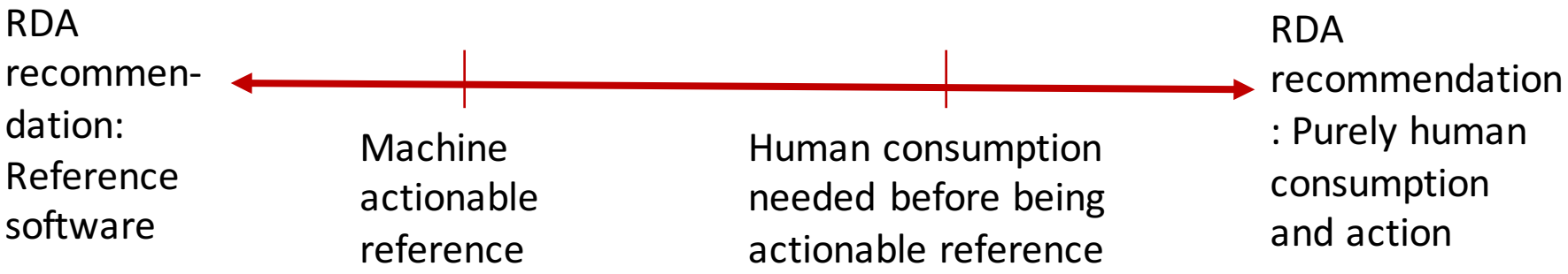
Data Fabric: building a testbed fabric inductively

Beth Plale
Indiana University

Objective

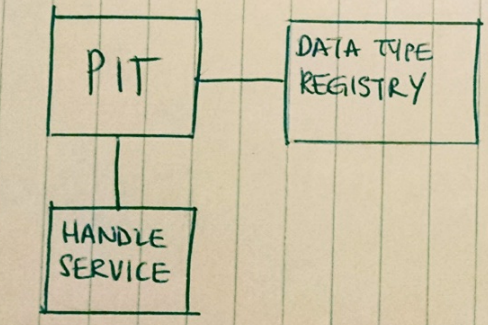
- Testbed composition: assemble a fabric inductively one RDA recommendation at a time (n=1, n=2 where n=fabric component)
- Why? Increase number of shared infrastructure practices in US that leverage RDA products (increased adoption and visibility of RDA in US)

Narrow definition of RDA recommendations that contribute to data fabric

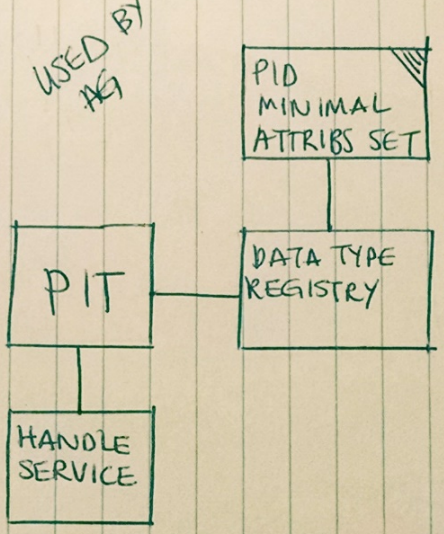


Build tiny fabric (start small)

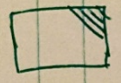
- Pulled components into single configuration for broad experimental use
- RDA PIT WG Recommendation and RDA Data Type Registry Recommendation are starting point because both are among few current RDA ouputs that have reference software

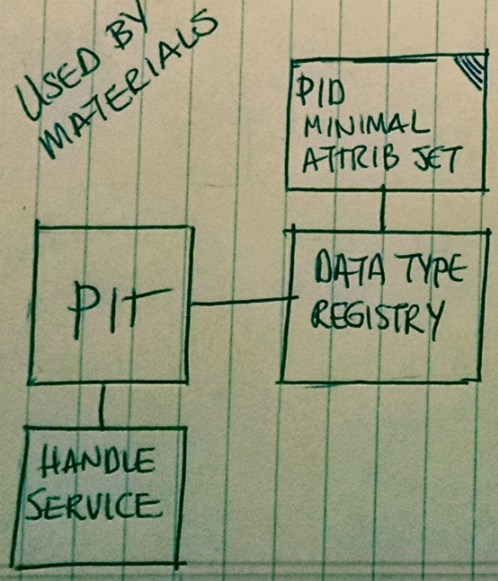


COMPOSITION (FABRIC) 1

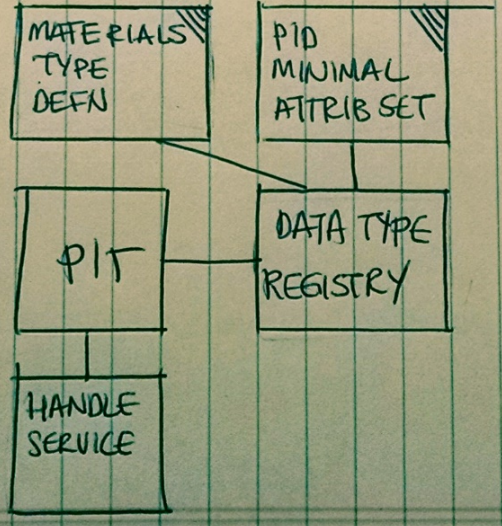


COMPOSITION 2

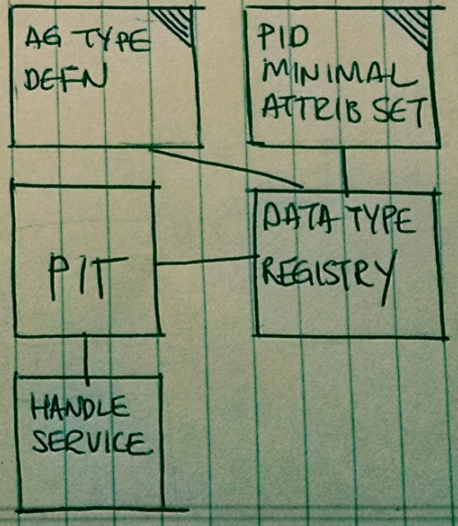
- COMPOSITION 1 IS NOT USEABLE AS IS
- COMPOSITION 2 ≡ COMPOSITION 3 SO SUBJECT TO "CORE" DESIGNATION
- COMPOSITION 4 AND COMPOSITION 5 START TO GET AT UNIQUE COMMUNITY INSTANCES OF A FABRIC
-  ; DOESN'T EXIST TODAY
- COMPOSITIONS 4 AND 5 HAVE $\Lambda = 5$



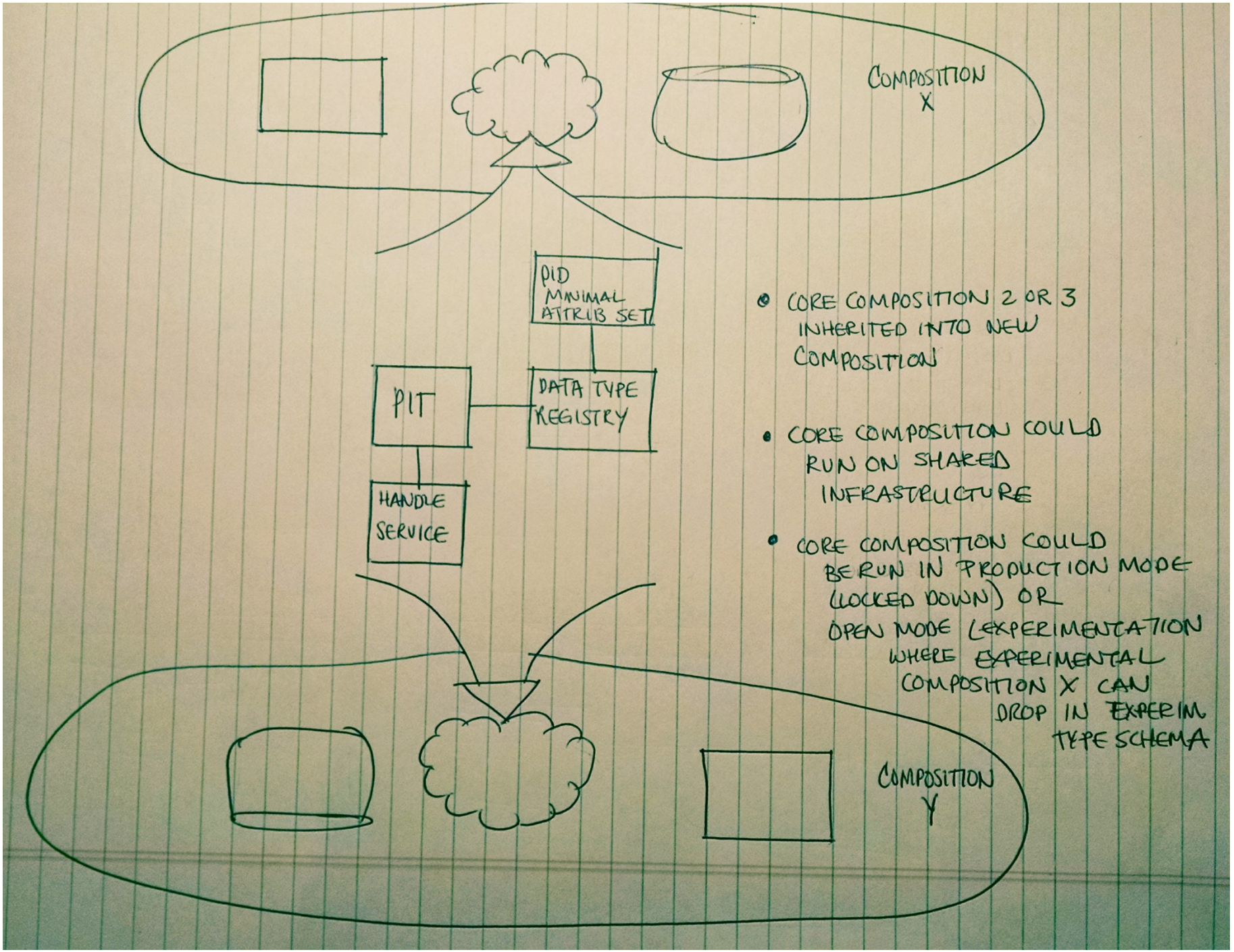
COMPOSITION 3



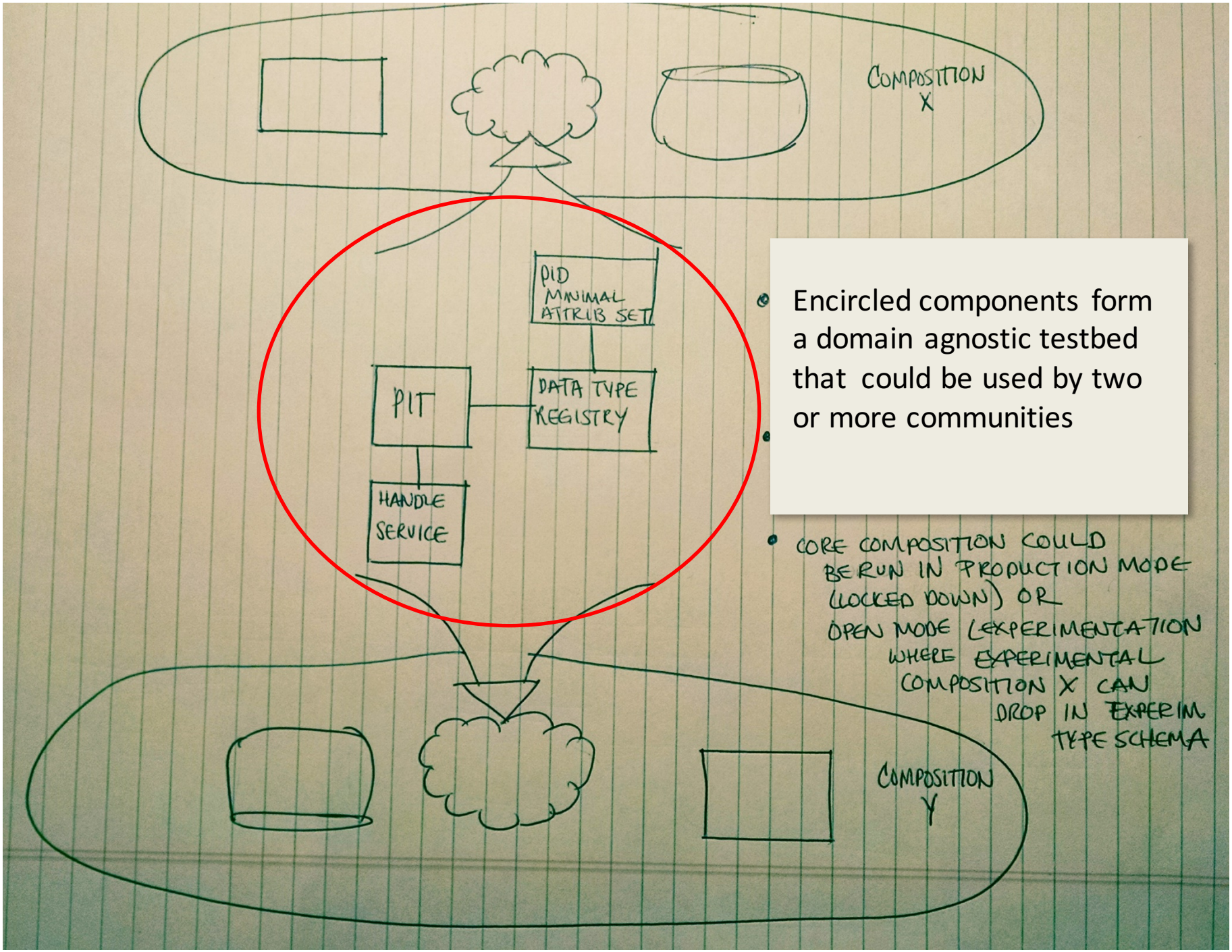
COMPOSITION 4



COMPOSITION 5

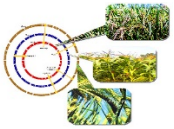


- CORE COMPOSITION 2 OR 3 INHERITED INTO NEW COMPOSITION
- CORE COMPOSITION COULD RUN ON SHARED INFRASTRUCTURE
- CORE COMPOSITION COULD BE RUN IN PRODUCTION MODE (LOCKED DOWN) OR OPEN MODE (EXPERIMENTATION WHERE EXPERIMENTAL COMPOSITION X CAN DROP IN EXPERIM. TYPE SCHEMA



- Encircled components form a domain agnostic testbed that could be used by two or more communities

- CORE COMPOSITION COULD BE RUN IN PRODUCTION MODE (LOCKED DOWN) OR OPEN MODE (EXPERIMENTATION WHERE EXPERIMENTAL COMPOSITION X CAN DROP IN EXPERIM TYPE SCHEMA



Bringing visibility to food security data results: harvests of PRAGMA and RDA

Beth Plale, Indiana Univ, USA; Jason Haga, AIST, Japan

- Launch use of two RDA products in Asia by utilizing PRAGMA community and tools to work with new rice genome group in Philippines and implement software services at AIST (Japan) using outputs of PID Information Types and Data Type Registries Working Groups
- Software will be installed additionally at National Data Service in US to stimulate US adoption

IRRI International Rice Research Institute

INDIANA UNIVERSITY

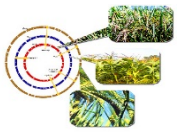
AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

Bringing visibility to food security data results: harvests of PRAGMA and RDA

Beth Plale, Indiana University, USA; Jason Haga, AIST, Japan

IRRI International Rice Research Institute

Rice genome variant discovery



PRAGMA/Rocks compute VMs

Persistent ID Types (PIT)
Data Type Registry (DTR)

PRAGMA Data Service

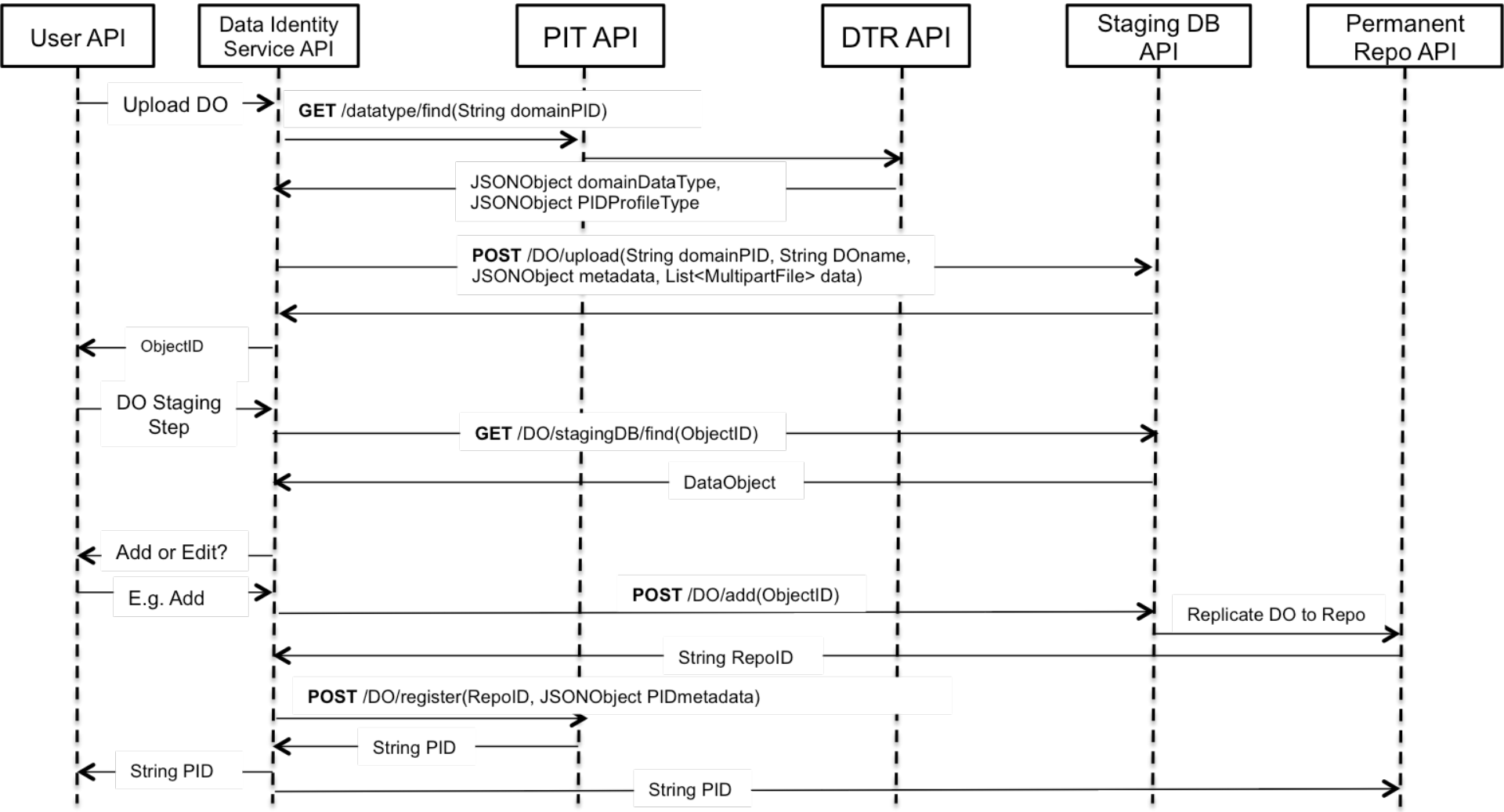
Future Uses

AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

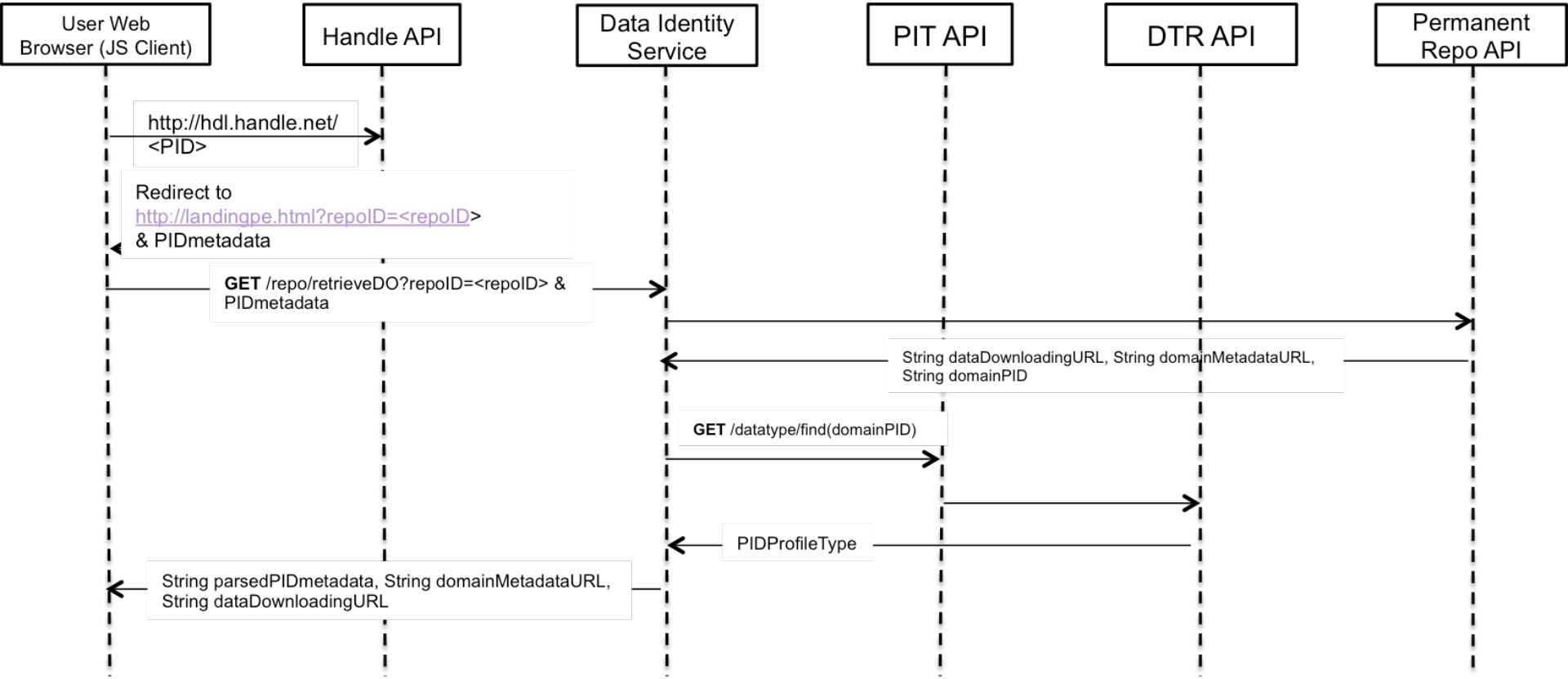
INDIANA UNIVERSITY

NATIONAL DATA SERVICES

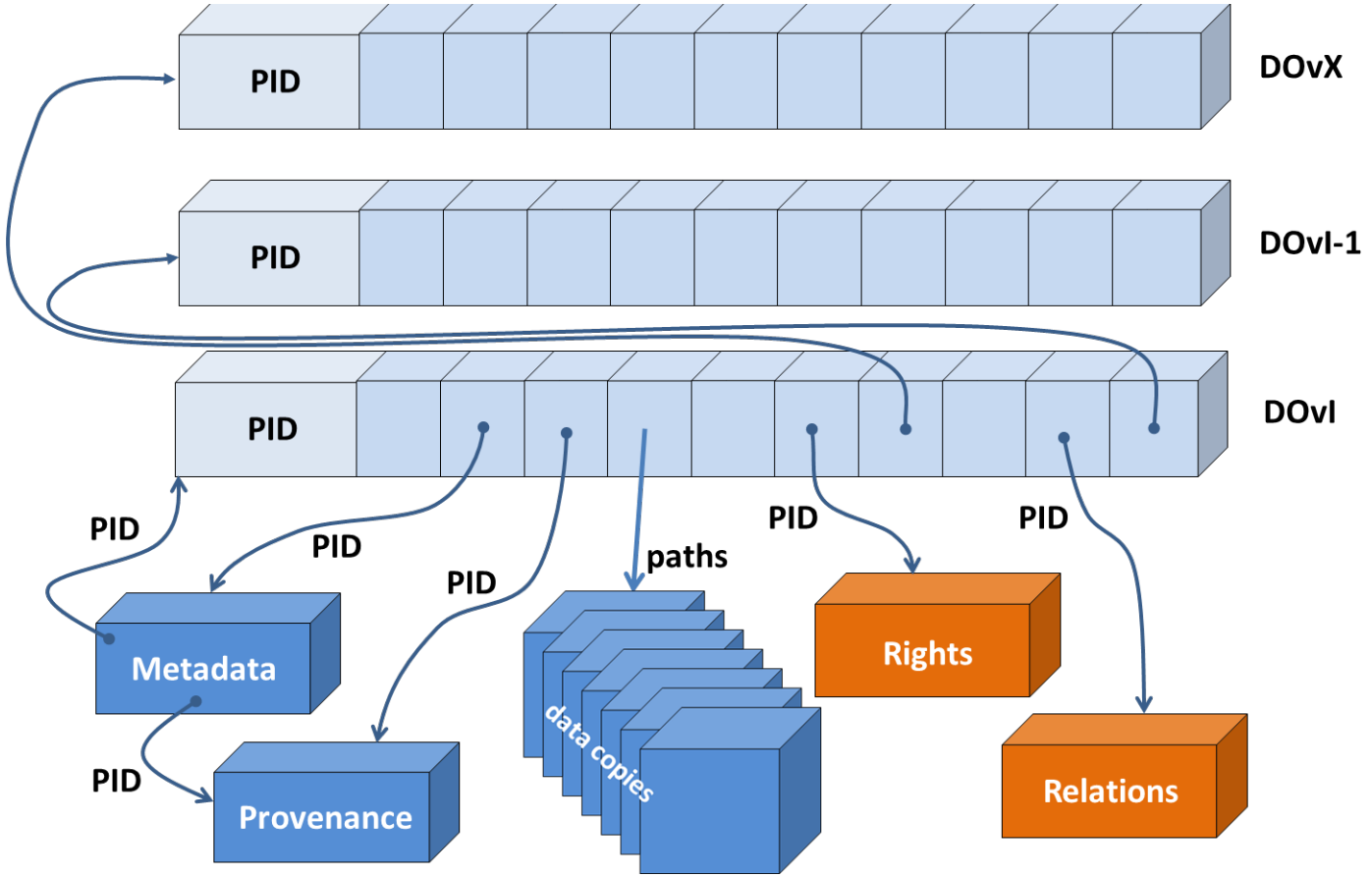
Component interaction: ingest



Component interaction: retrieval



PID minimal metadata



- Let's agree on types that are used to minimally define the metadata (attributes) associated with a PID

Next Steps

- Reflect feedback from Nottingham Jun 2016 meeting into prep for P8 session
- P8 Denver session:
  - Define minimal metadata for PID: what can be known about a record/dataset simply by resolving a PID
  - Discuss testbed
- Sep 2016: RDA Europa training: perspectives on PID