Using DOIP for Managing Scientific Data

Christophe Blanchi
DONA Foundation

GEDE Webinar
22nd February 2019

The DONA Foundation

- The DONA Foundation is Based in Geneva Switzerland.
- From its statutes it has two clear and distinct purposes:

1) Provide coordination, software, and other strategic services for the technical development, evolution, application, and other uses in the public interest around the world of the Digital Object Architecture (DOA) and related X.1255 standard with a mission to promote interoperability across heterogeneous information systems across many different countries, domains, and industries.

2) Administer and maintain the stable operations of the Global Handle Registry (GHR), a key component of the DOA, and shall authorize and coordinate its administration with multiple multi stakeholders across the world know as Multi-Primary Administrators (MPAs).

Motivations for the Digital Object Architecture

- Need to facilitate the secure discovery, access, exchange, and reuse of information represented in digital form across a wide range of types and sources of data.
- Make information a “First Class Citizen” in the Internet.
  - Information needs to be identifiable and locatable.
  - Information needs to be accessible in a consistent manner.
  - Information needs to be understandable to be reusable.
  - Information needs to be attributable, protected, secure, and trusted.
  - Information needs to be able to be used with the simplest to the most complex data and services.
  - Information needs to be able to persist over time.
  - Information need to be managed according to local policies.

Digital Object Architecture Overview

- The DOA is a 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 of which are:
  - Open Architecture. Two defined protocols: Handle Protocol, DOIP
  - Independence from the underlying technologies.
  - Minimized complexity for users.
- Provides a framework for managing digital information and services of all kinds whether in the Internet or not using the Digital Object (DO) as its basic structure.
- Built-in extensible typing system for describing new types of resources such as data, operations, services, interfaces, workflows, and protocols to name a few.
- Built-in Security and PKI.
- Distributed and highly scalable

DONA

Digital Object Architecture: 3 Components

Digital Object Architecture: Two DO Protocols

- Handle System Protocol – Access to a global resolution system
  - Allotment of globally unique identifiers known as handles.
  - Provides rapid secure identifier resolution into a handle record/digital object.
  - Provides distributed administration of handles and their records.
  - Has built in PKI and security.
- DOIP – Interoperable access to distributed, independently managed services and data sources.
  - DOIP is a protocol to interact with digital objects.
  - The DOIP Service (DO Service) can be integrated with existing systems.
  - Leverages existing transport encryption standards such as TLS.
  - Provides built-in security leveraging the PKI of the Handle System.
  - Leverages the handles system for identifier resolution.

Digital Object – The Base Common Data Model

- **id:** the identifier of the DO.
- **type:** The type of the DO.
- A sequence of bytes. DOIP defines a suggested structure
  - **attributes:** 0 or more key-value pairs.
    - **key:** ID of a DO Type describing the type of the value
    - v**alue:** the an instance of the associated type.
  - **elements:** 0 or more elements each consisting of:
    - **id:** id of the element. Must be unique in the DO.
    - **type**: the type of the element.
    - **Length(optional)**: the length of the data element.
    - **attributes**: 0 or more key-value pairs.
    - **data**: the payload of the data element.
  - **signature** (conditional)

Digital Object Interface Protocol : Request

- **requestId**: The identifier of the request provided by the client to keep track of responses.
- **clientId**: The identifier of the requesting client.
- **targetId**: The identifier of the DO on which to perform the operations
- **operationId**: the identifier of the operation to be performed on the DO.
- **attributes(optional)**: Stipulated by the Operation ID.
- **authentication(optional)**: used by clients to authenticate.
- **input(optional)**: Payload stipulated by the operation.

Digital Object Interface Protocol : Response

- **requestId**: The request identifier provided in the request by the client.
- **status:** The status code identifier that indicates the status of the request.
- **attributes(optional)**: Operations stipulated.
- **output**: The presence and nature of the output is specific to each operationID.

Two Synergistic and Complimentary Protocols

- The Handle System is a global, rapid, and secure resolution service that resolves handles into digital object structured handle records.
- For some applications, the Handle System is a complete solution.
- Other applications need additional service interoperability capabilities that DOIP provides.
  - Big Data: large data sets with different, non-interoperable access methods.
  - IoT devices: many different sorts of protocols with varying taxonomies.
  - Other uses cases require new and different security solutions.
- The DOIP relies on the Handle System for its PKI and to resolve a Digital Object identifier into information needed to connect to the appropriate DOIP Service.

Interacting with Digital Objects - Handles

55.123/AE0-123-5BA

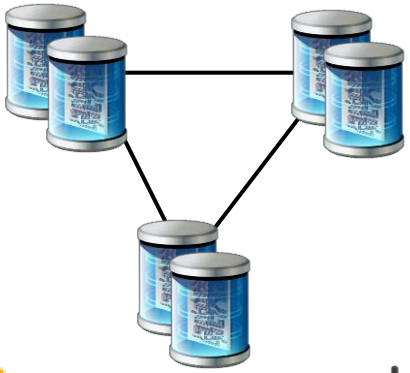
Client

Resolve ID

Handle Protocol

Handle Record
/ Digital Object

Handle Service



55.123/AE0-123-5BA	TYPE	4	0.TYPE/123-456
	META	10	<meta id="keywords" content="Data
	HASH	20	998fe8dd4a0accbe6e2...
	HS_SIG	100	eyJhbGciOiJSUzI1NiJ9.....

doi

EIDR

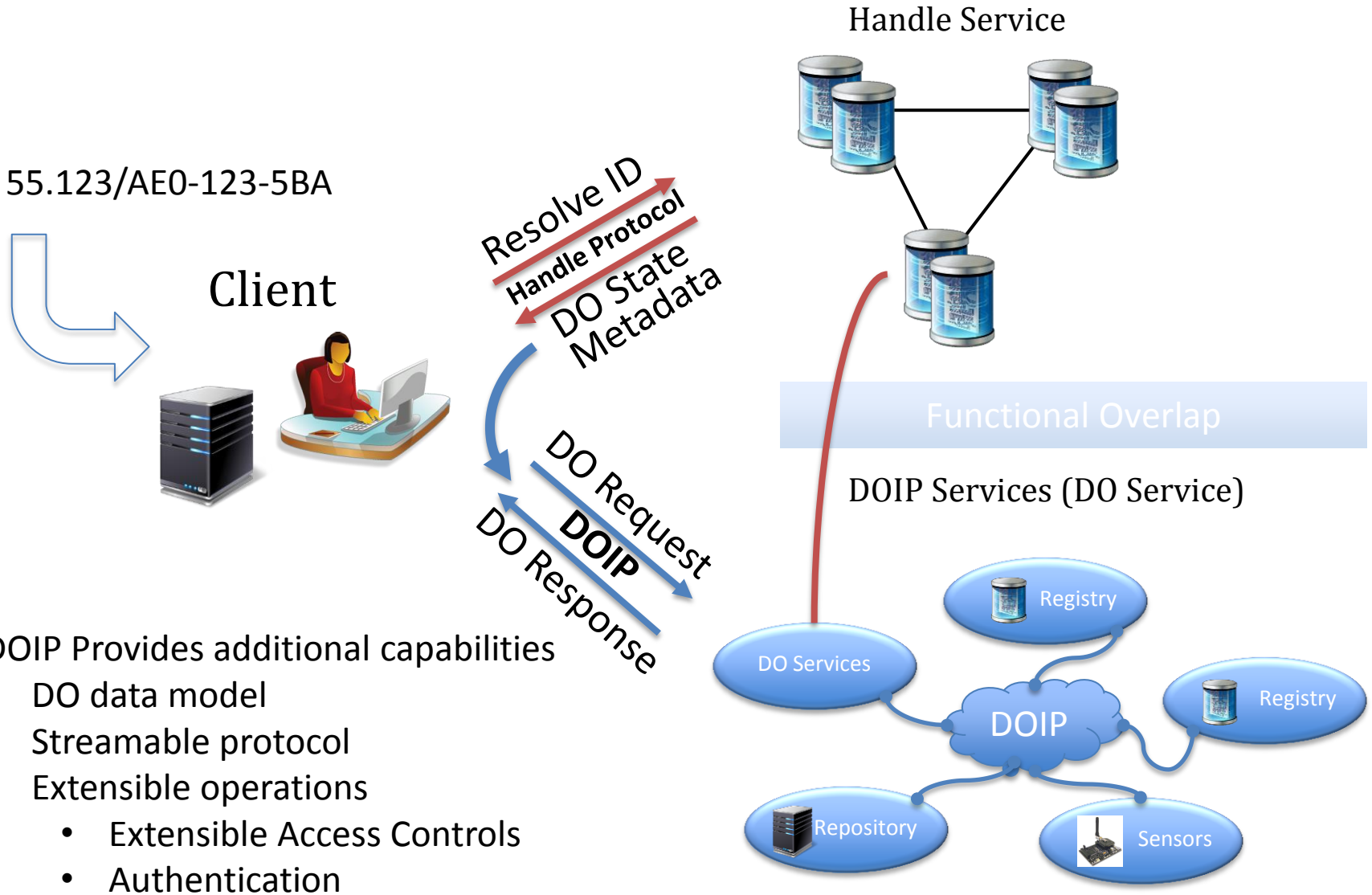
ePIC
Persistent Identifiers for eResearch

Type
Registry

Counterfeit and
Supply chain management

DONA

Interacting with Digital Objects – Handles and DOIP

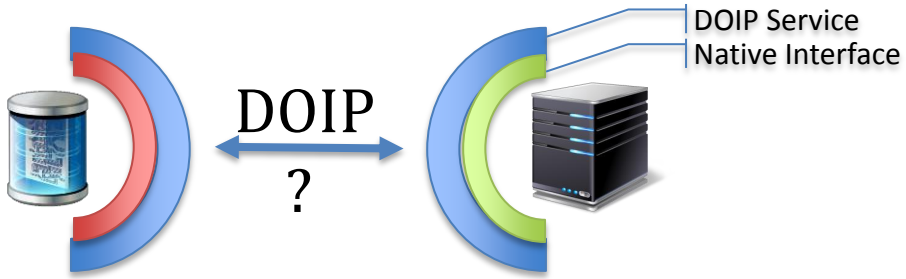


DOIP Provides additional capabilities

- DO data model
- Streamable protocol
- Extensible operations
  - Extensible Access Controls
  - Authentication

Integrating Existing Systems with DOIP

- DOIP is a simple protocol for exposing existing data, services, and things.
- DOIP is typically implemented on top of existing systems to expose needed data and services using registered types and operations.
- The Digital Object model is the basic data model for returning data and metadata. The use of extensible operation can extend this model.



- DOIP provides a common yet extensible basis for data interoperability:
  - Common name space.
  - Common basic data model.
  - Common global taxonomy, type, and typing framework.
  - Common global operation description framework

Additional DOIP/HS Considerations

- DOIP Services and Handle Services can operate in disconnected mode.
  - Requires careful namespace management.
  - For Data Acquisition: Disconnected LHS are authoritative and split across a large set of micro servers.
  - For Data Read only mode: Disconnected LHS are mirrors
  - Automatic synchronization takes place when a network is joined.
  - Mirrors provide consistent access but not always completely up to date information.

- This pattern can be used for various applications:
  - Low latency read/write requirements (5G applications)
  - On the field applications with spotty network.
  - “Sneaker net” applications.

Integration with Blockchain

- The Handle System and DOIP can be readily integrated in a bidirectional manner with Blockchain.
  - A handle provides a reference to an entry in the blockchain while also pointing to its identified digital object.
  - The entry in the blockchain includes a the handle. This is enables for large information to be referenced from the blockchain

- The blockchain is used to maintain the state of all changes made to the DOs.
- The DOA provides access controlled resolution, search, retrieval, and transformation over the data.
  - The handle and DOs can contain other integrity checks over their data such as multiple independently signed hashes, references to other related entities.
  - Current use case developed for ITU: membership voting.

DONA GHR Operations

- DONA' GHR has currently 9 registered MPAs and ITU.
- Provides a wide geographical coverage and redundancy of service.
- The GHR is now a self sustaining service.
- The current goal is to reach 12 MPAs by the end of 2019.

Thank you.

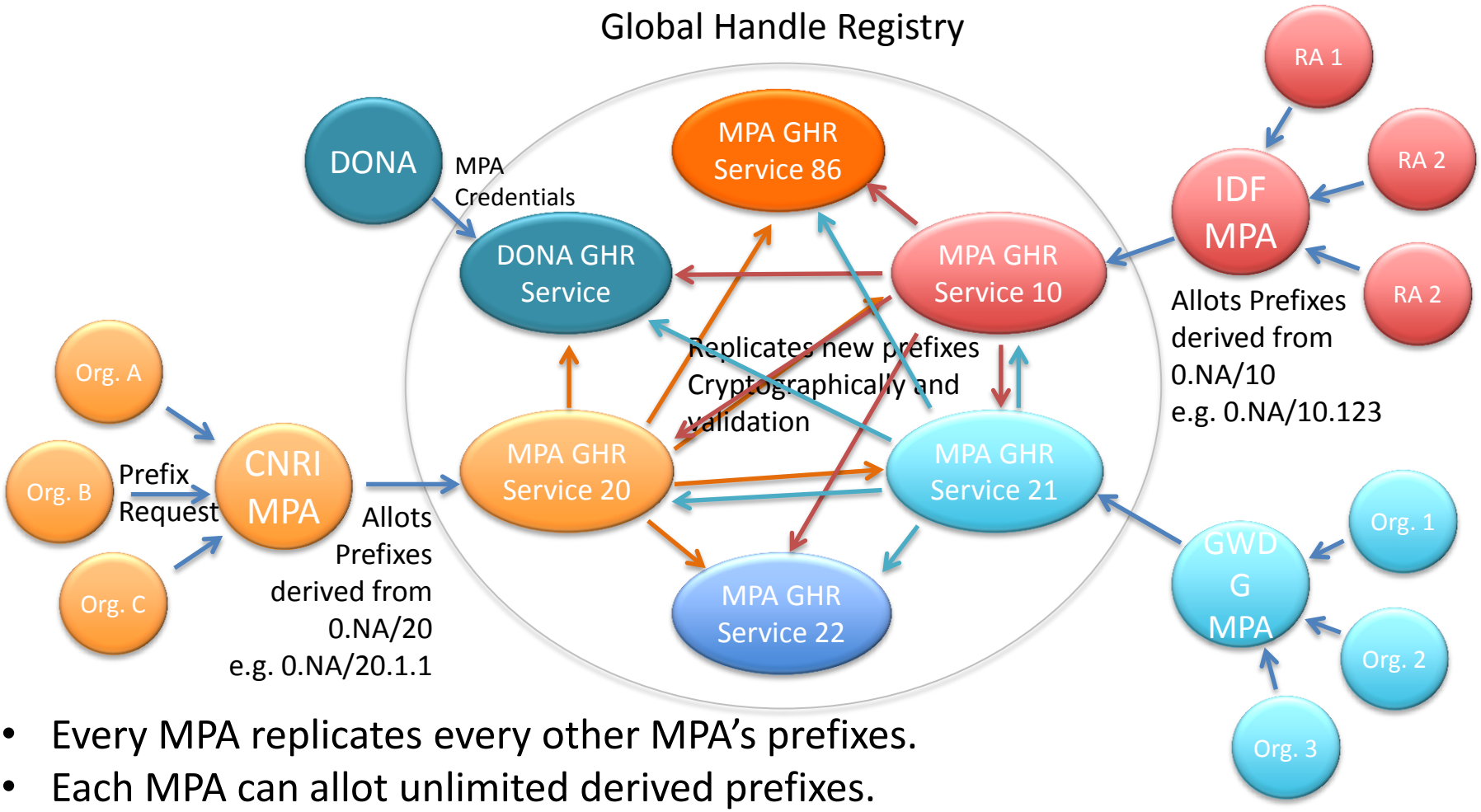
DONA DOIP Recent and Ongoing Activities

- CNRI transferred the rights to DOIP V2.0 to the DONA Foundation on the 19th of October 2018
- DONA Publically released the DOIP V2.0 Protocol on the 12th of November.
- DONA will publically release a free, open source, low level DOIP V2.0 client and server libraries targeted at developers before the end of the year.
- DONA, the MPAs and the community at large will collaborate to adopt and evolve DOIP.
- DONA is involved with the Research Data Alliance (RDA) to use the adoption of DOIP in big data project to establish interoperability between existing data repositories.

GHR Service Provider

- An organization that is credentialed and authorized by DONA to create derived prefixes from its allotted credential prefix is known as a Multi-primary Administration (MPA).
- Each such organization is allotted a credential (e.g. 0.NA/21) by DONA and authorized to provide GHR services.
- Each such organization can create and unlimited number of derived prefixes from its credential prefix and allot them to organizations that wish to provide local handle services.
- A maximum of 1 million 1 delimiter prefixes may exist on the GHR.
- All GHR Services verify and replicate any and all valid prefixes created/modified by other from all other MPAs and GSPs in accordance with DONA Foundation Policies and Procedures .

MPA GHR Operations



- Every MPA replicates every other MPA's prefixes.
- Each MPA can allot unlimited derived prefixes.
- An MPA can only allot derived prefixes from its allotted prefix.
- Only prefixes with 0 or 1 delimiter are replicated within the GHR.