iRODS

DataNet Federation Consortium & iRODS

Interoperability

Charles Schmitt
CTO & Director of Informatics
RENCI

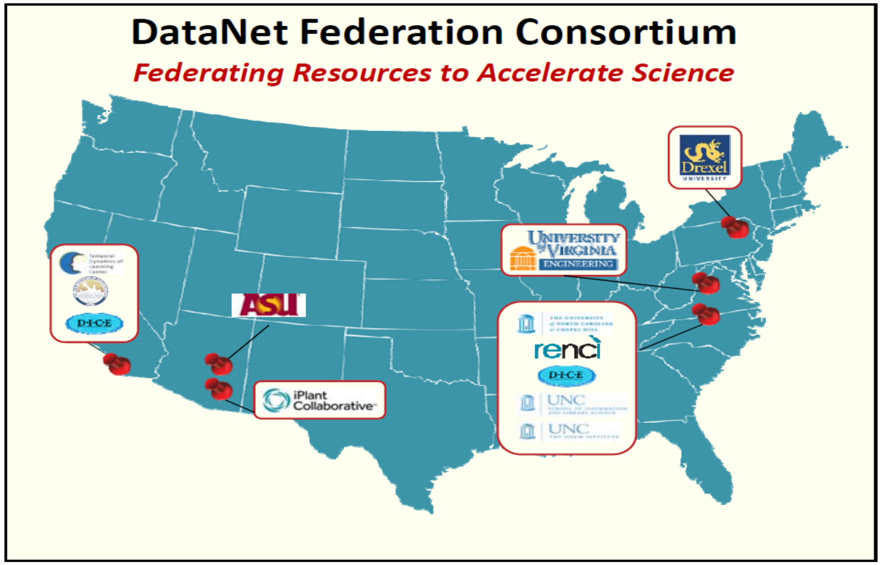
iRODS
CONSORTIUM

DataNet Federation Consortium (DFC)

NSF Sustainable Digital Data Preservation and Access Network Partners (DataNet)

Goals

- Enable scientific collaboration
- Support access to live research data
  - Data Sharing
- Support reproducible data-driven research
  - Workflow Sharing
- Establish national data cyber-infrastructure
  - Persistent and Extensible Architecture



DataNet Federation Consortium

Federating Resources to Accelerate Science

Temporal Dynamics of Learning Center
D·I·C·E
ASU
iPlant Collaborative
University of Virginia Engineering
The University of North Carolina at Chapel Hill
renci
D·I·C·E
UNC School of Information and Library Science
UNC The Odum Institute
Drexel University

renci

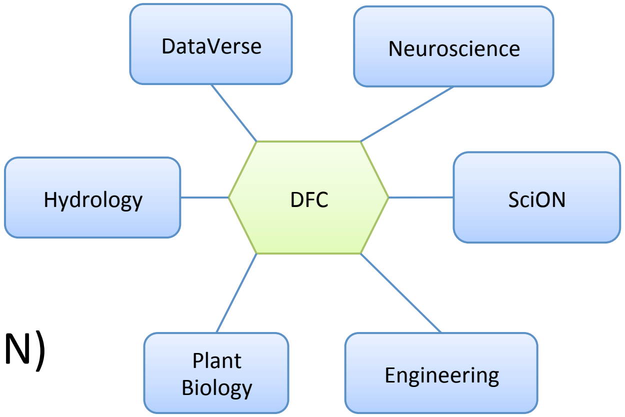
DFC DataNet FEDERATION CONSORTIUM

NSF

DFC Communities

Research Communities

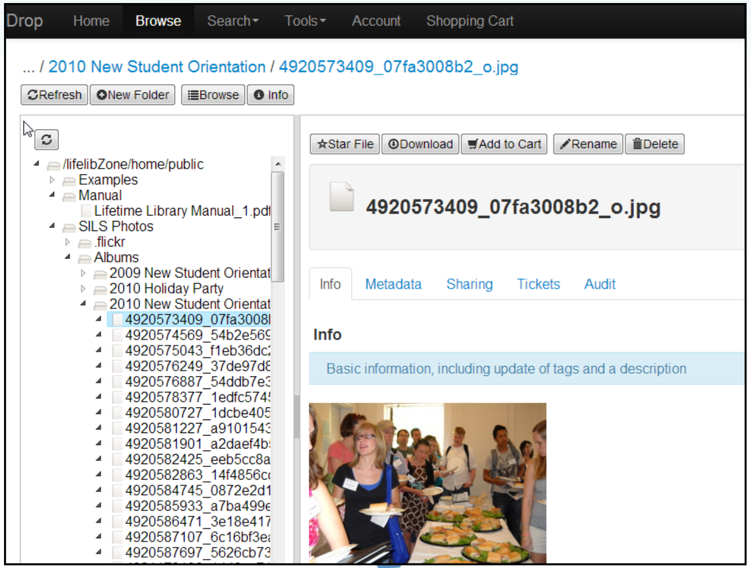
- Hydrology (UVA)
- Social Science (DataVerse)
- Neuroscience (TDLC)
- Ocean & Environment (SciON)
- Engineering (Drexel)
- Plant Biology (iPlant)



Communities of Practice

- Science and Engineering
- Technology and Research
- Facilities and Operations
- Policies and Standards
- Education and Outreach
- Sustainability

iRODS - Integrated Rules Oriented Data System



Drop Home Browse Search▾ Tools▾ Account Shopping Cart

... / 2010 New Student Orientation / 4920573409_07fa3008b2_o.jpg

Refresh New Folder Browse Info

- /lifelibZone/home/public
  - Examples
  - Manual
    - Lifetime Library Manual_1.pdf
  - SILS Photos
    - .flickr
    - Albums
      - 2009 New Student Orientat
      - 2010 Holiday Party
      - 2010 New Student Orientat
        - 4920573409_07fa3008l
        - 4920574569_54b2e56§
        - 4920575043_f1eb36dc
        - 4920576249_37de97d8
        - 4920576887_54ddb7e3
        - 4920578377_1edfc574
        - 4920580727_1dcbe405
        - 4920581227_a9101543
        - 4920581901_a2daef4b
        - 4920582425_eeb5cc8a
        - 4920582863_14f4856c
        - 4920584745_0872e2d1
        - 4920585933_a7ba499e
        - 4920586471_3e18e417
        - 4920587107_6c16bf3e
        - 4920587697_5626cb73

Star File Download Add to Cart Rename Delete

4920573409_07fa3008b2_o.jpg

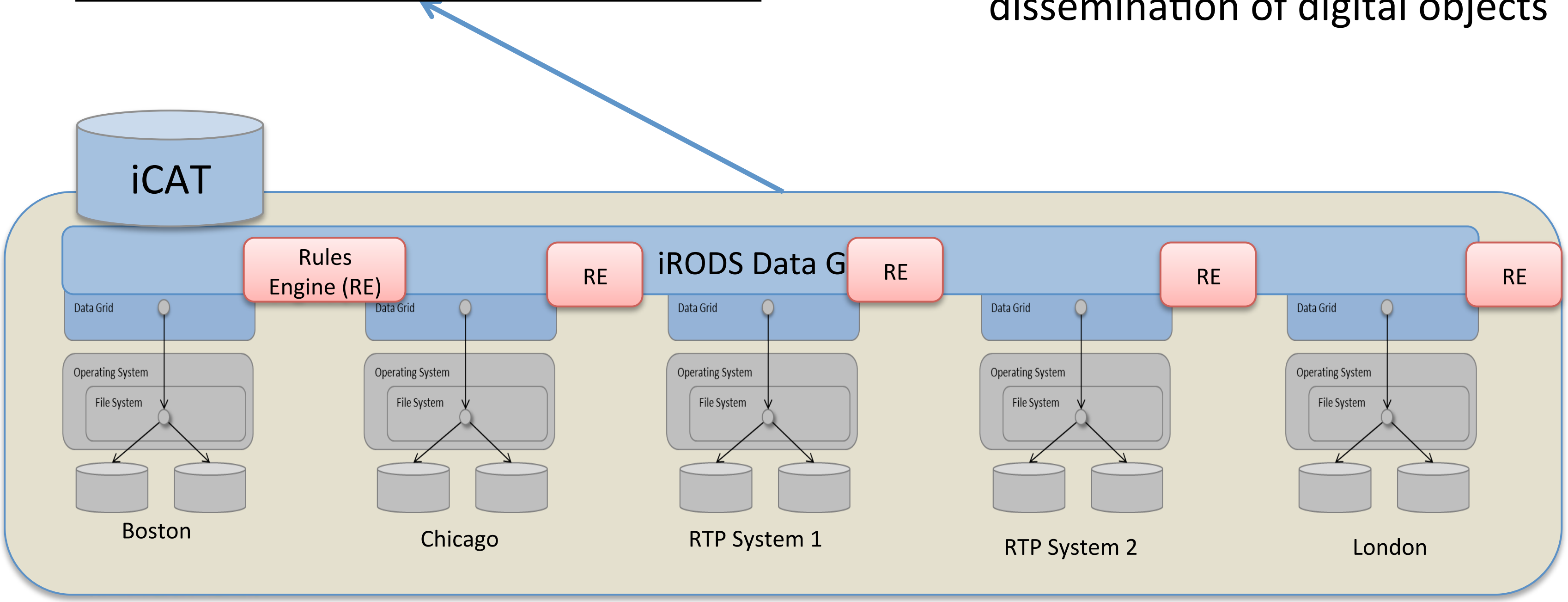
Info Metadata Sharing Tickets Audit

Info

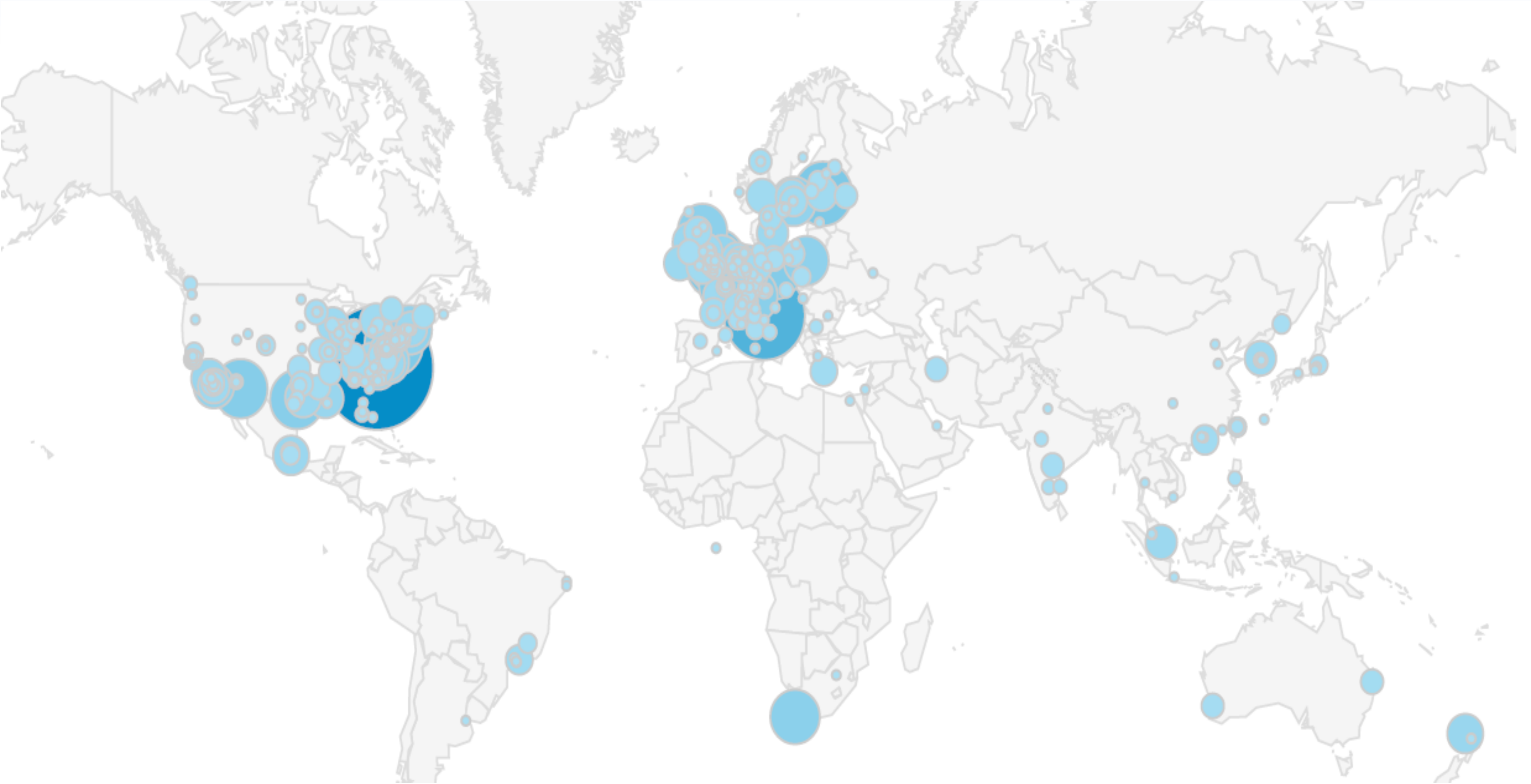
Basic information, including update of tags and a description

Policy-based Data Management Platform

- Virtual, Distributed Object Space
- Centralized Metadata
- Distributed Rules Engine allows:
  - Enforcement of policies
  - Automation of tasks
  - Orchestration of workflows
  - Controlled ingest, processing, and dissemination of digital objects



iRODS Download Base



iRODS 4.0 downloads: 04/01/14-10/16/2014

renci

iRODS EcoSystem

- Federal Users
  - National Aeronautics and Space Administration (NASA)
  - National Oceanic and Atmospheric Administration (NOAA)
  - National Optical Astronomy Observatory (NOAO)
  - US Geological Survey (USGS)
- Resellers/Redeployers/Partners
  - DataDirect Networks
  - Distributed Bio
  - Computer Sciences Corporation (CSC)
  - EMC
  - Seagate
- Commercial Users
  - DOW Chemical
  - Beijing Genome Institute
- Research Institutions
  - Broad Institute
  - International Neuroinformatics Coordinating Facilities (INCF)
  - Wellcome Trust Sanger Institute
  - Computer Center of the French National Institute of Nuclear and Particle Physics (CC-IN2P3)
  - CineGRID
- Hundreds of academic institutions worldwide
- Supported by the iRODS Consortium

DFC Models of Federation

Strong Federation

- Full and complete protocol-level federation across iRODS data grids
- Seamlessly Move from one grid to another
- “Mu Casa Su Casa”
- Used in DFC to federate Science and Engineering domain grids

Weak Federation

- DFC to External Resources
- Micro-services and Workflows
- DFC needs to ‘know’ the external protocol - plug-ins & wrappers
- Still seamless - external access problems hidden from user
- Used in DFC
  - To access THREDDS (netCDF), Sensor system, federal data resources
  - To connect to SEAD, TerraPop, and DataONE
  - To access Amazon Web Services (EC2, S3) and External workflows

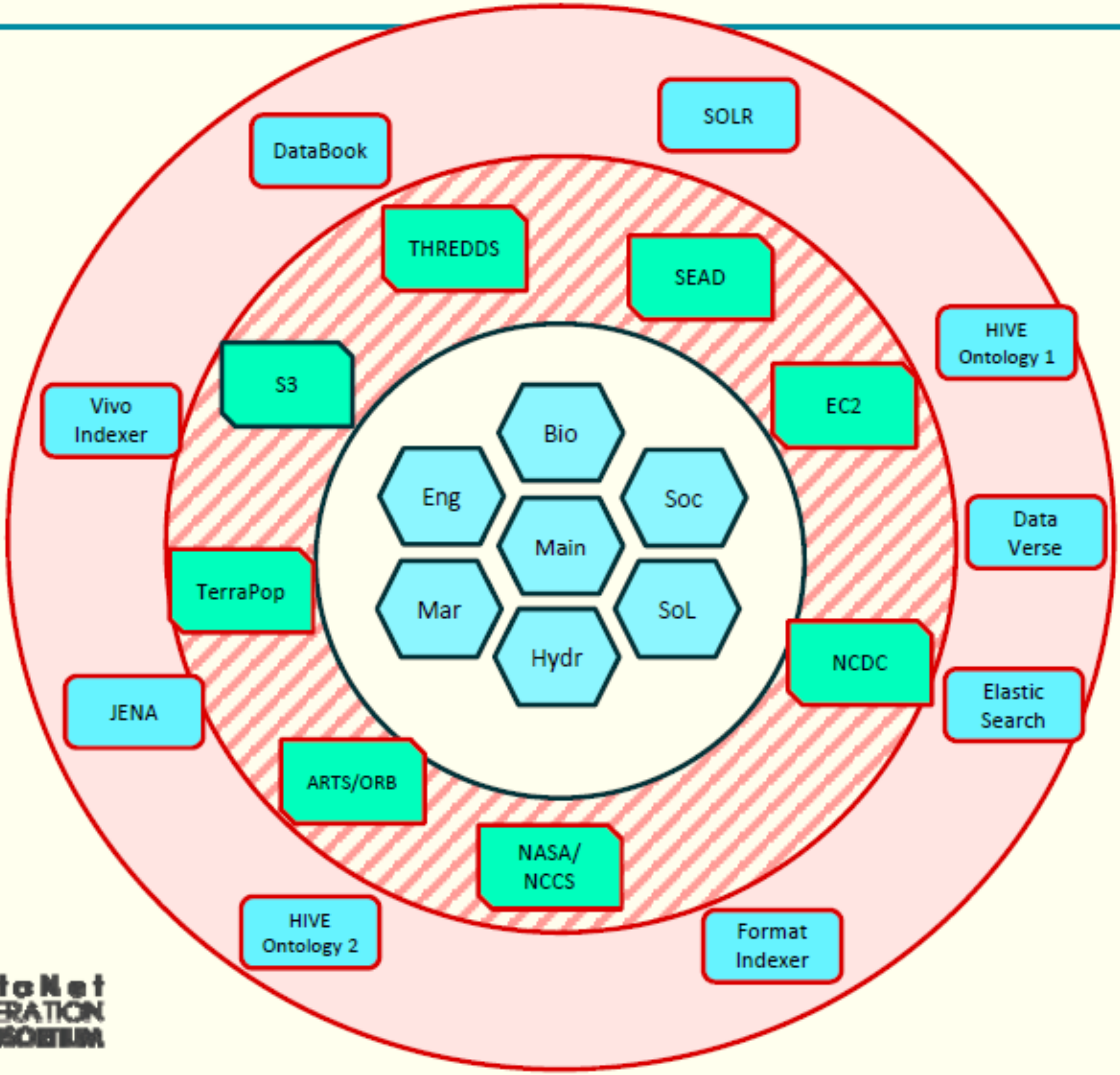
renci

DFC Models of Federation

Asynchronous Federation

- Defined & Developed to meet needs of DFC
- Motivation Example: Provision access to indexing services
- Multiple indexing services – each with its own protocols
- DFC does not want to federate each separately – access is not seamless
  - Promote a common connectivity - based on message bus
  - Any indexer who can ‘talk’ this common connectivity can play
- Other examples: Metadata services, Ontologies, Formatting Services

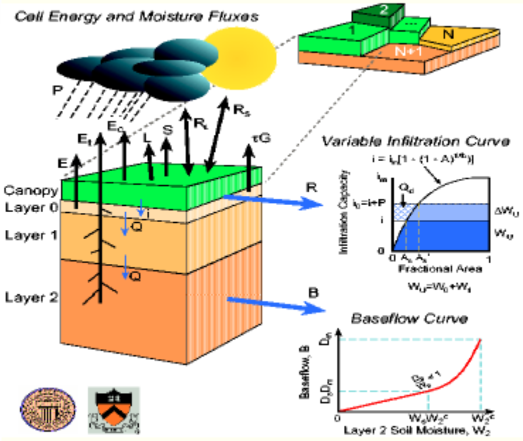
Three Federations in DFC



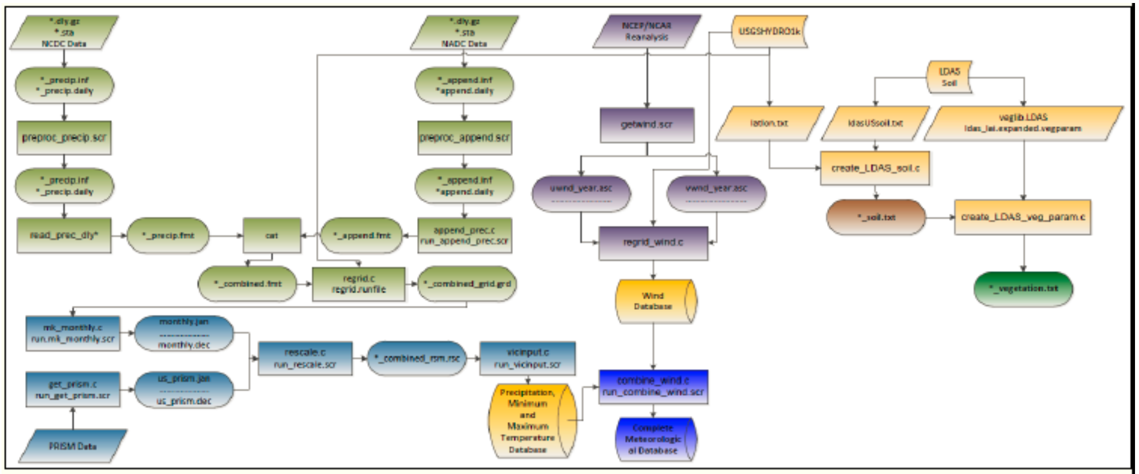
DFC DataNet FEDERATION CONSORTIUM

NSF

Interoperability Example



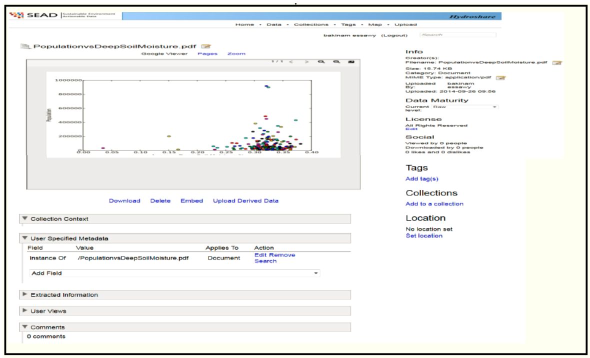
Source : Gao et al. (2009)



Source : Billah et al. (2014)

DFC Hydrology workflow (UVA)

- Input data from TerraPop data systems (weak federation)
- Data processing across differenet iRODS data systems (strong federation)
- Output data to SEAD hydrology system (weak federation)
- Data generation/usage events can be monitored through DataBook (asynch federation)



PopulationvsDeepSoilMoisture.pdf

Collection Context

User Specified Metadata

Field	Value	Applies To	Action
Instance Of	/PopulationvsDeepSoilMoisture.pdf	Document	Edit Remove Search

Extracted Information

User Views

Comments

0 comments

Info

Creator(s):
Filename: PopulationvsDeepSoilMoisture.pdf
Size: 15.74 KB
Category: Document
MIME Type: application/pdf
Uploaded By: bakinam essawy
Uploaded: 2014-09-26 09:56

Data Maturity

Current level: Raw

License

All Rights Reserved

Social

Viewed by 0 people
Downloaded by 0 people
0 likes and 0 dislikes

Tags

Collections

Location

No location set

SEAD Hydroshare Project Space

Data Systems

- Purpose
  - Reason a collection is assembled
- Properties
  - Attributes needed to ensure the purpose
- Policies
  - Controls for enforcing desired properties
- Procedures
  - Functions that implement the policies
- Persistent state information
  - Results of applying procedures
- Property verification
  - Validation that state information conforms to purpose

- *These change throughout the lifecycle of the collection*

- *These are defined by the user/community of interest*

Multiple aspects of interoperability in Data Systems

- Protocols & Interfaces
  - Services, APIs
- Authentication & Authorization
- Data
  - Discovery, Link, Retrieval, Submission, Manipulation, Syntax, Semantics
- Metadata
  - Discovery, Link, Retrieval, Submission, Manipulation, Syntax, Semantics
- Workflows and Computation
  - Automated control and data flows between systems
- Policies
  - Usage rights, audits, embargoes, verifications, encryption, FISMA, leakage protection