Catalyzing an NDS ecosystem

Ian Foster
Director, Computation Institute
Argonne National Laboratory
& The University of Chicago
@ianfoster
ianfoster.org
We aim to reduce sources of friction that currently impede data-driven science ...

... without creating new ones
Let’s make a few simple things easy

For example:

• “Dr. U shares data with colleagues”
• “Project V organizes data across partners”
• “Institution W sets up an institutional repository”
• “Dr. X searches across a dozen databases”
• “Agency Y creates a secure data commons”
• “Publisher Z links papers to research repositories”
Three elements of an NDS ecosystem

Provide basic services for data sharing, publication, discovery, reuse; authentication, authorization, profile, group management; cross-domain search

Vanilla services

Pluggable framework

Foundational infrastructure

Deploy initial infrastructure to allow communities to jump-start disciplinary services: Scratch storage, repository storage, cloud computing services for hosting portals and services

Identify interfaces and conventions to allow disciplinary extensions, replacements, and contributions to NDS infrastructure and services
A generic data solution based on rough consensus and running code

National Data Service Portal

Collection creation & sharing
- Create collections
- Share privately
- Extract/assign metadata
- Assign DOI
- Find appropriate repository

Cross-disciplinary search service

Identity & group management

Data movement & access

Data coming from...
- Lab experiment
- Simulations
- Other computations

Data coming from...

Repository

Scholarly Journal links to data

NationalDataService.org
Organized as an extensible framework

National Data Service Portal

Collection creation & sharing
- Create collections
- Share privately
  - Extract/assign metadata
  - Assign DOI
  - Find appropriate repository

Cross-disciplinary search service

Project repository
- Curates standard products
- Enforces embargo policies
- Subscription, delivery

Data discovery service
- Custom schema
- Semantic search
- Fine-grain access control

Identity & group management

Data movement & access

Lab experiment
- Lab notebook tool
- Access service
- Mass data import tool

Simulations
- Computations
- Archive data

Community repository

Repository

Scholarly Journal links to data

NationalDataService.org
We want your help

Use cases
Infrastructure
Communities
Services