Trusting Digital Entities

by
Robert E. Kahn
Corporation for National Research Initiatives
Reston, VA

A presentation for Data61
Melbourne, Australia
March 14, 2018

What is the Internet?

- Original FNC Definition still applies:
  - Global Information System that makes use of IP (its logical extensions or follow-ons), TCP (its logical extensions or follow-ons, and other IP compatible protocols), and which supports applications based on the above.
- Overall Architecture is still intact despite increases in the underlying technology by factors of 1 – 10 Million (computation, communication and storage)

Bindings to Technology vs. Information

- Arpanet – 16 bit addresses → wires
- Internet – 32 bit IP addresses → machines
- Web - URLs → <IP Address/filename>
- DO Architecture – DO Identifiers → DOs
  - DO Architecture describes a means of managing information over both short and long time frames in which Digital Objects are the basic structures.
  - Compatible with the current Internet and builds upon it.

Fundamental Properties of the Digital Object (DO) Architecture

- Logical Extension of the Internet
- Based on the same architectural ideas embedded in the Internet's architecture, and which have sustained its evolution, the three most important characteristics being:
  - **Open Architecture** (defined protocols & interfaces)
  - **Independence** from the underlying technology
  - **Minimized Complexity** for users
- The DO Architecture enables interoperability across heterogeneous information systems, whether in the Internet or not.
- It is a non-proprietary architecture and is publicly available.

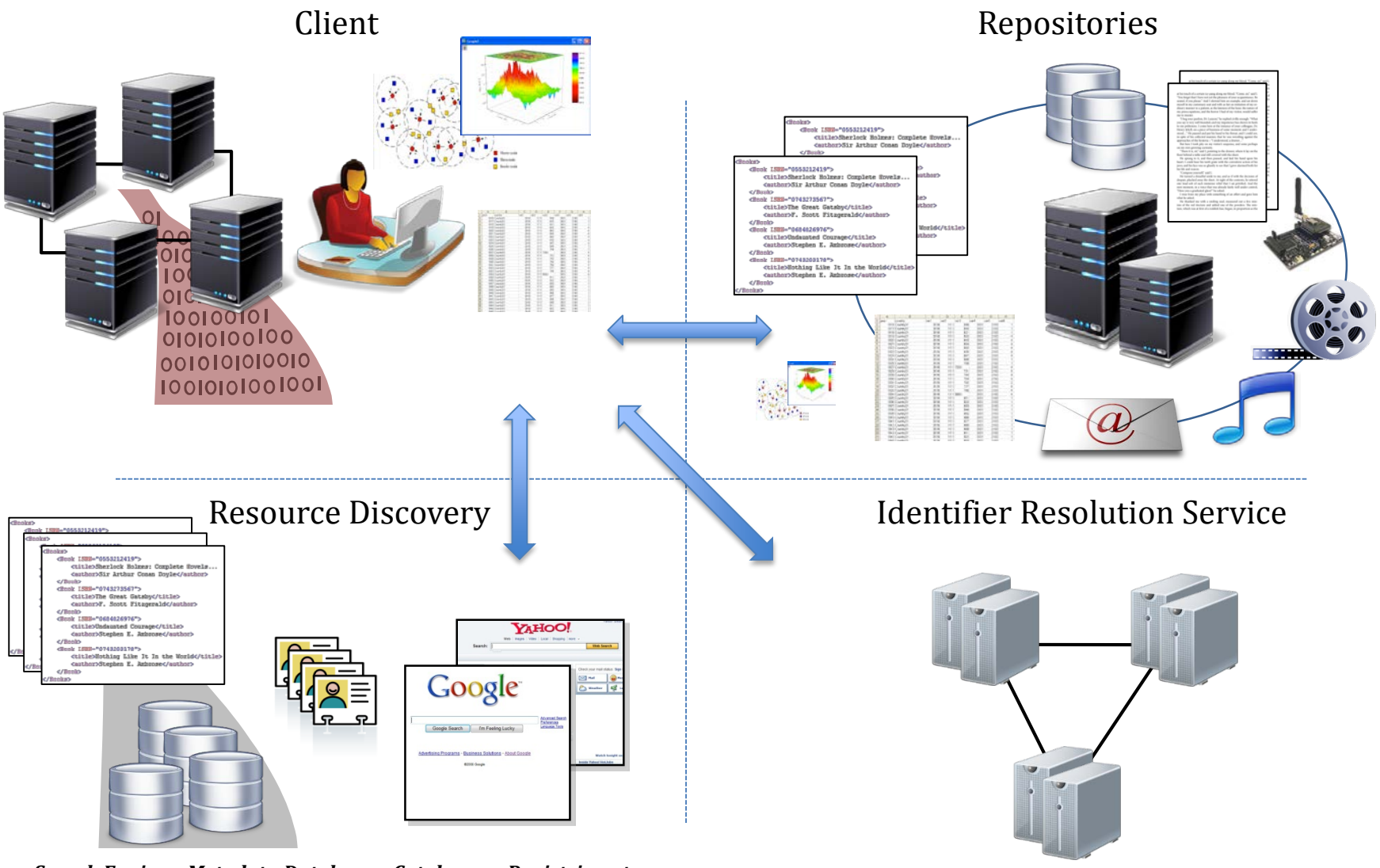
Basics of the DO Architecture

- **Digital Objects** are its basic structures (also known as Digital Entities). Each DO consists of information represented in digital form, and having an associated unique persistent identifier.
- The DO Architecture consists of three components:
  - a resolution component that resolves identifiers to "state information" about the desired information - - a resolution request yields a handle record;
  - Repositories that store DOs and enable access via their identifiers; and
  - Registries that store metadata about DOs and are used for searching.
- Resolvable "data types" are critical to understanding a DO by computer or otherwise.

ISO Effort to define the structure of Types

- Not intended to define specific “types”
- But rather what a type specification should look like, where each type is represented as a separate DO with its own unique persistent identifier
- Every element of every DO consists of a pair of (type, value) entries.
- And every type is represented by its identifier.

Digital Object Architecture: Information Management on Networks



Search Engines, Metadata Databases, Catalogues, Registries, etc.

Critical Role of Identifiers in the DO Architecture

- Identifiers are used to designate users, system resources, networks, services and desired information of all kinds represented in digital form and structured as digital objects.
- The resolution system provides important real-time information to client software.
- Everything being identified has a public/private key pair; and the public key is accessible by resolving its identifier.
- This enables an integrated PK Infrastructure (PKI) that is essential for purposes of providing security and generating trust.

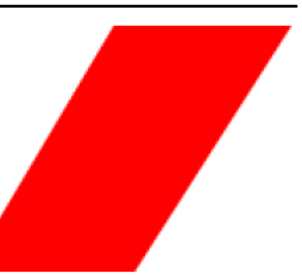
More Background on DO Architecture

- Started with the work of Bob Kahn and Vint Cerf at CNRI on mobile programs in the 1980s (i.e., Knowbots)
- Elaborated upon in the early 1990s in the Computer Science Technical Reports (CSTR) project.
- In 1997, the Cross-Industry Working Team (XIWT) supported the concept of digital objects and “stated operations” on digital objects, and noted the importance of chaining operations and managing value.
- The DO Architecture received the Digital Id World Award in 2003 for balancing innovation with reality.

Cross-Industry Working Team

Managing Access to Digital Information:

An Approach Based on Digital Objects and Stated Operations



3Com
Alcatel Telecom
American Management Systems
Apple Computer
AT&T
BBN
Bell Atlantic

May 1997


Digital Object Architecture
Digital Identity World 2003 Award
For balancing innovation & reality

DOIP Protocol

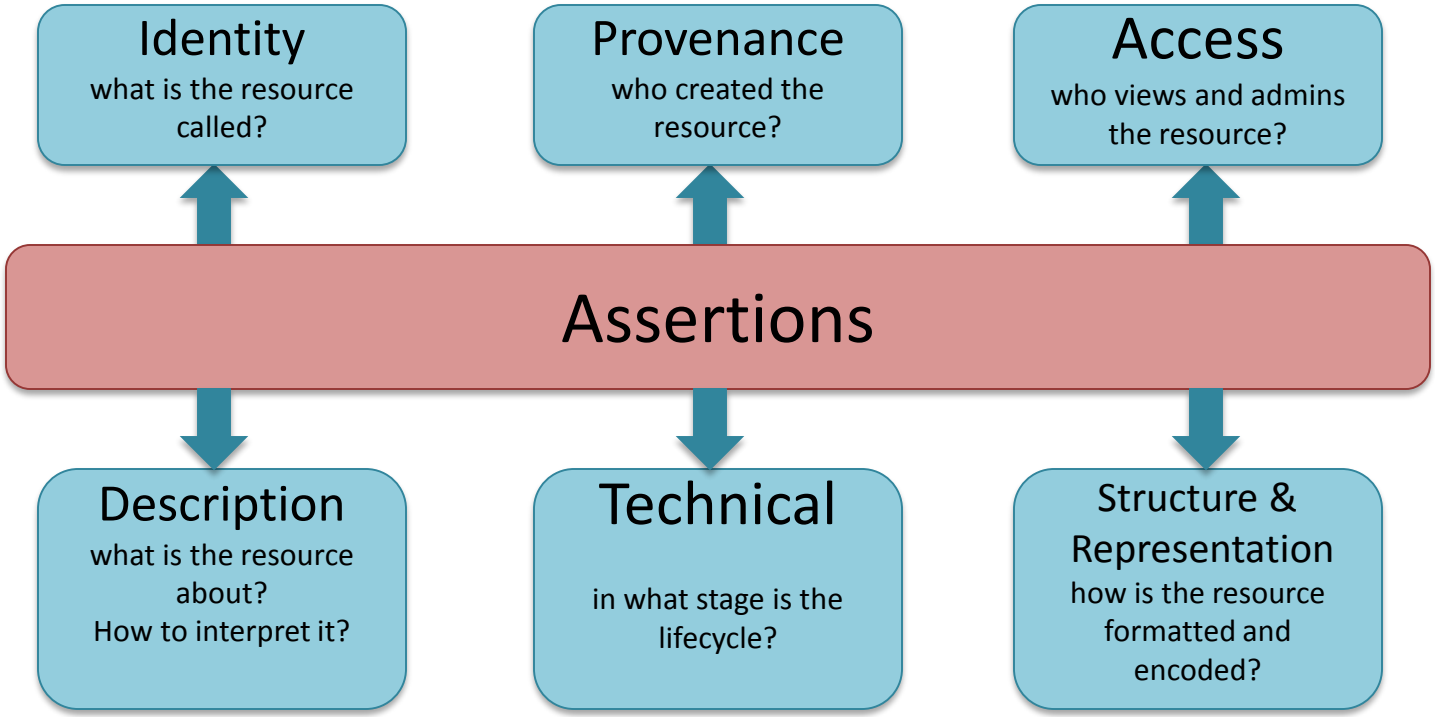
- The Digital Object Interface Protocol (DOIP) is a simple, but powerful conceptual protocol for software applications ("clients") to interact with "services" which could be either the digital objects or the information systems that manage those digital objects.
- The DOIP enables a user (or another DO) to interact with a DO based on the use of associated identifiers
  - Each action is represented by a DO; and the interface conveys the action's identifier (ID1);
  - Each target of an action is also a DO; and the interface conveys that identifier as well (ID2);
  - The formal specification is written as a schema that is incorporated in a program typically run by a repository that serializes structured data.

Framework for Discovery
ITU-T Recommendation X.1255

- Based largely on the Digital Object Architecture, ITU-T Recommendation X.1255: “Framework for discovery of identity management information” was approved in September 2013.
- Focused specifically on discovery and access to information in digital form, structured as digital objects, X.1255 is applicable to operational requirements for information management more generally.
- For purposes of X.1255, a digital object is defined as a digital entity; and the Recommendation describes a data model and interface protocol.
- Since the notion of a DO and a DE are nominally identical, from here on I will simply refer to both of them as Digital Entities (DEs).

What is Metadata

- People commonly define metadata as “data about data”
- A more complete definition:
  - Metadata is a set of (structured) assertions about an entity
  - Multiple parties may make those assertions
  - Veracity of those assertions is usually outside the scope of metadata
- Those assertions could be about



What is a Block?

- Blocks are not new!
- Historically, a block was viewed, essentially, as a sequence of bits, usually with a defined beginning and end.
- In the past, it may or may not have been uniquely identified other than by its arrival sequence in time.
- Blocks were also linked with other blocks
  - **In the programming field, blocks were often linked or chained using pointers**
  - **In the communications field, they were usually linked in some time sequence and often involved encryption.**
- Blocks were not usually managed separately from the application that invoked them- -but they could be.

Accessing Information about a Block

- This is the province of what is now called **metadata**.
- Part of the metadata may be self-contained within the block. This is sometimes referred to as key metadata.
- The amount of information that may be termed metadata about a block can be enormous and would normally be managed separately from the block.
- The use of blocks in information management was pioneered by CNRI in connection with mobile programs in the Internet.

General Observations about Blocks & Blockchains

- As previously mentioned, the context for the development of blockchain technology has been around for many years. Indeed, every block is an example of a Digital Object.
- A blockchain represents a particular way of structuring a Digital Object that comprises multiple DOs.
- DOs are stored in Repositories, which may be replicated (otherwise known as mirroring).
- Various mechanisms can be invoked to cross-check the multiple repository entries, if deemed necessary to augment trust.

Managing Mutable & Immutable DEs

- Blocks may be immutable; blockchains may not change, but are inherently mutable as they need to change when they are updated.
- Examples of blocks are transactions, contracts, bills of lading, digital cash; for example, see "Representing Value as Digital Objects: A Discussion of Transferability and Anonymity."
- Immutable objects can be authenticated without reliance on external parties.
- Mutable objects rely on external mechanisms to validate.

Authenticating a DE

- If a DE has been signed by a user or a system resource with its private key, the DE will identify the signatory (or it will be conveyed separately by the access protocol); and the DE can then be validated from the signature.
- Parts of a DE can be signed or encrypted, if desired.
- If so, the usual approach is to treat that part as a separate DE in its own right and link to it from the other DE by including its separate identifier in the first DE.
- Alternately (and often in addition) the handle record obtained by a piece of client software for a given DE will contain the authentication information for that DE; and the handle record may be signed by the server from which it was obtained.

Authenticating an Immutable DE

- If a DE is known to be unchangeable in a given context (e.g., a contract, digital cash, or bond), then a simpler mechanism is available to authenticate the DE.
- Namely, the identifier can contain a powerful cryptographic hash of the DE (with an associated methodology to use the hash for authentication) so that the DE can be validated no matter how it is provided or obtained.
- In this case, the user need not rely on anything other than the strength of the encryption mechanism.
- And, a failure in any one case (perhaps due to a loss of a private key) will not compromise the rest of the system.
- Replication of the DE in multiple instances of a repository will increase the likelihood that a valid version of the DE can be accessed in the unlikely event of repository failures.

Trusting the Resolution Mechanism

- A key part of the Internet is the IP Addressing mechanism that is used to route packets from source to destination.
- Similarly, a key part of the DO Architecture, which is a logical extension of the Internet, is the identifier/resolution component.
- It is used to map identifiers for *digital entities* to useful state information about them: that system must be trustworthy as well.

Two Stage Resolution

- CNRI implemented a two stage resolution system in the early 1990s in which identifiers have the structure "**prefix/suffix**". This implementation is in widespread use with more than a billion digital entities identified. The prefix, which is allotted to a specific party that wants to create resolvable identifiers, is unique to that party; and that party would start its identifiers with its prefix and add whatever suffix it wishes.

- Derived prefixes may be created by the party using a "dotted" convention. For example if prefix 35 is allotted, 35.1 or 35.HQ.1 may be derived from 35. The zero and one delimiter prefixes are retained in a distributed registry called the Global Handle Registry (or **GHR)**. Multiple organizations around the world operate the GHR and coordinate with each other in maintaining its integrity.

- The actual identifier records - such as those corresponding to 35.1/abc - are retained in one or more local services they run, or contract to have run for them, and also managed by the party that created them. The system is inherently distributed. The local services can also be mirrored for reliability and security, as desired; and most organizations choose to do so.

What changes are in process?

- Fundamental changes took place in the Internet as the number of devices exceeded what were then a staggering number – like 100 Million.
- Today, it is envisioned that the number of devices in the IoT (or cyber/physical systems more generally) may come close to 100 Billion in the not too distant future.
- This will stress almost every aspect of the Internet - and especially those that involve information management.
- Many organizations are rallying behind the use of blockchains to provide trust, but this is but one of several alternatives; its use will provide its own challenges for managing information. Issues of interoperability as well as scalability, efficient performance and graceful degradation must be balanced against the need for architectural changes to provide enhanced defenses.

As the Internet Confronts Increased Complexity

- Mobile program technology may soon be needed in the context of implementations of the DO Architecture.
- Trust in the system of information management and the digital entities it manages are critically important especially when the DEs have value, as is the case with crypto-currencies.
- The need to protect rights, values and other interests that may be embodied in DEs, coupled with the sheer volume of information that will be available in digital for, requires a new paradigm for information management.
- The Digital Object Architecture can provide a sound basis for moving forward.

Some Background Reading

- Kahn, Robert E., Vinton G. Cerf, "An Open Architecture For a Digital Library System and a Plan For Its Development," The Digital Library Project Volume I: The World of Knowbots, (DRAFT) March 1988, http://hdl.handle.net/4263537/2091.

- Kahn, Robert E., Robert Wilensky, "A Framework for Distributed Digital Object Services," *International Journal on Digital Libraries*, (2006) 6(2): 115-123, https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf. (First published by the authors May 13, 1995, "A Framework for Distributed Digital Object Services", http://hdl.handle.net/4263537/5001).

- Managing Access to Digital Information, Cross-Industry Working Team, May 1997, http://www.xiwt.org/documents/ManagAccess-1.pdf.

- Denning, Peter J. and Robert E. Kahn. "The Profession of IT: the Long Quest for Universal Information Access". *Communications of the ACM*, December 2010, Vol. 53, No. 32, pp. 34-36, http://doi.org/10.1145/1859204.1859218.

- ITU Recommendation X.1255, "Framework for discovery of identity management information," was approved on September 4, 2013 (the work is based largely on CNRI's Digital Object Architecture; and Robert E. Kahn, CNRI's President, served as Editor), http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=11951&lang=en

- Kahn, Robert E., "The Role of Architecture in Internet Defense", America's Cyber Future: Security and Prosperity in the Information Age, Center for a New American Security (CNAS), Volume II, Chapter XII, May 2011, http://www.cnri.reston.va.us/papers/CNAS_CyberSecurity_Kahn.pdf.

Some Background Reading (Cont'd)

- Braswell, Jefferson, Lannom, Larry, Milne, Alistair, Northey, Jim, Paskin, Norman and Traub, Ken, "Response to the Financial Stability Board's Request for an Engineering Study on the Best Approach to Managing the Structure and Issuance of Legal Entity Identifiers (LEIs),"(2012), http://doi.org/10.2139/ssrn.2197269.

- Kahn, Robert E., and Patrice A. Lyons, "Representing Value as Digital Objects," *Journal on Telecommunications & High Technology Law*, Vol. 5, Issue 1 (2006), http://www.jthtl.org/content/articles/V5I1/JTHTLv5i1_KahnLyons.PDF.

  A patent application (US 20030233570 A1), titled ***Authenticating and using digital objects,*** *and based in part on the ideas expressed in this article, was filed by CNRI. It specified that the technology may be applied in managing, inter alia, the issuance and authentication of financial instruments). The application was later abandoned when the claims were rejected by the U.S. PTO as covered by the now expired, CNRI Patent No. 6,135,646,* ***System for Uniquely and Persistently Identifying, Managing, and Tracking Digital Objects.***

- Lyons, Patrice A. and Robert E. Kahn, "The Handle System and its Application to RFID and the Internet of Things", in *RFIDs, Near-Field Communications and Mobile Payments; A Guide for Lawyers*, edited by Sarah Jane Hughes, Cyberspace Law Committee, 2013, http://hdl.handle.net/4263537/5046.

- *Definition of "block" as a "Digital Entity,"* Contribution of Corporation for National Research Initiatives (CNRI), DLT-I-048, ITU Focus Group on Application of Distributed Ledger Technology (Bern, Switzerland; 2018), http://www.cnri.reston.va.us/documents/DLT-I-048_CNRI_Contribution_FG_DLT.pdf