OPEN ACCESS

CrossMark

Conference Abstract

An Introduction to Cordra

Robert Tupelo-Schneck [‡]

‡ Corporation for National Research Initiatives, Charlottesville VA, United States of America

Corresponding author: Robert Tupelo-Schneck (schneck@cnri.reston.va.us)

Reviewable   v 1

Abstract

Cordra is a digital object server that can function as a key infrastructural piece in FAIR DO (findable, accessible, interoperable and reusable digital object) implementations. Cordra manages JSON records and payloads as typed digital objects identified by handles. Cordra is neither a database nor an indexer, but it integrates the two and provides a unified interface.

Cordra is intended to support both quick prototyping as well as production systems.

For prototyping, Cordra makes it easy to get up and running rapidly with a digital object server. A potential Cordra administrator can download Cordra and very quickly have a server which supports creation, search, and retrieval of digital objects with resolvable identifiers. The server supports Digital Object Interface Protocol (DOIP) and HTTP APIs out of the box, as well as an immediately usable prototype user interface. Cordra saves substantial development time as it comes with ready-made functionality ranging from user authentication and access control to information validation, enrichment, storing, and indexing. By default, Cordra is configured to store objects on the local file system of the machine and use embedded Apache Lucene for indexing. Simply by editing type definitions in Cordra's user interface, the administrator can start changing the behavior of the APIs and user interface in real time for experimentation, including adding custom operations.

For production use, Cordra allows intensive extension and customization of the processes underlying the digital object server: how digital objects are stored and indexed, how they are validated and enriched, how users authenticate, when and to whom to give access to objects, and what custom operations can be performed. In production Cordra is run at

scale, supporting high reliability and performance; among other options Cordra supports MongoDB and Amazon S3 for storage, and Elasticsearch and Apache Solr for indexing. By definition of the underlying types and operations, Cordra is intended to serve directly as the API backend for a production application.

This talk will cover basic Cordra features as well as customization/configuration basics. Examples of current use will be shown, including the use of the Digital Object Interface Protocol (DOIP), for which Cordra serves as a reference implementation.

Current users of Cordra include the Derivatives Service Bureau (DSB), which uses Cordra as part of its backend to manage the automated generation of International Securities Identification Numbers (ISINs) for OTC derivatives in the financial services sector; and the British Standard Institute (BSI) whose Identify service for construction product manufacturers aims to assign a Universal Persistent Identification Number (UPIN) "for every product that is specified and incorporated in a building structure". The DSB, Cordra users since 2017, has a production system with over 80 million identified digital objects which receives millions of searches each month. BSI.Identify has a system where Cordra's DOIP interface is directly accessible as the service's public API.

Keywords

digital objects, digital object servers, repositories, infrastructure, handles, DOIP, PIDs

Presenting author

Robert Tupelo-Schneck

Presented at

First International Conference on FAIR Digital Objects, presentation