Brief Introduction to Cordra & DOIP

RDA FAIR DO Fabric
Robert Tupelo-Schneck & Larry Lannom
Corporation for National Research Initiatives
24 March 2022
Cordra – what is it?

• Digital object server – provides access to data at the object level
• Reference implementation of DOIP (Digital Object Interface Protocol)
• See cordra.org for details
  • “Highly configurable software for managing digital objects at scale”
  • Open source
  • V2.4.0 released 20 Feb 2022
    • Base server modules
    • Cordra client library for Java
    • Cordra client library for JavaScript
    • DOIP (Digital Object Interface Protocol) Software Development Kit
    • Documentation and examples

• Small but growing developer community
  • Slack workspace (linked from cordra.org)
  • Email support@cordra.org
Cordra – where did it come from?

- At the implementation level, Digital Object Architecture has three parts
  - ID system
  - Repository
  - Registry (aka Repository for metadata)
- 15 year old CNRI registry project for DoD (Advanced Distributed Learning)
  - Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA) project
  - No longer in use, but the name stuck
- Fast forward a decade and multiple registry projects later
  - Every new registry was a new effort
  - Lead developer at the time wanted a turn-key registry that needed only a schema to provide basic registry functions.
  - A repository for metadata (objects) can be configured to serve up objects of any type
- Every new registry project has brought increased robustness and utility
  - Movie/TV industry (EIDR)
  - Financial sector (International Securities Identification Numbers (ISINs) for OTC derivatives.)
  - DiSSCo (biodiversity)
  - BSI.identify (construction products)
  - NIST (Material Science)
Cordra – what can you do with it?

• Produce a turn-key registry by feeding it a schema
• Run a combination registry/repository + handle server
• Supports MongoDB, Amazon S3, Elasticsearch, and Solr out of the box
• Data life cycle management
  • Validate input against schemas
  • Generate ids
  • Dynamic updates at read or write time
  • Add custom operations that read from or write to digital objects
  • PKI or password based authentication
• Declare access control at the object level
• Scale horizontally
Digital Object Interface Protocol (DOIP)

- At its core, DOA consists of two protocols
  - Resolve identifiers to get state information, e.g., object access logistics: handle system
  - Talk to the object to request operations: DOIP
- Abstract protocol
  - Request operations on objects
  - Requestor, Operation, Target Object are all specified as identifiers
  - Additional segments can include authentication, attributes, and input, depending on the operation and access permissions
- DOIP v2.0 specification (available at DONA.net) defines
  - Object serialization
  - Basic and extended operations
  - Role of types (every object has a type)
  - Communication protocol
- DOIP SDK for Java also available at DONA.net
Cordra Scalability

• Cordra makes it easy to either develop on a single machine for convenience, or deploy multiple instances of Cordra to a distributed environment.
  • Backed by scalable distributed storage (e.g. MongoDB, S3).
  • Backed by scalable distributed index (e.g. SOLR, ElasticSearch).
  • Synchronization using ZooKeeper and Kafka guarantee atomic operations and ensure that all stored objects are indexed.

• Benefits of a distributed deployment include:
  • No single point of failure
  • Increased request rate
  • Lower request latency when under load

• Size and performance considerations:
  • Production system in use today with 100 million digital objects. System tested with 1 billion objects.
  • Create requests (which include validation, storage and indexing for each request) measured at over 900 requests/second.
Demo