The Key Role of Registries and Registry Standards in the Transition to a Federated Network of Repositories

Adrian Burton and Chris Blackall*

Australian Partnership for Sustainable Repositories (APSR)

- Registries and Registry Services
- Federated Networks of Repositories and Mass Data Storage Facilities

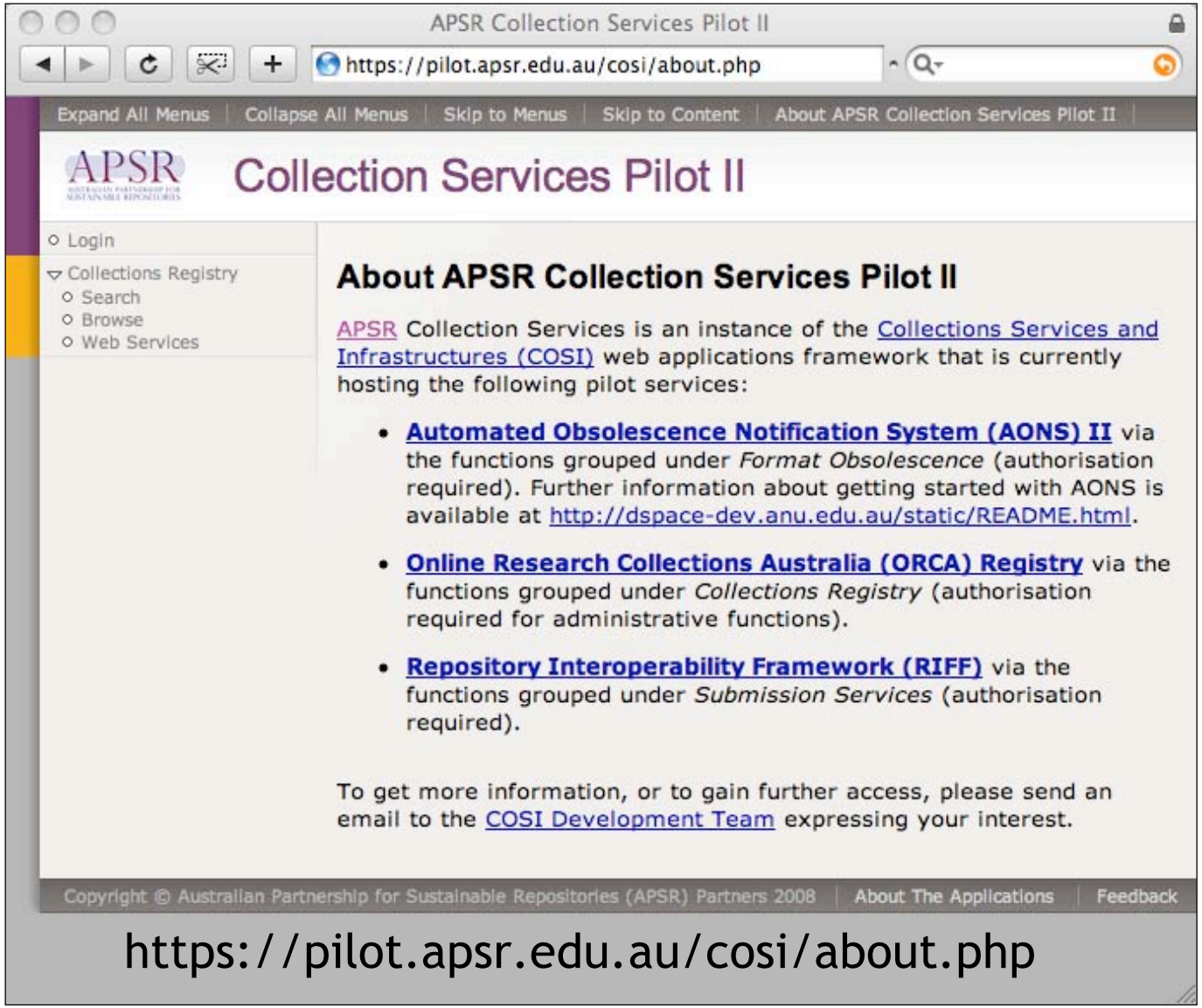
Presentation Outline

1. Live presentation of the ORCA (Online Research Collections Australia) Registry
2. Discussion of the role and benefits of registries in the context of *federated* networks of repositories and mass data storage facilities
3. Current progress in developing global approaches to implementing registries

ORCA Registry

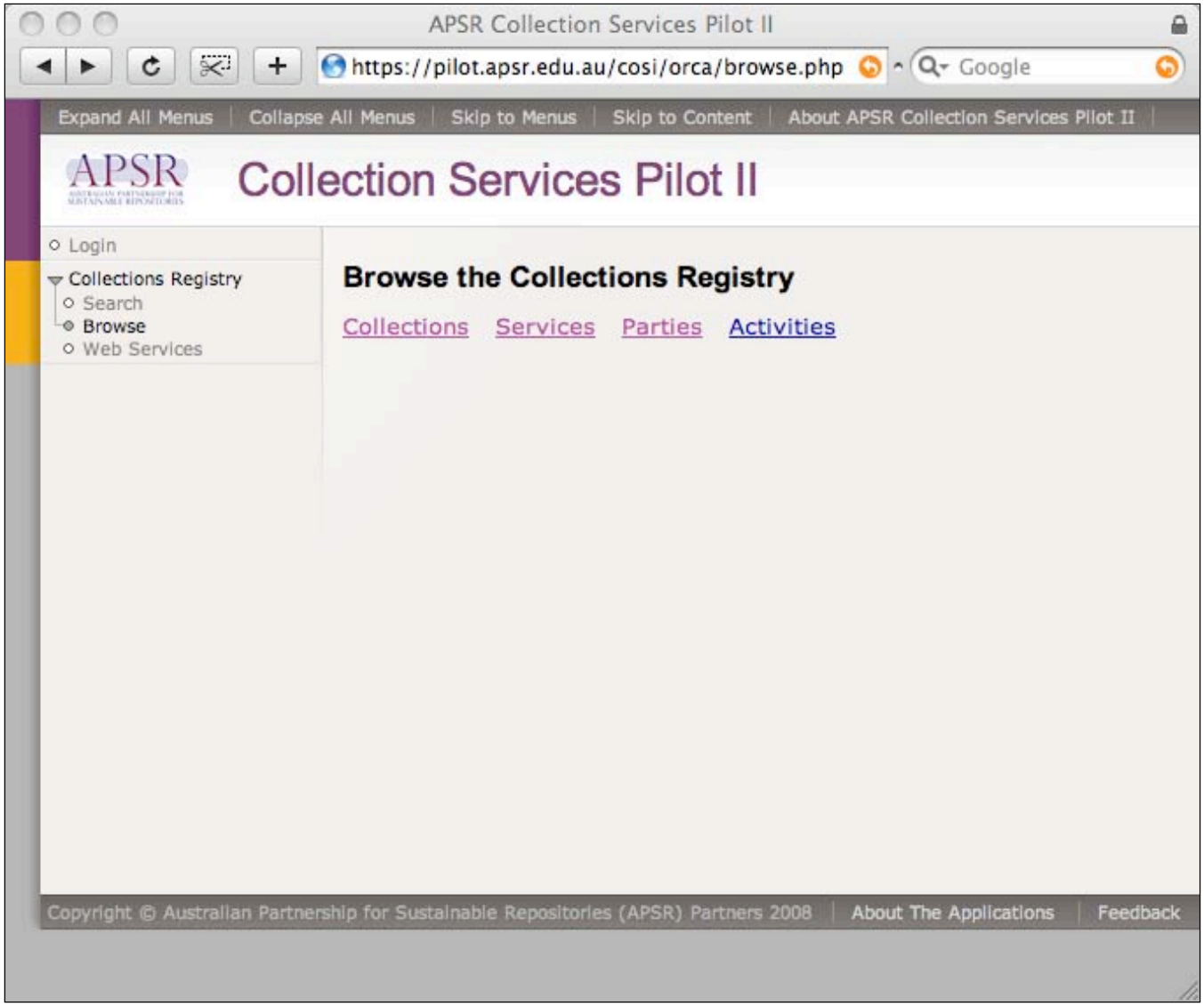
- It's a pilot not a production service
- It's not a discovery service (it's for machine-to-machine transactions)
- When people interact, its repository/collections managers and upstream metadata aggregators
- All metadata records are automatically harvested
- It harvests collection-level metadata records, not item-level records
- It provides no capacity for record owners to edit records (a feature not a bug)
- It's pointless unless part of a federated architecture of repository services
- It works on the principle of 'register locally, discover globally'

ORCA Registry Screen (I)



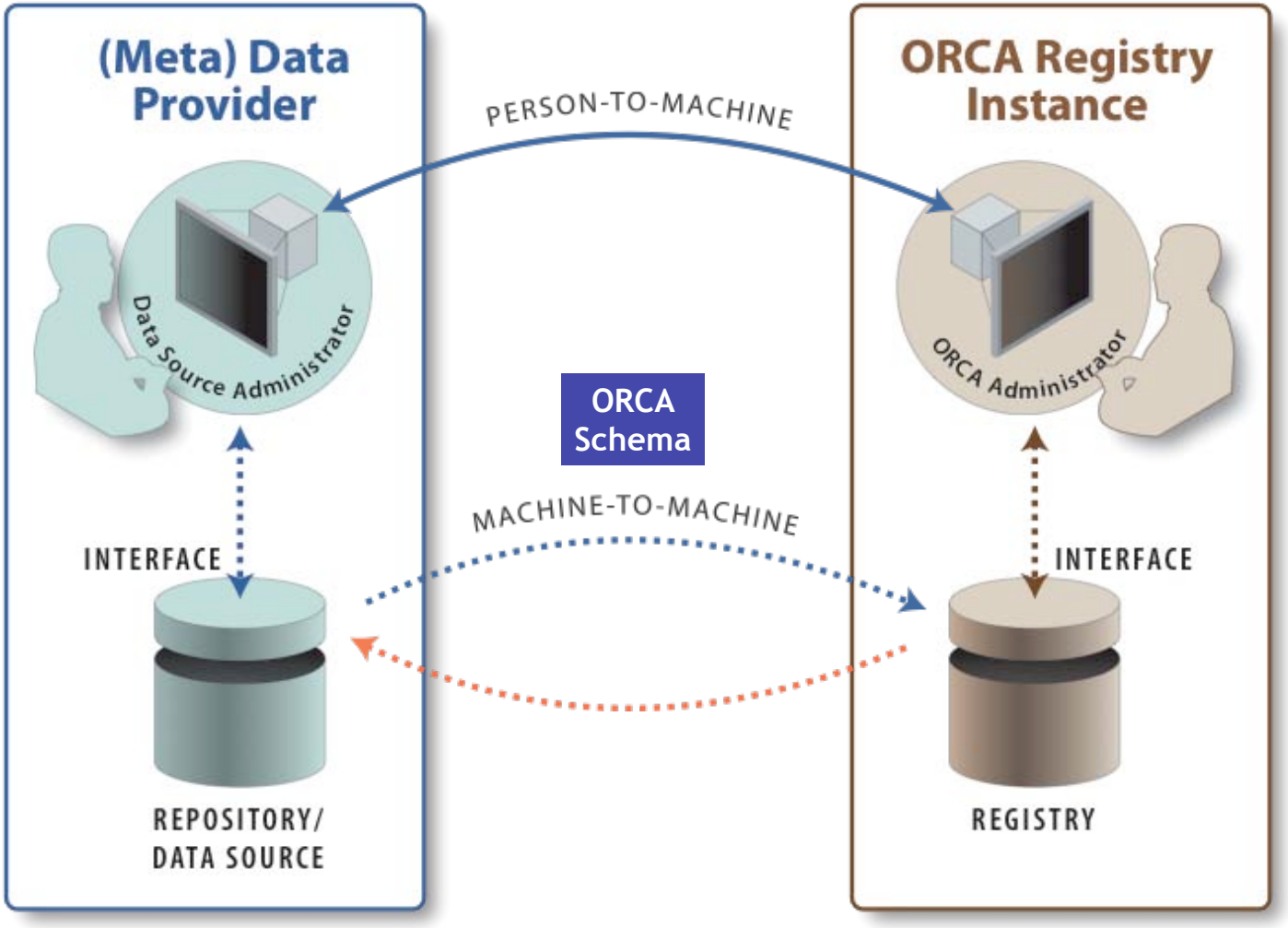
https://pilot.apsr.edu.au/cosi/about.php

ORCA Registry Screen (II)



ORCA Registry Screen (III)

Online Research Collections Australia (ORCA) Registry



For more ORCA info…

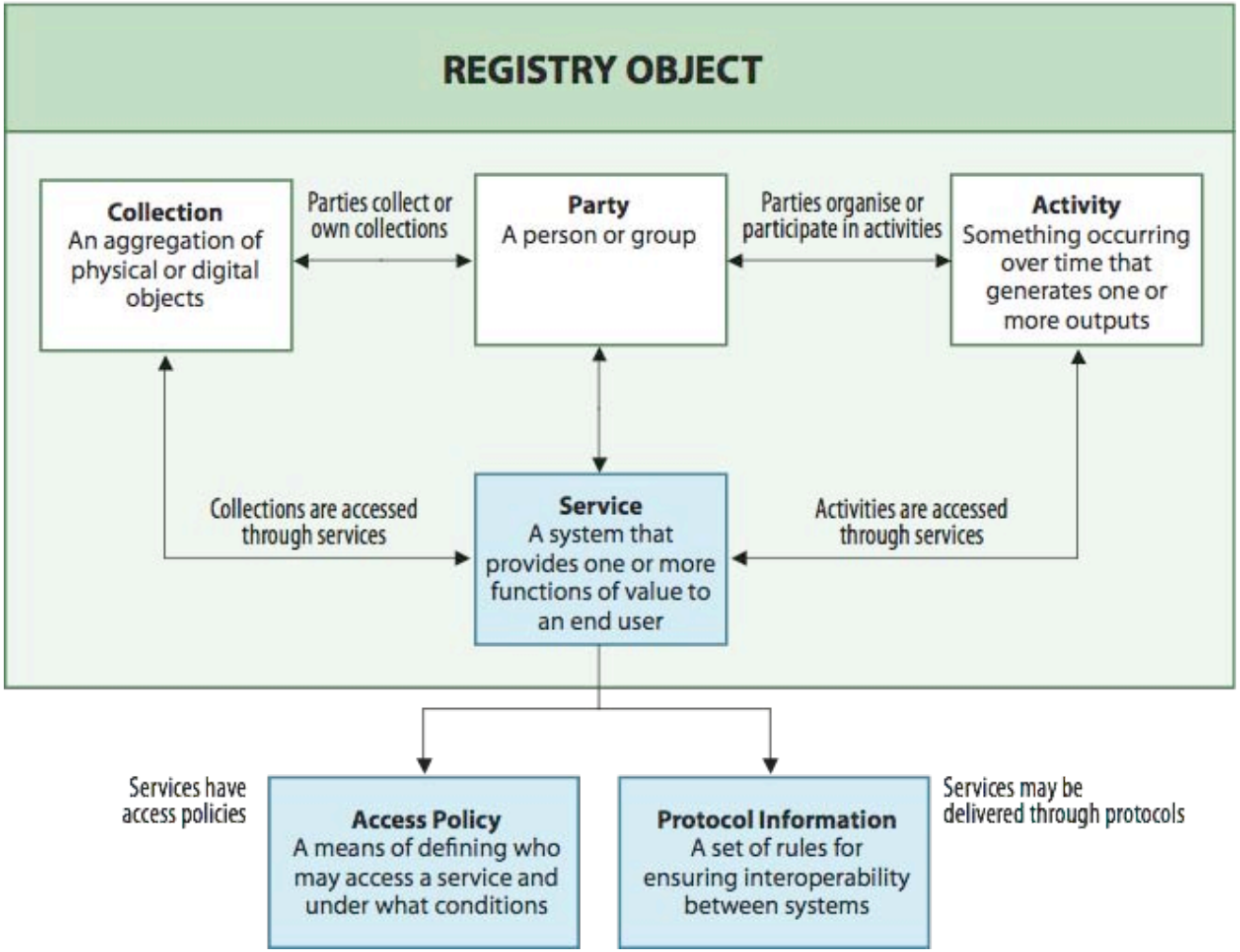
ORCA Registry technical details (I)

- PHP/PostgreSQL Web application (Apache lic.)
- Part of our own Web application framework (COSI-Collections and Services Interface)
  - Supports REST-type service calls and messaging (i.e. not WS-I standards and protocols)
- An ORCA Registry instance can configured for local and/or a national/global network topology
  - ORCA Registry instances have REST service interfaces for for metadata harvesting and providing

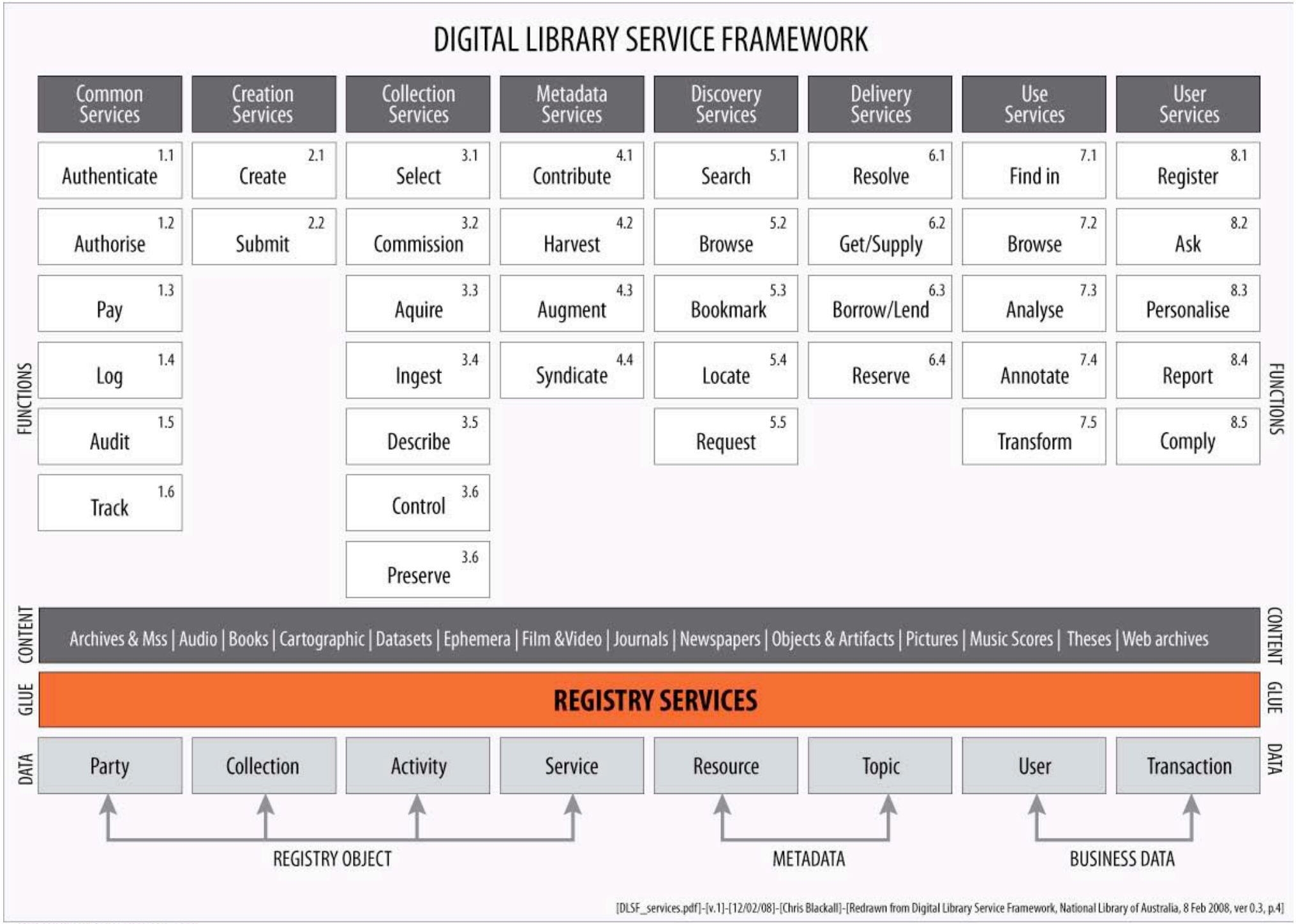
ORCA Registry technical details (II)

- First implementation of the draft ISO2146 (Registry Services for Libraries and Related Organizations)
- ORCA Registry uses ISO2146 XML binding for the basis of the DBMS data model and the metadata interchange profile (expressed as a .xsd schema)
  - Working on a ORCA Schema profile for OAI-PMH (June 2008)
- ISO2146 is compatible with the Dublin Core Collections Application Profile
- Interoperable with IESR and OCKHAM metadata
- Latest draft has input from APSR to align with current research information systems (i.e. CERIF compatibility)
- Credits: James Blanden, Scott Yeadon, Judith Pearce, Adrian Burton and Chris Blackall

ISO2146 Registry Services for Libraries and Related Organizations

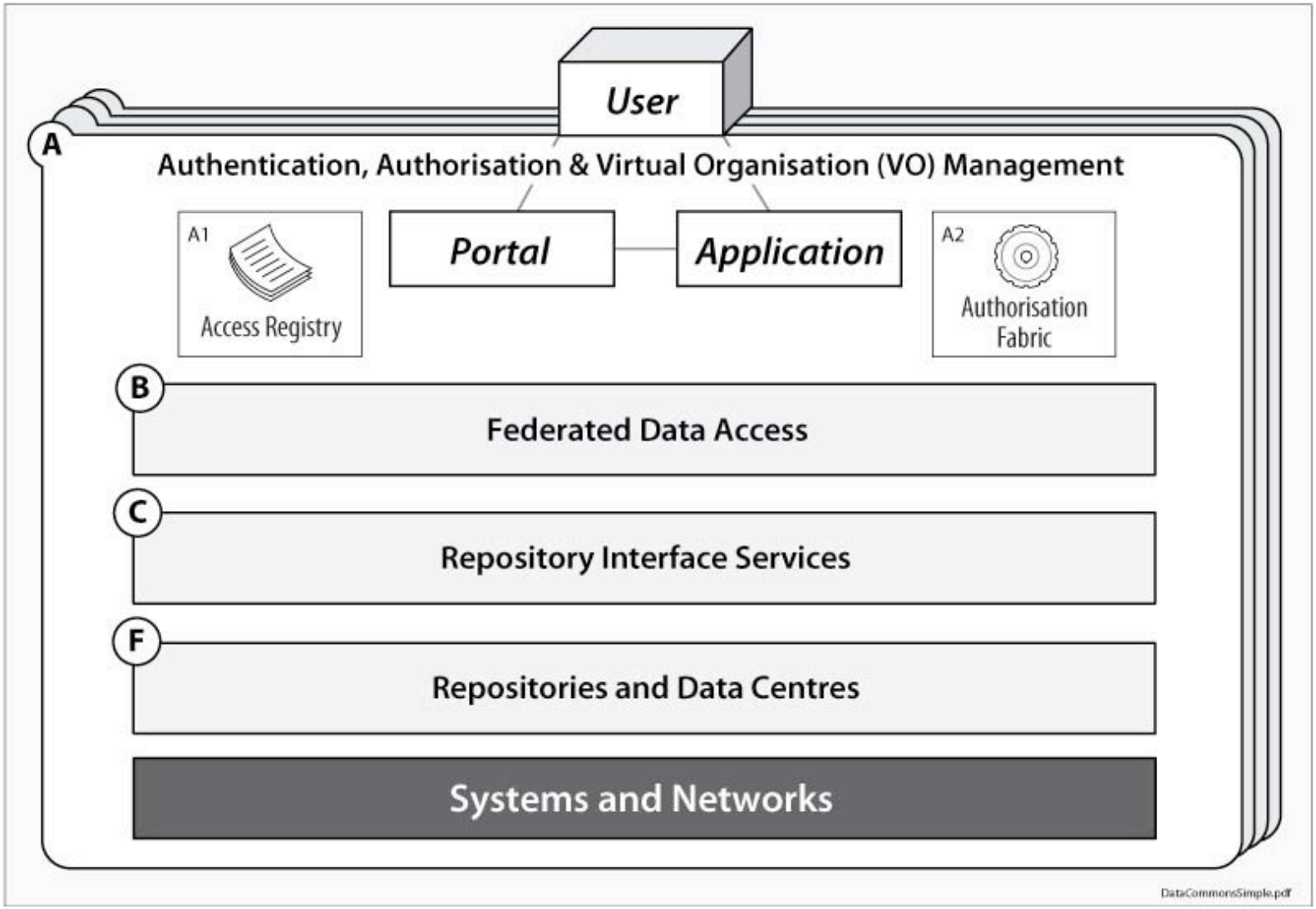


http://www.nla.gov.au/wgroups/ISO2146/

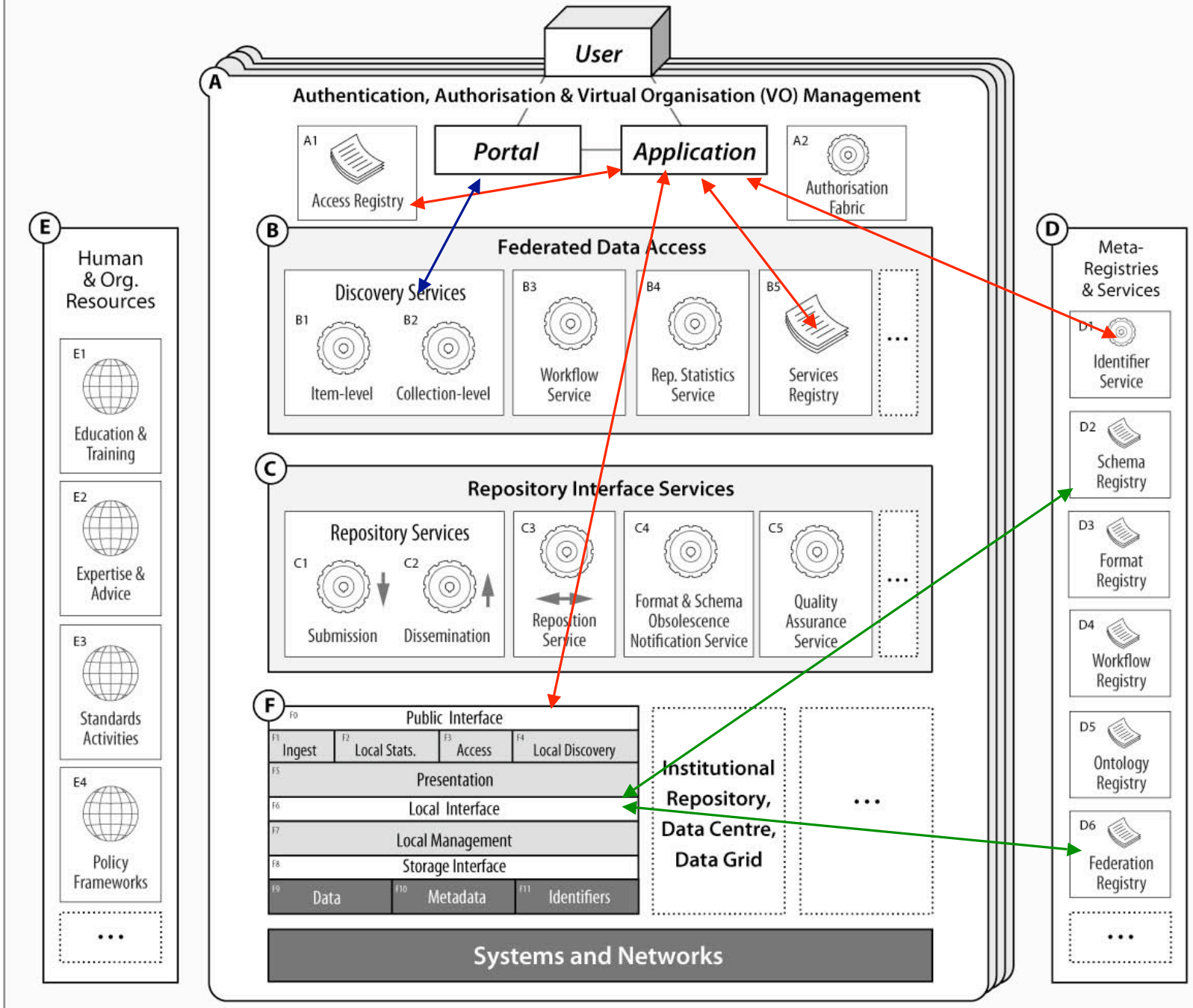


[DLSF_services.pdf]-[v.1]-[12/02/08]-[Chris Blackall]-[Redrawn from Digital Library Service Framework, National Library of Australia, 8 Feb 2008, ver 0.3, p.4]

Australian Data Commons Architecture (I)

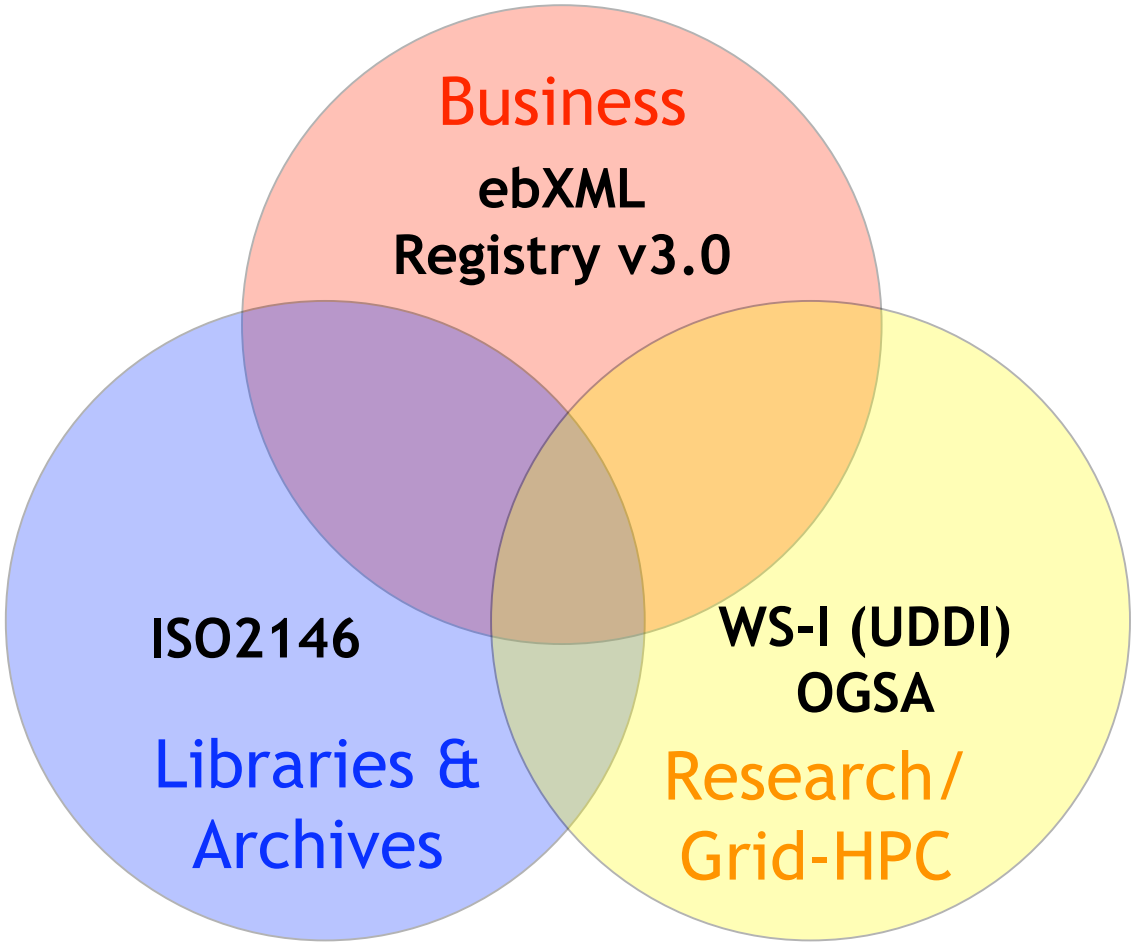


AUSTRALIAN DATA COMMONS ARCHITECTURE



[DataComServiceFwork.pdf]-[v.1]-[12/02/08]-[Design: Chris Blackall]-[Earlier version reproduced in 'Towards and Australian Data Commons' (2007); diagram produced by M. Bucchorn, A. Burton and C. Blackall]

Registry technology/standards options



WS-I (UDDI): Web Services-Interoperability (Universal Description, Discovery and Integration)
OGSA: Open Grid Services Architecture

Andy Powell (eduserv foundation)

Instead of developing *institutional* repositories, “we need to focus on building and/or using global scholarly social networks based on global repository services.”

Andy Powell, *Repositories thru the looking glass*, 2008. http://efoundations.typepad.com/efoundations/2008/02/repositories-th.html

Lorcan Dempsey, OCLC

"Registry services will need to get more common if we are going to have efficient interactions within networks of library providers and consumers"

Lorcan Dempsey. *Registries: the intelligence in the network*. August 20, 2006. http://orweblog.oclc.org/archives/001105.html

What are Global Repository Services? (I)

- Global repository services are **not** equivalent to networked repositories
  - IRs are typically implemented as enterprise applications that are focused on institutional needs rather that researcher needs
  - Low levels of interoperability between leading repository platforms
  - Service interfaces are available in IRs, but little in the way of shared services infrastructure
  - Commercial Web 2.0/Cloud Computing providers offering services

What are Global Repository Services? (II)

- Federated Networks of Repositories (and Mass Data Storage Facilities)
  - **Shared authentication and authorization services** (that enable users to log on to anywhere within the Federation and access Federation content)
  - **Shared discovery services** (that enable users to discover content unique to the Federation)
  - **Shared content creation and workflow services** (that enable users to create new content and content metadata that adds extra value to the Federation)
  - **Shared current research and statistics services** about research outputs, impact and quality (that enable users to make strategic decisions about current and future research)
  - **…**

Caveats

- In order the benefits to be achieved, we (repository developers) must develop systems that:
  - enable users of low to moderate IT skills to easily submit new content into repositories/data centers
  - reduce the time and effort required for users to capture and describe content metadata (preferably within the application that the content was originally created)
  - provide usage feedback services, such of usage statistics, impact measures and so on
  - Provide user-friendly metadata editing services to improve the quality of metadata records over time.

Federated Networks of Repositories

- "Towards an Australian Data Commons" (2007)
  - Australian Access Federation (AAF)
  - Australian National Data Service (ANDS)
  - Australian Research Collaboration Services (ARCS)
  - National Computing Infrastructure Services (NCIS)
- DRIVER (European Union)
- JISC Information Environment (UK)
  - Shared Infrastructure Services (SIS) Projects (2008)
  - Information Environment Services Registry (IESR)
- Datanet Projects (National Science Foundation-USA 2008)
- Global Registries Initiative (APSR, IESR, OCKHAM and others)

"The simpler the user-experience, the greater the scope, complexity and abstraction of the underlying information infrastructures"

ATM Machine
(3D Model)