



Community Integration Toolkit

Sustainability and Evolution of the European Data
Infrastructure

Alison Kennedy
EPCC, University of Edinburgh
a.kennedy@epcc.ed.ac.uk



What is it?

- Community Integration Toolkit is a set of “tools” to help communities “join” EUDAT
- Called a toolkit as it contains all you need and is laid out in a way that makes it easy to find what you are looking for
- ...but communities can chose what tools to use for a particular job



What does it cover?

- Provides targeted information, advice, guidance and training, installation guides, service level agreements...
 - Whatever you tell us you'd like to be included
- ... to allow communities/centres to successfully accomplish the tasks necessary to create and configure an EUDAT “node”, customised to the local requirements
- ...and it documents the interfaces to which you can link your own PoP front end, community software and services

Background (1)

- CIT is one of the planned outputs from the Sustainability and Evolution workpackage
- One of a range of activities being undertaken to ensure the sustainability of EUDAT e.g.
 - Rolling plan – to signpost our plans and priorities for the next 3 years
 - Costing models – to ensure that the costs of sustaining the European Data Infrastructure are well-understood
 - Outreach activities – to ensure complementarity and co-operation with other infrastructures and projects working in similar or related areas

Background (2)

- CIT will bring together input from many different areas to present what-do-do and how-to-do-it information to potential EUDAT communities and users in a coherent and easy to understand, step-by-step way
- CIT work led by an experienced, customer-focused (former) operations director (Richard Blake from STFC) who understands service delivery and customer service requirements

Background (3)

- Our starting point for the CIT: sustainability depends on economies of scale, leverage and widespread uptake
 - adding users from existing communities, adding new communities and service providers (horizontal integration)
 - broadening the range of services offered, including services developed outside the project (vertical integration)
- The CIT needs to facilitate both types of integration

Rolling Plan Services Classification (1)

- **Core services:** services internal to the CDI essential for delivering the principal data management functionality of the whole infrastructure.
 - common across all centres that are “part of” the EUDAT CDI
 - Deploying and federating them are what it means to “join” EUDAT.
- **Basic services:** services which provide basic access to the EUDAT CDI for external users.
 - EUDAT provides a set of simple, stable interfaces to the CDI which can be built on to provide more complex advanced services in the future.
 - includes the simple upload and download of data and metadata.

Rolling Plan Services Classification (2)

- **Advanced services:** services which build on both core and basic services to provide more complex functionality.
 - can be extended to meet more and more complex use cases and requirements although as CDI services they should remain generic.
 - Does *not* include community-specific software tools but rather covers complex use cases for data in general.
 - Examples include data query, subsetting, workflow integration, computational hosting etc.

Where are we just now?

- The focus is on developing the core services to provide the essential EUDAT functionality of safe, policy-driven replication and registration and persistent identification of data.
 - Tight integration model
 - iRODS-based replication subsystem
 - PID registration and metadata harvest subsystems
- “Early adopter” users working very closely with data centres in “islands”
 - This is a suitable way to develop services but it’s not sustainable for roll-out (labour intensive, assumes too much local expertise)

Where do we want to be?

- Offering a loose integration model
- With a web service interface layer
- Offering ingest and retrieve services (where detailed knowledge of underlying technology is neither assumed nor needed)

CIT will support this transformation from a prototype to a production infrastructure.

Initial release Sept 2013, second release Sept 2014

What's the CIT view of EUDAT?

- The view of EUDAT that the CIT envisages:
 - EUDAT is a domain of registered data
 - EUDAT nodes have a common software stack
 - EUDAT's primary offering is safe replication and PID registration
 - EUDAT is policy driven but communities have the ability to customise these policies
 - Communities do not need to have expertise in particular technologies – EUDAT will present well-documented interfaces to underlying “technology sub-systems”

What should it cover?

- Covers “What?”, “Why?” and “How?”
- Installation Guides
- User Documentation
- Training (not technology specific)
- Software downloads/APIs
- Policy information
- Technical support information
- Testing plans etc.

No pain, no gain!



Our objective is to minimise the pain, while the technical workpackages maximise the gain.

Intention

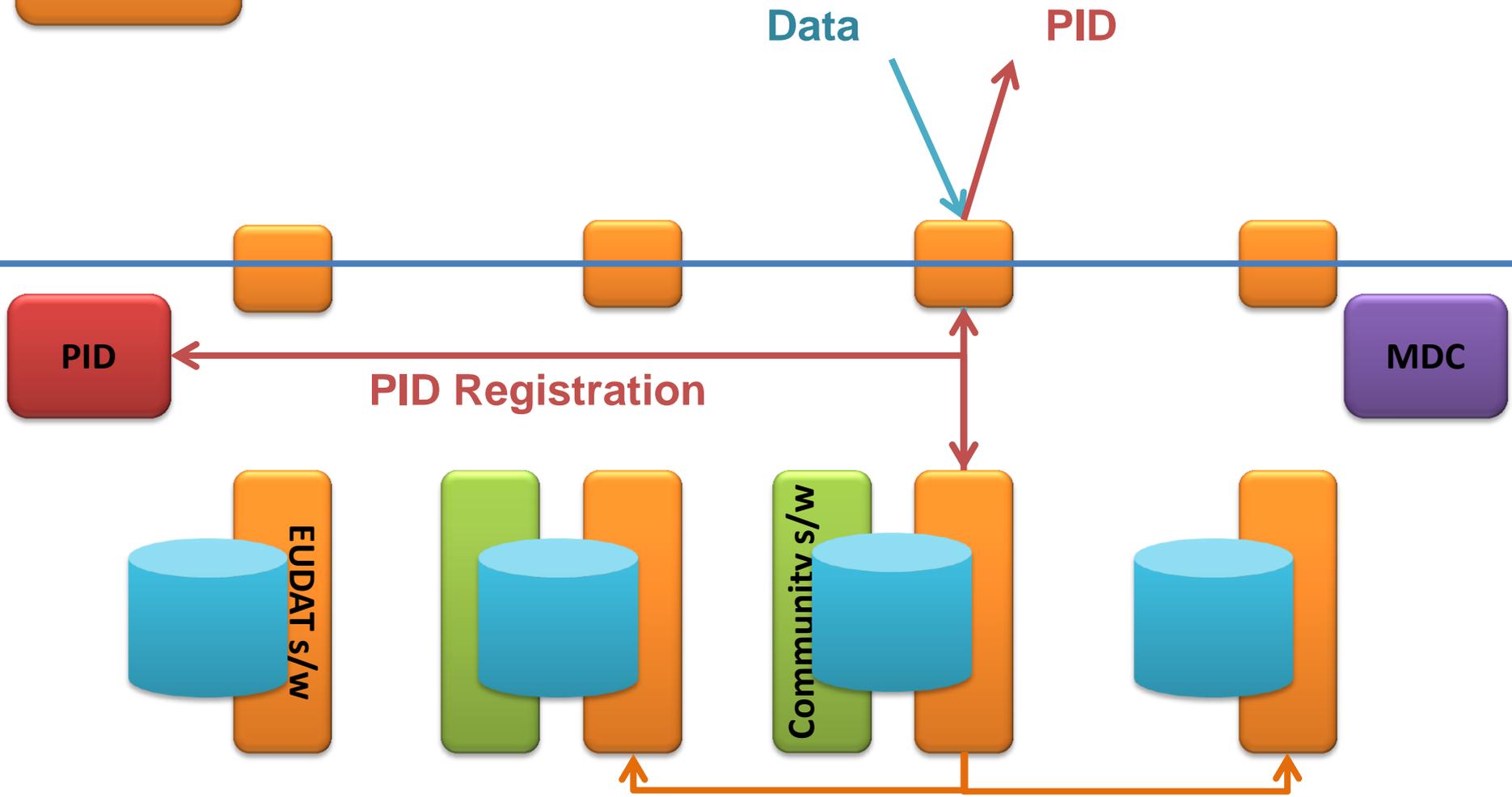
- Every EUDAT node will have the same software stack e.g.
 - Authorisation
 - Data Ingest (Data in, PID out)
 - Data Retrieval (PID in, data out)
 - Policy manager (setting and changing policies where these are configurable)
 - Query interface (links to metadata catalogue)

Simplified service development diagrams

- The following two slides depict simplified service diagrams (work in progress!)
- Will be expanded and documented in the Rolling Plan
- Will provide a roadmap for service development and CIT

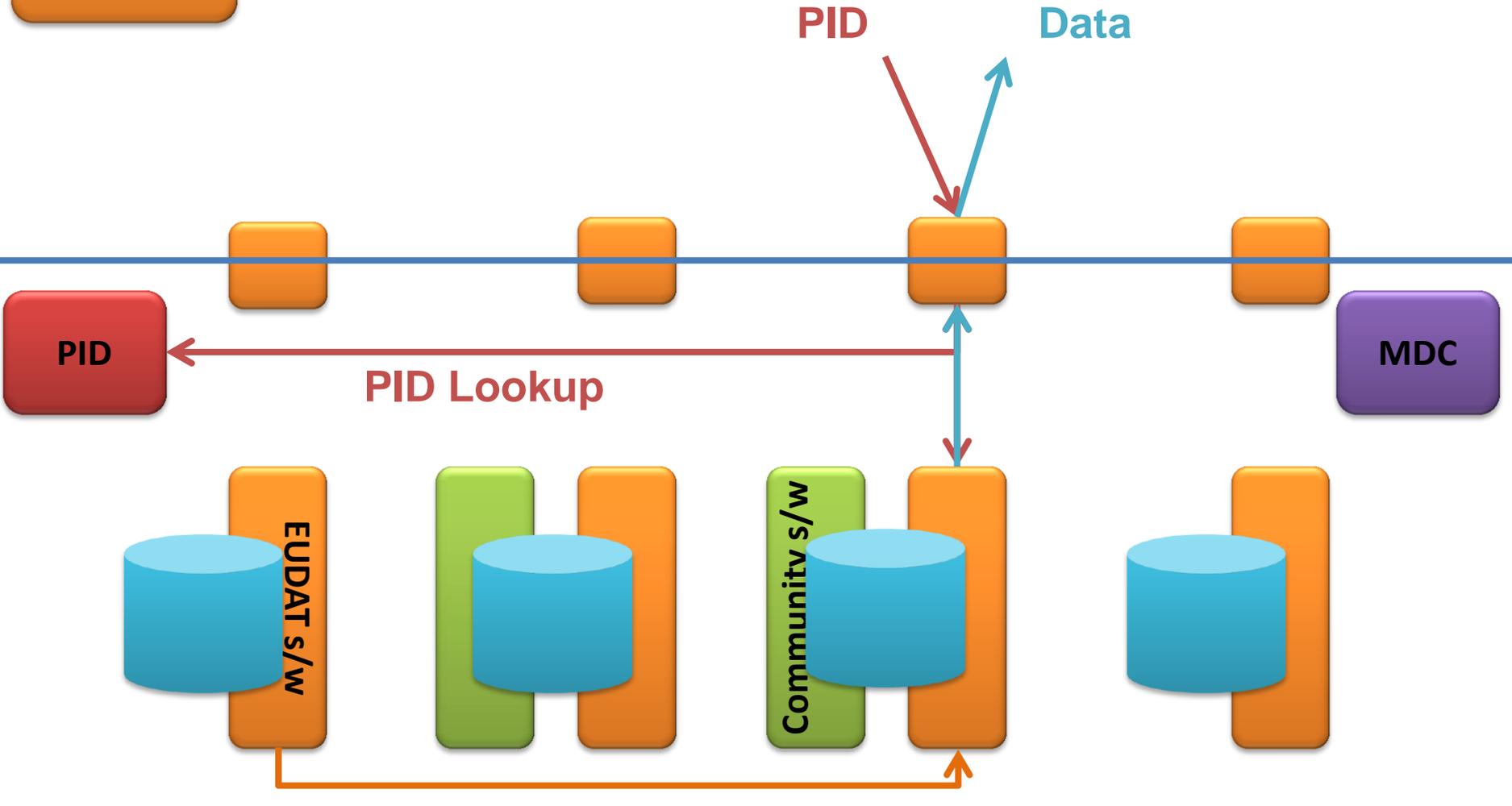


INGEST





RETRIEVE



Conclusion

- Community Integration Toolkit aims to simplify and document what it means to “join” EUDAT and how to do this
- It’s a project-wide initiative which brings together information on anything which helps centres or users to integrate and to use the services
- EUDAT infrastructure and core services are still under development, so there will be two versions of the CIT – Sept 2013 and Sept 2014
- Rolling Plan, publically available in Q2 2013, will signpost EUDAT’s next steps in more detail