

Mercury: Metadata Generation, Search and Data Discovery System

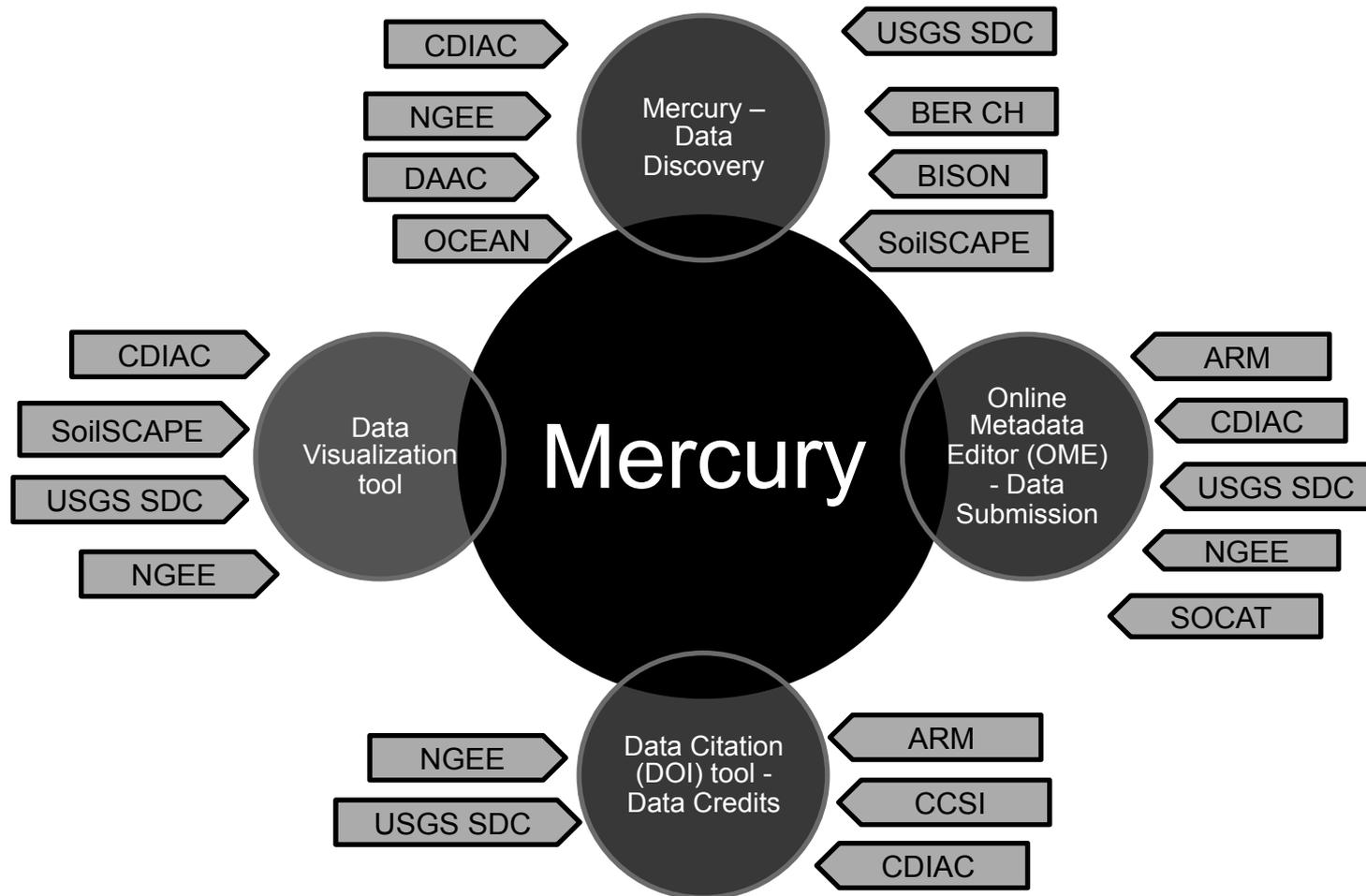
Ranjeet Devarakonda
R&D Staff and Systems Architect
Environmental Sciences Division
Oak Ridge National Laboratory
devarakondar@ornl.gov



Mercury: Metadata Generation, Search and Data Discovery System

- Mercury is a combination of three main tools
 - Data/Metadata registration Tool
 - Search and Discovery Tool
 - Data Citation Tool
- In existence since early 2000
- Many DOE and non-DOE data centers use Mercury. Including NASA DAAC, ARM, NGEA Arctic, SPRUCE, and CDIAC
- Leverage investment in existing information systems and research
- Based Open source software and some custom ORNL developed Software. Free to use under Copyright policy# TXu 1-873-934 CR No. 50000002.

Mercury Supported Projects



Outline

- Data/Metadata Registration
 - Introduction and Capabilities
- Data Search and Discovery
 - Brief Introduction
 - Architecture and Operation
- Data Citation
 - Introduction and Capabilities

Outline

- Data/Metadata Registration
 - Introduction and Capabilities
- Data Discovery
 - Brief Introduction
 - Architecture and Operation
- Data Citation
 - Introduction and Capabilities

Online Metadata Editor: Introduction

- Allows PIs and data providers to easily document their data
- Based on easy-to-use web form, but standards compatible
 - Controlled vocabulary, thesaurus and semantics support. Example: CF variables.
- Has built-in QA/QC and approval workflow. Example: Approval workflow and QC on Date, Location, etc.
- Scalable and customizable for project specific requirements
 - Currently being used in climate, biogeochemical, biological, ecological and ocean sciences

Carbon

- Atmospheric
- Plant
- Soil
- Ocean
-

Invasive Species

- Introduced species
- Invasive plants
- Invasive Animals
-



Projects using OME

DOE BER

- **ARM** (<http://archive2.ornl.gov/armome/>)
- **NGEE Arctic** (<http://ngee-arctic.ornl.gov/ngeemetadata/>)
- **CDIAC** (<http://mercury.ornl.gov/OceanOME/>)

Other Projects

- U.S. Geological Survey - Core Science Analytics, Synthesis, and Libraries
- Surface Ocean CO2 Atlas (SOCAT) OME
- NASA Distributed Active Archive Center (ORNL DAAC)
- Climate.data.gov

Outline

- Data/Metadata Registration
 - Introduction and Capabilities
- Data Discovery
 - Brief Introduction
 - Architecture and Operation
- Data Citation
 - Introduction and Capabilities

Data Discovery Using Mercury: Introduction

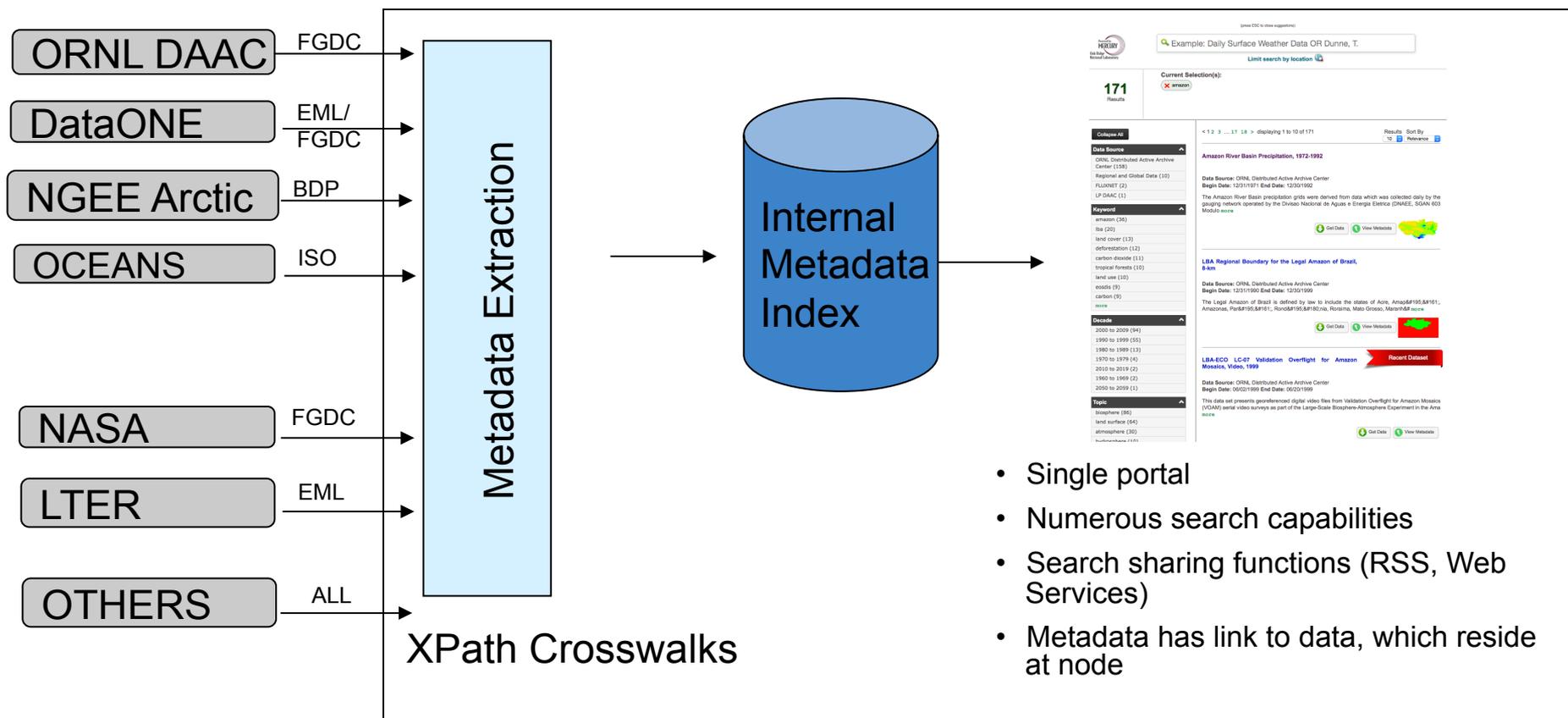
Mercury: Data Search and Discovery Tool

- Distributed metadata management, data discovery and access system.
- Provide a single portal to information contained in disparate data management systems
- Supports various metadata standards including ISO 19115, CF, FGDC, EML, GCMD, DIF, DC
- Allow PIs and database managers to distributed their data while maintaining complete control and ownership



<http://mercury.ornl.gov>

Mercury System Architecture



Normalization

Fields: During indexing and querying

- Keywords: e.g. U.S. DOE vs DOE
- Spatial Coordinates: e.g. 180 N vs 90 N
- Temporal: e.g. 04/2010, 2010 April
- Other fields

External Mapping File

information\ management\ system=Maps and Data
raster\ digital\ database=Maps and Data
comma\ delimited\ text\ and\ ms\ excel\ spreadsheet\ in\ yearly\ files=Maps and Data
pamphlet=Publications
vector\ digital\ data;\ esri\ shapefile=Maps and Data

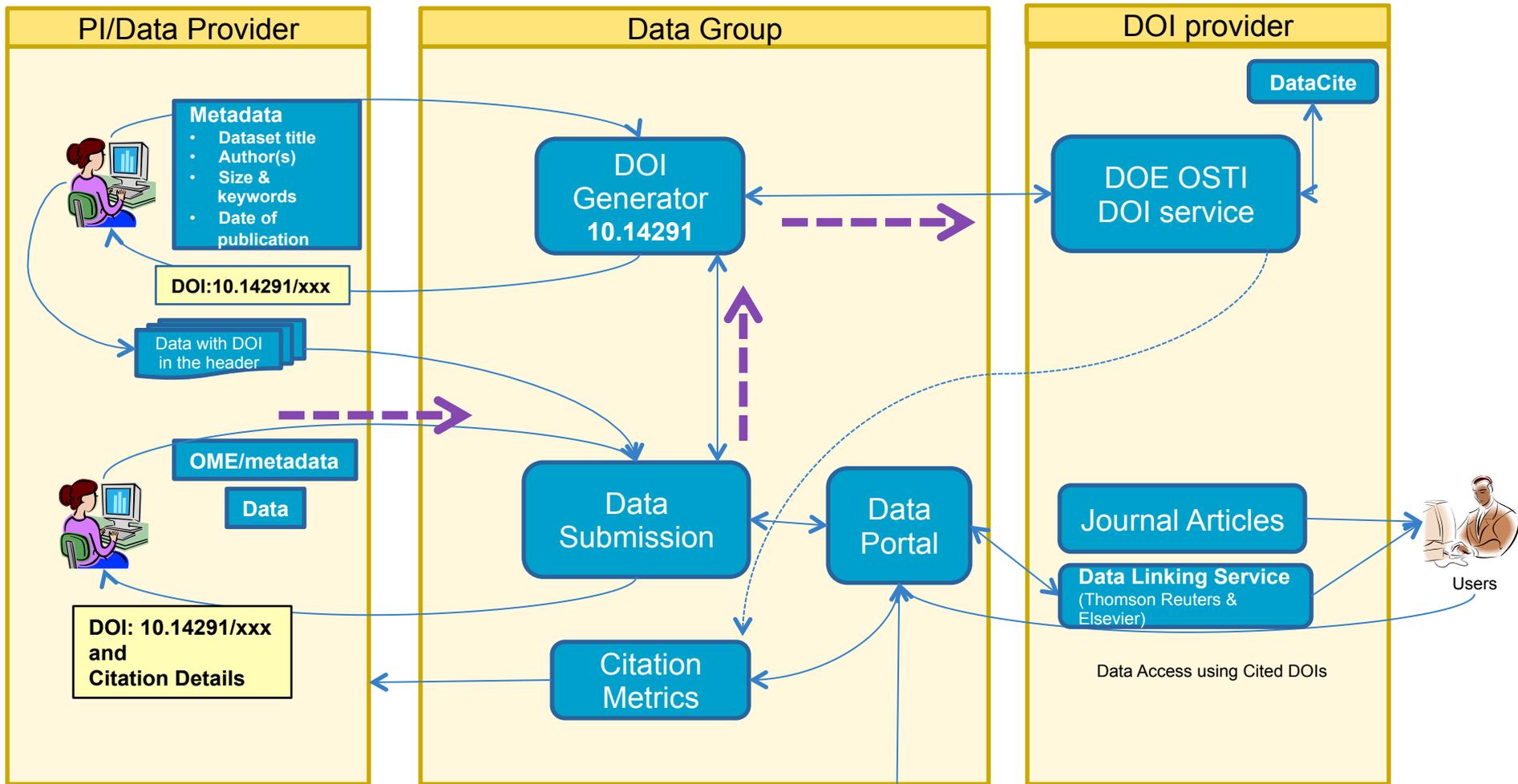
Outline

- Data/Metadata Registration
 - Introduction and Capabilities
- Data Discovery
 - Brief Introduction
 - Operation
- Data Citation
 - Introduction and Capabilities

Data Citation using Digital Object Identifier

- A Digital Object Identifier (DOI) is one type of unique, persistent identifier that is permanently assigned to a specific electronic resource.
- Uses OSTI's DOI service to mint DOIs with DataCite
- To generate the DOIs, we prepare citation metadata for their digital content; this includes information about creator, title, place where data resides (URL) and publication date.

DOI registration workflow



Acronyms

- ARM: Atmospheric Radiation Measurements
- BDP: Biological Data Profile
- CDIAC: Carbon Dioxide Information Analysis Center
- CF: Climate and Forecast Convention
- DataONE: Data Observation Network for Earth
- DC: Darwin Core
- DIF: Directory Interchange Format
- DOE: U.S. Department of Energy
- DOI: Digital Object Identifier
- EML: Ecological Metadata Language
- FGDC: Federal Geological Data Committee
- GCMD: Global Change Master Directory
- ISO 19115: International Organization for Standardization
- LTER: Long Term Ecological Research
- NASA: National Aeronautics and Space Administration
- NGEE Arctic: Next-Generation Ecosystem Experiments
- OCEAN: Ocean Carbon Dioxide
- OME: Online Metadata Editor
- OSTI: Office of Scientific and Technical Information
- SPRUCE: Spruce and Peatland Responses Under Climatic and Environmental Change