E-RPID: Enhanced Robust Persistent Identification of Data

DATA TO INSIGHT
CENTER
INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE

Yu Luo, Beth Plale
Indiana University Bloomington, USA
Rob Quick
Science Gateways Research Center, Indiana University Bloomington, USA

Problem

Open science is a worldwide activity with the mission of making research data more easily used for the benefit of science economies, and societies. A foundational first step towards making research data more usable is it having a unique and persistent identifier (PID).

Overview

Unique identification of the data products used in, or produced by, a scientific study:

1. strengthens science reproducibility,
2. makes it easier for a researcher to distinguish one dataset from its earlier version, and
3. makes citing of datasets within manuscripts and grant proposals consistent with citing of other papers.

A recent contribution of Open Science are the FAIR principles (for data to be more Findable, Accessible, Interoperable, Reusable) which bring clarity how data is made discoverable, accessible, interoperable, and re-usable.

The Enhanced Robust Persistent Identification of Data (E-RPID) project is a community testbed for experimenting with persistent IDs. A **Handle** is a widely used, globally unique persistent ID (PID). It is highly suited to identifying digital data objects. Handle are resolved by a global resolving system. DOIs are built on Handles.

The testbed, by running its own Local Handle Server (LHS), supports experimentation with PIDs. The PID Kernel Information activity of the Research Data Alliance is a user of E-RPID. PID kernel information allows a small amount of information to be stored with a PID record at the LHS for experimental purposes.

Architecture and Approach

The E-RPID project supports test services for managing Persistent Identifiers (PIDs). It integrates existing [1] services, some lightly modified, with new software needed for existing software products to function properly. The testbed services contribute to a FAIR data architecture, particularly the Digital Object Architecture framework [2,3].
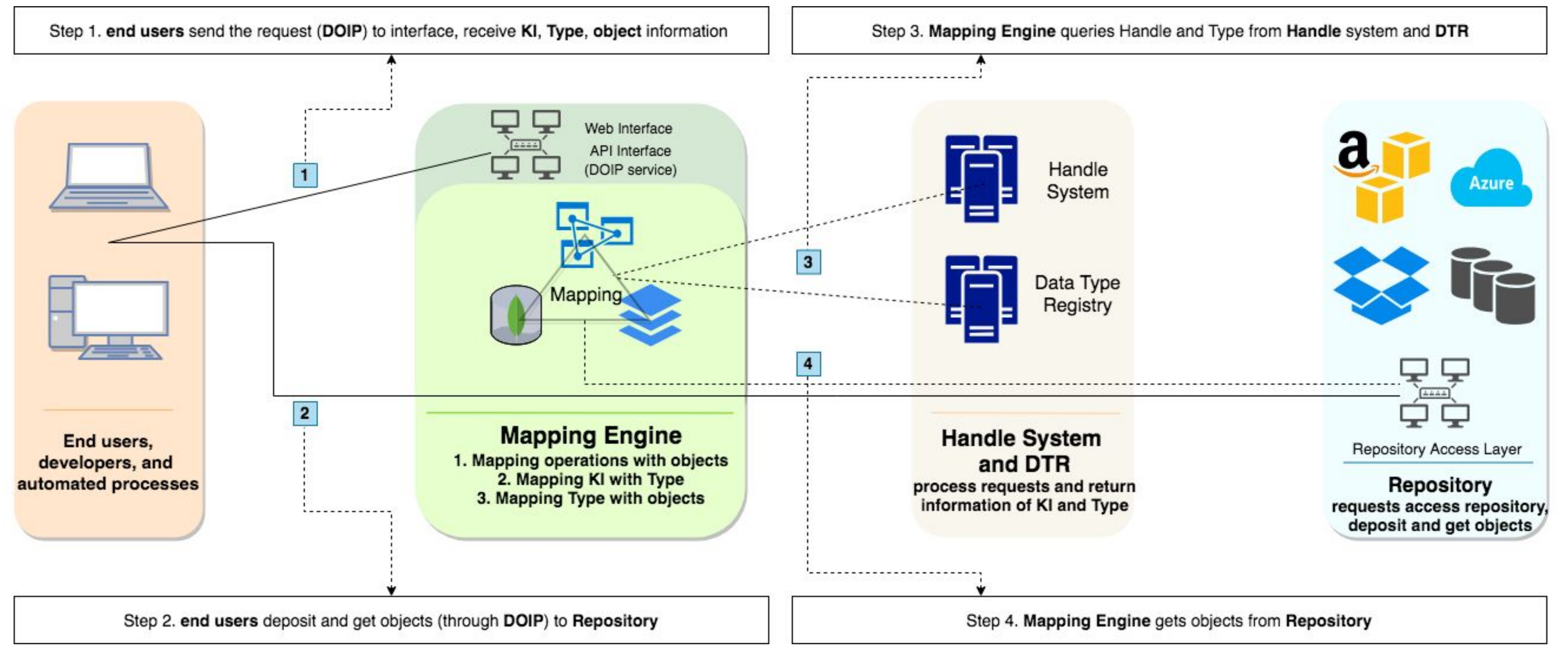
Users of the E-RPID testbed are experimenting with high-level data filtering, programmatic interoperability mechanisms, computational workflow provenance information recording, and the ability to assign long-term (at least within the testbed duration) identities to each of these aspects. The E-RPID testbed is provisioning new services in addition to the Handle service and Data Type Registry. These are

- Mapping and brokering tools enhanced processes to create a virtual object level over existing information structure, and
- An object interface protocol (Digital Object Interface Protocol [5]) that optimizes interaction with the object layer by understanding how to query each object for the set of operations that apply to it.

Acknowledgements

References

[1] Robust Persistent Identifier Testbed (RPID) Project, https://rpidproject.github.io/rpid, [Online Access, Mar. 08th, 2019]
[2] Digital Object Architecture, https://www.dona.net/digitalobjectarchitecture, [Online Access, Mar. 08th, 2019]
[3] Wittenberg et al. (2018) Digital Objects as Drivers, https://www.rd-alliance.org/sites/default/files/Digital_Objects_as_Drivers_towards_Convergence_in_Dat a.pdf
[4] RDA Recommendation on PID Kernel Information, DOI:10.5281/zenodo.1462543
[5] Digital Object Interface Protocol, https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf, [Online Access, Mar. 08th 2019]