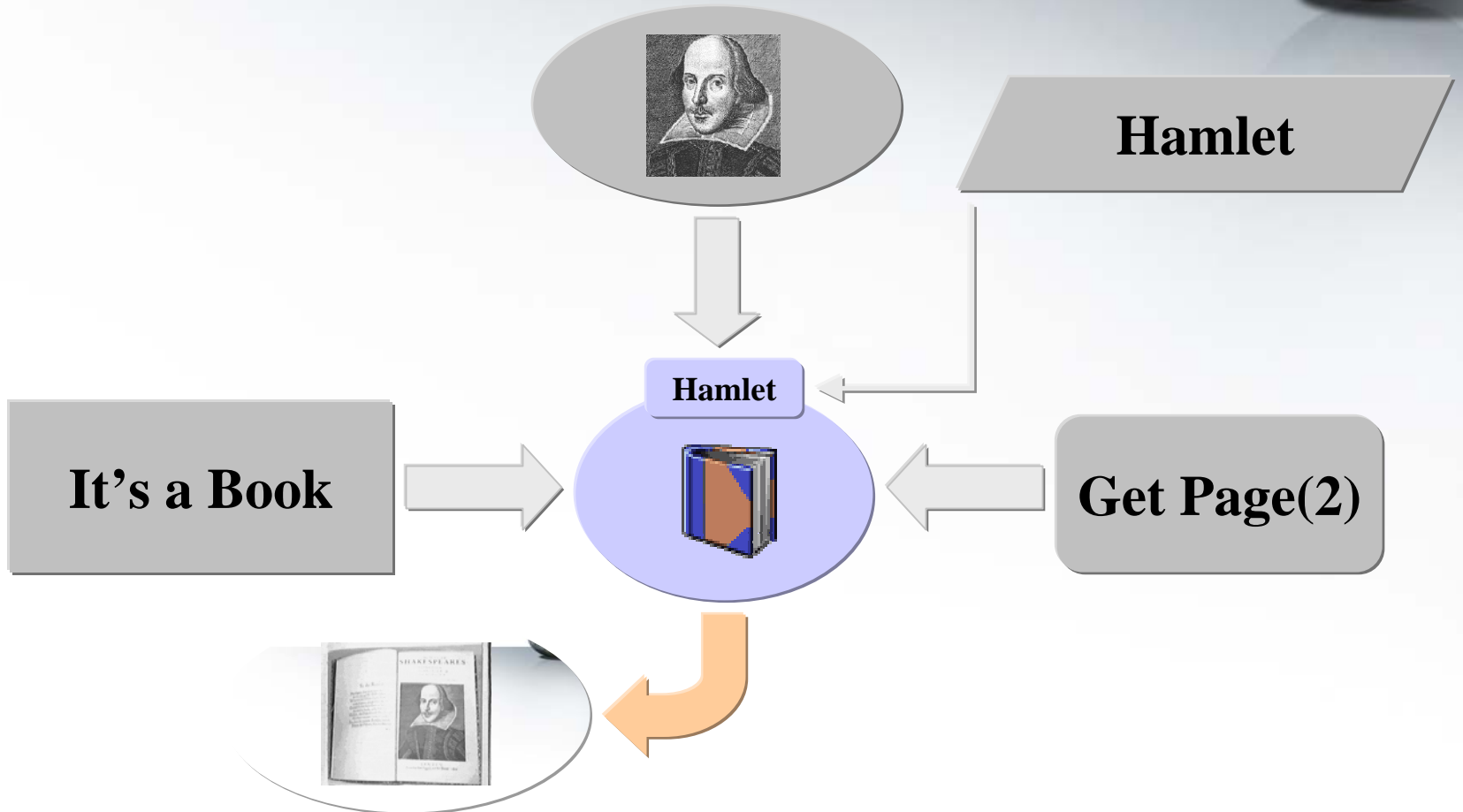
Type Resolution

- 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)

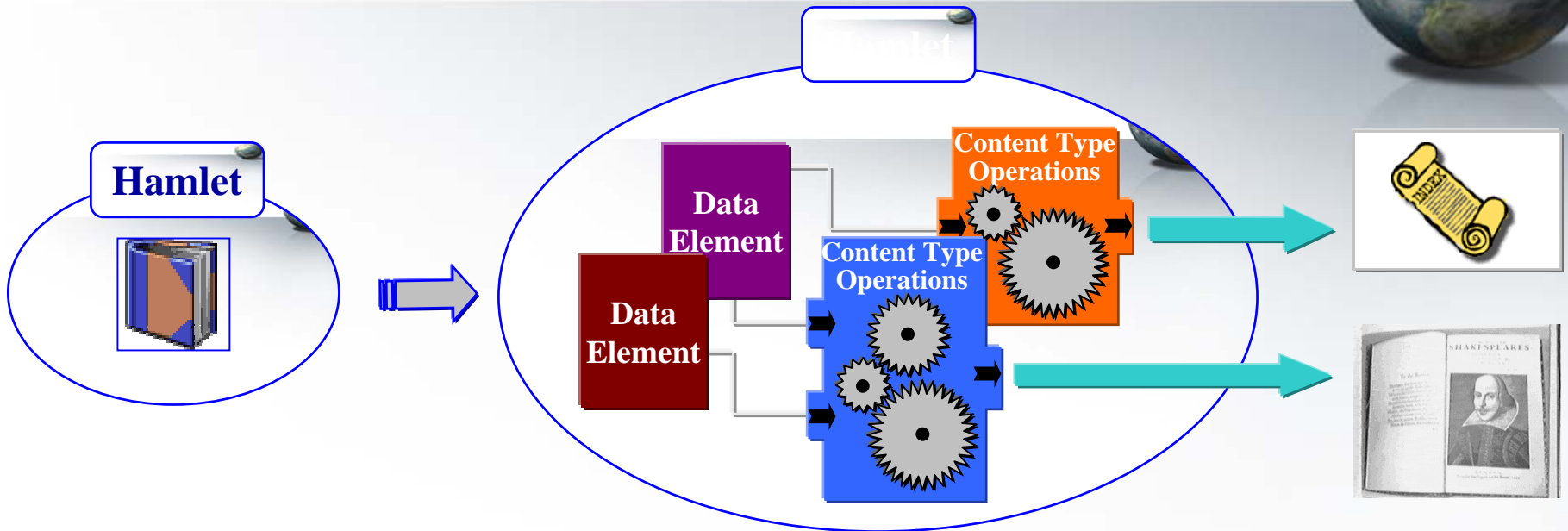
Digital Object Overview



Digital Object Overview



Digital Object Overview

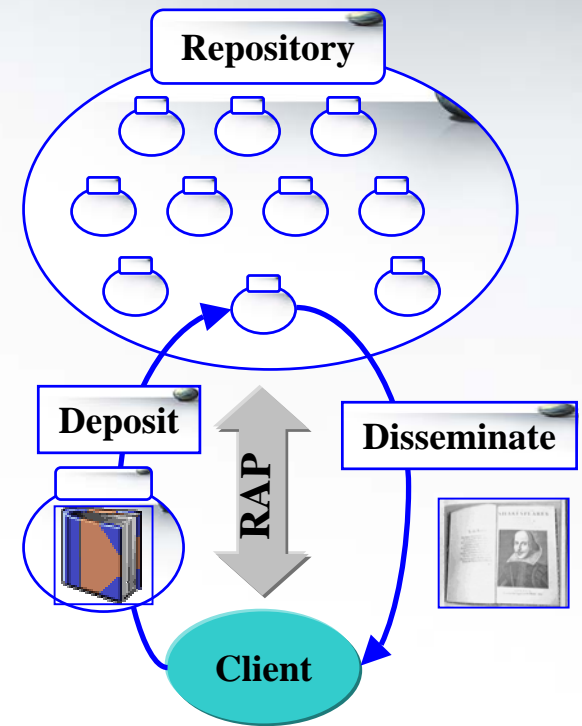


- Digital objects are uniquely identified in a given identifier space.
- Data elements reference sequences of typed data.
- A Digital Object can have zero or more *Content Types* to reflect intended uses by its creator.
- Content Type Operations are accessible as DOs

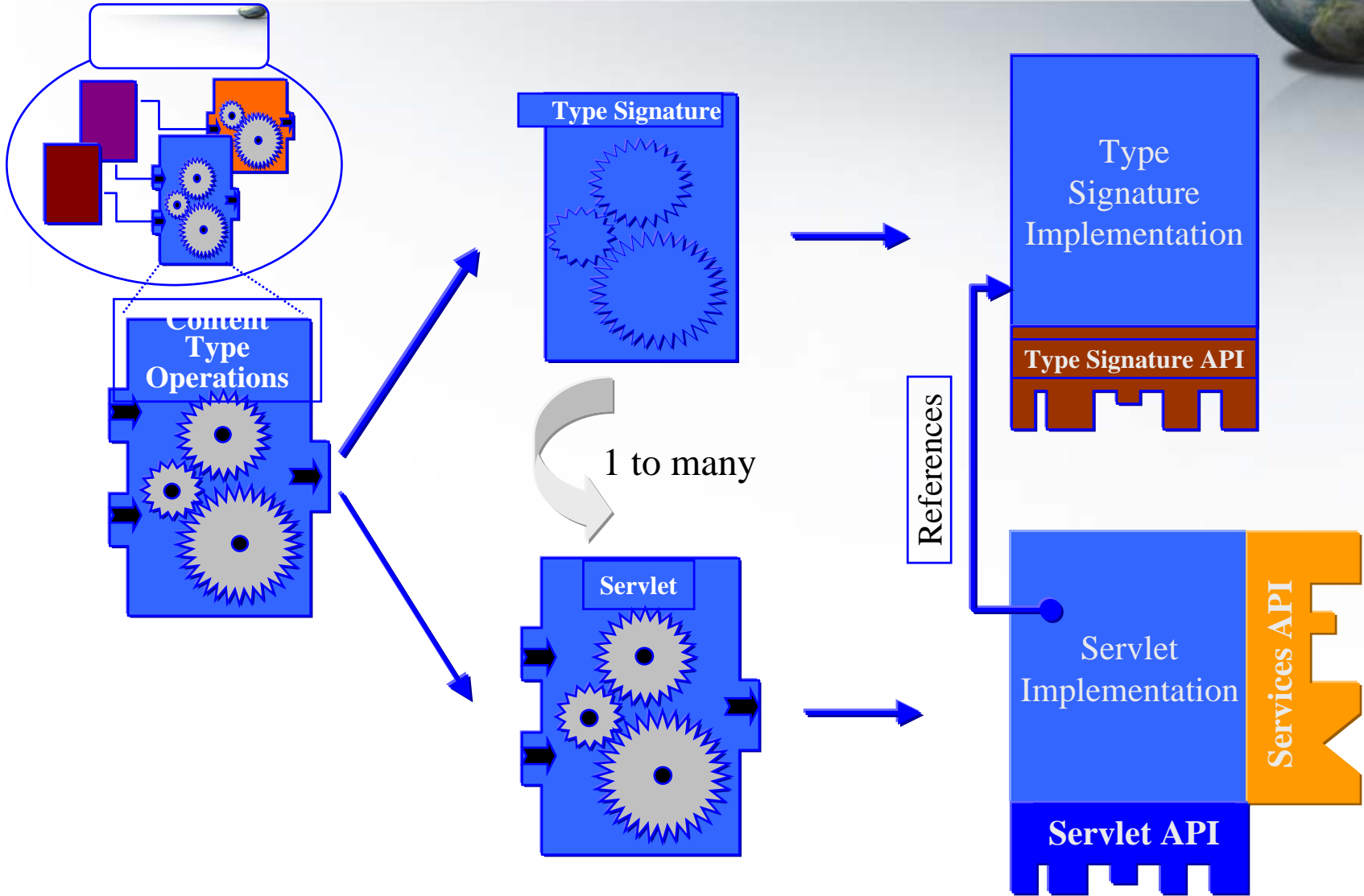
Digital Object Repository



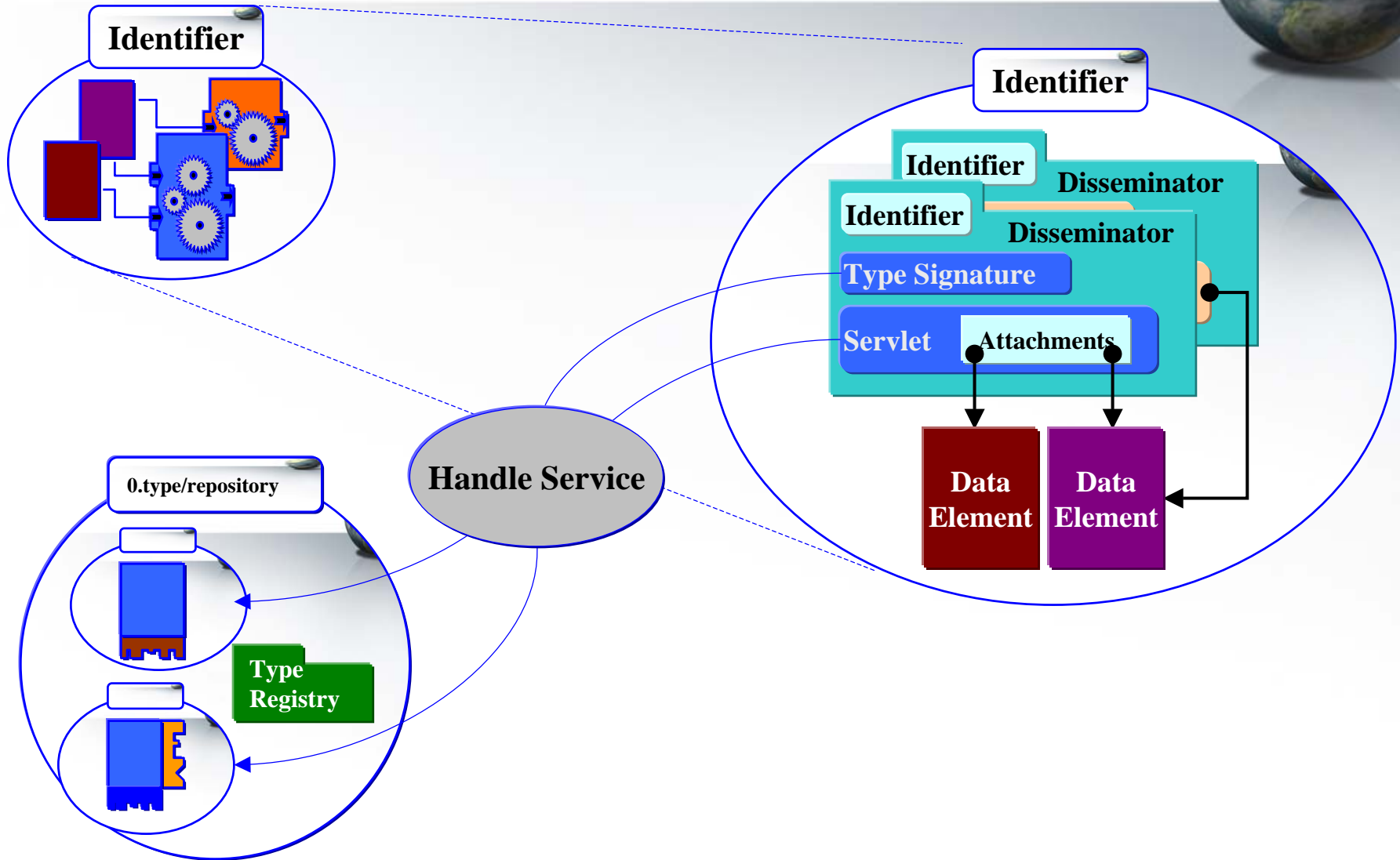
- Provides distributed Digital Object storage.
- May itself be a Digital Object.
- Provides a dynamic acquisition and execution mechanism for the mobile code that implements the content type operations.
- Exclusively accessed using the Repository Access Protocol (RAP).



Content Type Extensibility



Digital Object Structure



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[®] (GHR) administrator
- Site is under re-development to accommodate widespread use via automated means
- Experimental Repository software also available on-line

Business Potential



- Selling infrastructure technology
- Providing identification, management and Metadata services
- Enabling third-party value-added capabilities
- Helping organizations manage their own information better & offer new types of services
- Stimulating access to “surface information” and “embedded information” with appropriate access controls and conditions of use

Conclusions



- Managing Digital Objects for long-term access is the challenge
- Technology Components are available from R&D
- Interoperability is a critical objective
- Applications (with user-friendly interfaces) need to be developed & deployed
- Metadata registries need to be created and maintained
- Third-party value-added providers will ultimately shape the long-term evolution
- Infrastructure for managing information over indefinite periods can fundamentally alter the net and how we use it
- With profound impact on both business and society



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
- These concepts will diffuse down to most aspects of network management