Overview of the Digital Object Architecture

1. Overview

The Internet was designed and implemented as a general purpose platform to provide global connectivity among computers, devices and networks of all kinds. It did not single out any particular application for special treatment, but rather provided a means by which any application could be made available publicly or for use only by authorized users for any legal purpose.

In the mid 1980s, it became apparent that information management was a kind of specialized application that had the potential for being a platform that made use of the existing Internet infrastructure, but offered a new set of capabilities that were likely to be of widespread utility. The Digital Object Architecture ("architecture") resulted from that insight and its development proceeded in parallel with an alternate effort called the World Wide Web. Since that time, both capabilities have become widely used, although the web acquired more rapid acceptance since it focused primarily on public information, limited security and near-term access.

The Digital Object Architecture was designed to enable all types of information, whether public, private, or some combination thereof, to be managed in a network environment over potentially very long time frames; and it includes the capability for integrated public key management within the same infrastructure used to manage other information. Indeed, any information expressed in digital form can be securely managed within the architecture.

The core components of the architecture have all been implemented and made available on the internet by Corporation for National Research Initiatives (CNRI). The infrastructure aspects of the architecture allow interoperability to be easily achieved technically by parties that wish to enable it administratively. The components may be used, virtually without change, for most forms of information management, subject only to limits of bandwidth, available storage, and computational demands. From an infrastructure perspective, the technology for identity management differs little from the technology for any other kind of information management, and thus readily enables system security and interoperability. Further, the technology for an "Internet of Things" is essentially about managing information about the things and could thus be managed using components of the architecture.

Most detailed system choices are left to the parties making use of the system. For example, the terms under which information can be provided may be included within the information that can be accessed, but it is supplied by the party establishing the specific terms. Existing identifier systems may be used within the architecture, which can also support interoperability between the different systems. Existing identity management systems, each of which uses its own version of unique identifiers as part of the process by which identity is established, may be accommodated within the architecture.

2. Core components

The basic architecture contains three distinct component types, of which an unlimited number of instances of each type is possible. The architecture is independent of the underlying technology used to implement the components, just as the Internet architecture is independent of the choice of component networks that comprise it. In other words, the implementation of the architecture is scalable along dimensions of size, and is evolvable with technology, so that, if

properly managed, information created now will be accessible within the architecture independent of changes in the underlying technology for the indefinite future. The principal components of this open architecture developed to date are:

2.1 **The Digital Object Repository** (or dorepository), which stores digital objects and provides access to them. There can be an unlimited number of such repositories in the system. The dorepository consists of software that manages the objects and the storage system being used to hold the objects (presumably using standard commercial products) chosen by the party running an instance of the dorepository. Each digital object is assigned a unique persistent identifier upon deposit (or even prior to deposit) and all access to the dorepository is based on the use of identifiers. The set of dorepositories is interoperable subject to secure administrative controls that are built into the interface protocol. The interface protocol itself is extensible and its functions are designated by various identifiers. One important aspect of this architecture is that digital objects may be moved from one storage system to another and even from one dorepository to another with a few instructions from its administrator. In the process, metadata, including provenance information and access control information, are all preserved. The dorepository software may be downloaded from the http://www.dorepository.org site.

2.2. **The resolution system**, a principle function of the Handle System, maps known identifiers into handle records containing useful state information about the digital object being identified. Entries in the handle records are expressed in Unicode, using UTF-8 encoding, and can support multiple languages. Handle records typically contain location information, such as one or more IP or MAC addresses, and may also contain other important "state information" such as certain terms and conditions for use of the object, information to verify the authenticity of the object, or public keys, where appropriate. If the identifier is allotted to information about an individual or other resource for purposes of identity management, the handle record would contain the public key assigned to that individual or resource, and other parties would know to access the Handle System to obtain the relevant public key.

The Handle System has been deployed on the Internet for over 15 years; the system is in daily use and is extremely reliable. The Handle System technology and software, including the Handle System interface specification, is available at the http://www.handle.net internet site. Local handle services may be provided by any organization that wishes to do so. For parties wishing to provide identifier and/or resolution services using the Handle System technology, there is a one-time requirement to register and obtain an authorization number, known as a prefix, that, together with a suffix typically assigned by that party, forms the digital object identifier (also known as a "handle") for that information. The use and management of prefixes ensures that duplicate identifiers are not placed in the system. The use of unique persistent prefixes facilitates efficient resolution from local handle services.

2.3. **The Digital Object Registry** (or doregistry) is used to define collections of digital objects that may each exist across one or more repositories and which permit browsing and searching. The results of a registry search is a set of identifiers, perhaps embedded in a displayable mark-up format to be viewed by a user or mapped into a parsable digital string that is sent back to the requesting program. The registry uses the dorepository to store its metadata records, and thus inherits the same security capabilities that would be

manifest when a user tries to access private information in a repository. While earlier versions of the doregistry have been developed to meet the requirements of the distributed learning community, a generic version of the doregistry software has been developed and is available at http://www.doregistry.org.

The Digital Object Architecture provides a basic information infrastructure that can facilitate interoperability between or among different systems, processes, and other information resources, including different identity management systems. Much as the Internet did not supplant existing packet networks, but rather enabled them to work together, the digital object architecture does not necessarily replace existing information systems, but rather enables such resources to become interoperable, provides for long term persistence of the information (if properly managed), and facilitates secure sharing of such information. Many other aspects of interoperability involving digital objects are possible.

3. Security Considerations

Security in the Handle System is provided by several means. The basic administration of handle records is done under a PKI regime. Each handle record specifies one or more administrators for that record and those administrators can be authenticated using public/private key technology. The public keys are held in handle records associated with the administrators. In addition, each handle service has its own public/private key pair which can be used to verify that interactions with the identified service have not been compromised. The public keys for such handle services are currently in the Global Handle Registry (GHR) as are most of the public keys for administrators, although administrator keys can be elsewhere in the system. Management of the public keys in the GHR will change as the global management of the system evolves.

Providing security at the level of individual digital objects in dorepositories is in addition to other requirements to provide separate security for information on other parts of the infrastructure (other than, perhaps, physical security for the infrastructure components). The architecture enables PKI authentication for all parties involved in transactions as well as encrypted communication. Of particular importance is that interactions with digital objects can be authorized on a per object basis, or parts thereof. Finally, access to information in an instance of the doregistry is also subject to the same type of security control so that users can only search over information that they are authorized to access.

4. Registration Agencies

Certain groups that have implemented the components of the architecture have developed membership systems that rely on Registration Agencies to provide trustworthy identifier and other information management services in the Internet. A prominent example of such a system, called the International DOI Foundation (IDF), was established by the publishing community. The identifier/resolution system deployed by the IDF and its registration agencies, called the DOI® System, is a branded version of the Handle System developed by CNRI. Information about the IDF is available at http://www.doi.org.

Corporation for National Research Initiatives (July 28, 2012)