EMBEDDING PERSISTENT IDENTIFIERS INTO ORGANIZATIONAL INFORMATION AND DATA SERVICES AT THE NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Matt Mayernik, Deputy Director, NCAR Library, 0000-0002-4122-0910
Jennifer Phillips, Director, NCAR Library, 0000-0002-4861-0292
Greg Madden, Chief Information Officer, UCAR|NCAR, 0000-0001-8316-9749

CNI Spring 2023 Membership Meeting
April 3-4, 2023

NCAR
UCAR

Outline

- Organizational background
- Immediate motivation
- Assigning PIDs
  - DOIs & ARKs
  - RORs & ORCIDs
- Using PIDs
  - Linking scientific papers to underlying data, software, and other resources
  - Gathering impact metrics
- Connecting internal systems
- Summary & lessons learned

NCAR
UCAR

NCAR & UCAR

- NCAR is a Federally Funded Research and Development Center

- UCAR manages NCAR via a Cooperative Agreement with the NSF

- UCAR has 120+ member colleges and universities

- Provide facilities and community coordination, and conduct research

- ~1200 employees

NCAR
UCAR
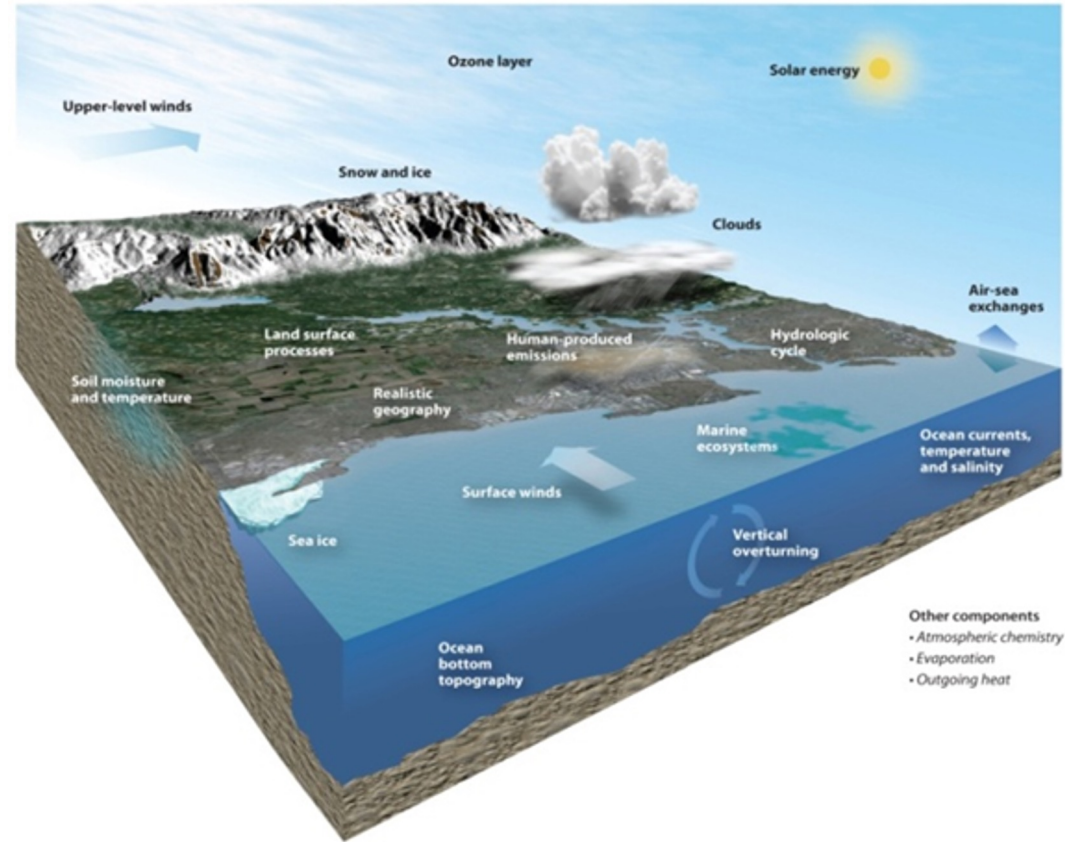
Our Immediate Motivation

- Matt will talk more about the larger-scale motivations for our work on PIDs and specifically ORCIDs
- Immediate motivation is that we received a UCAR President's Strategic Initiative Fund grant to:
  - Improve the adoption of ORCIDs across NCAR|UCAR
  - Complete some initial integrations of ORCIDs into systems
  - Help with NCAR's ongoing Research Impacts Initiative
  - And drive towards a virtuous cycle of integrations, shown on the next slide

Potential Research Product Integrations

1	**OpenSky**	• ORCID used to maintain database of organizationally-affiliated research output
2	**Geoscience Data Exchange (GDEX)**	• ORCID used as an authentication service for collaborators & ORCID included in metadata records
3	**Research Data Archive**	• ORCID used as an authentication service for collaborators & ORCID included in metadata records
4	**Climate Data Gateway**	• ORCID used as an authentication service for collaborators
5	**EOL Field Data Archive**	• ORCID included in metadata records
6	**DASH Search**	• ORCID included in metadata records

Potential System Integrations

1	**Integration System**	• Enables easier integrations to additional applications • (MuleSoft)
2	**Research Administration System**	• Open data, secure data, and cyberinfrastructure planning • State-of-Research reporting • (Kuali Research)
3	**Organizational Financial System**	• Integration from RA System to Financial System enables better tracking of spending • (Workday Financials)
4	**Organizational Human Resources System**	• Ties ORCID to researcher identity in the HR system • (Workday HCM)
5	**Research Information Management System**	• Research Impact Metrics • Sponsor reporting • (Selection in progress)
6	**Identity Governance and Directory Services**	• Enables use of ORCIDs as federated identity for use by collaborators • (Fischer Identity, Active Directory)

Assigning Persistent Identifiers (PIDs)

"The term 'persistent identifier' is usually used in the context of digital objects that are accessible over the Internet. Typically, such an identifier is not only persistent but also actionable…"

McMurry, J. A., et al. (2017). Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLOS Biology*, 15(6): e2001414. https://doi.org/10.1371/journal.pbio.2001414

PID Assignment

UCAR has been assigning Digital Object Identifiers (DOIs) and Archival Resource Keys (ARKs) since 2012 (numbers as of 3/24/2023)

DOIs
- Dataset - 10,026
- Text - 1,228
- Software - 41
- Physical Object - 25
- Collection - 5
- Service - 4
- Model - 2
- Computational Notebook - 1
- Interactive Resource - 1
- Event - 1
- Other - 1

DataCite

ARKs
- Article - 23694
- image - 4008
- archival document - 3609
- conference material - 3526
- newsletter - 1918
- technical report - 666
- manuscript - 584
- chapter - 566
- book - 174
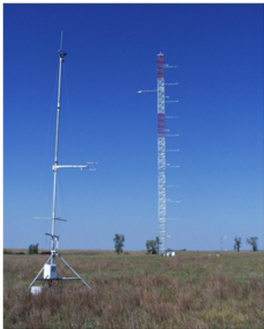- animation/moving image - 134
- other - 240

ARK

PIDs for Facilities and Instruments

https://doi.org/10.5065/D6ZC80XJ

https://doi.org/10.5065/D6WM1BG0

Integrated Surface Flux System

ISFS DESCRIPTION

NSF/NCAR C-130

Aircraft Overview

The Lockheed C-130 "Hercules" aircraft is a four-engine, medium-size utility aircraft that has proven to be one of the most well-known and versatile aircraft ever built. The NSF/NCAR aircraft is a model EC-130Q, similar to the more common model C-130H model except for electrical and air-conditioning modifications. The aircraft is an all-metal, pressurized, high-wing monoplane powered by four Allison T-56- A-15 turboprop engines. It is equipped with dual-wheel, tricycle landing gear with the main gear wheels arranged in tandem and the nose gear arranged side-by-side. The C-130, maintained and

C-130

NSF/NCAR C-130

NSF/NCAR C-130 Investigator Handbook
Airborne Instrumentation
NSF/NCAR C-130 Request Guidance
Aircraft Schedules
Request the NSF/NCAR C-130
Contact RAF

https://ncar.github.io/FAIR-Facilities-Instruments/

NSF Awards #2226396, 2226397, 2226398

NCAR UCAR

PIDs Generated Externally

Organizational Identifiers - Research Organization Registry (ROR) IDs

- UCAR and NCAR have IDs in ROR
- Two of seven NCAR top-level laboratories have RORs
- No UCAR sub-units have RORs

Personal Identifiers - ORCIDs

- Current staff - est. 400+
- Past staff - est 300+

Using Persistent Identifiers (PIDs)

"In order for scholars to benefit from the use of persistent identifiers, there needs to be a clear case made to them about the benefits of taking the time to do so."

Klump, J., Murphy, F., Weigel, T. and Parsons, M.A., (2017). Editorial: 20 Years of Persistent Identifiers – Applications and Future Directions. *Data Science Journal*, 16, p.52. http://doi.org/10.5334/dsj-2017-052

Networked Science
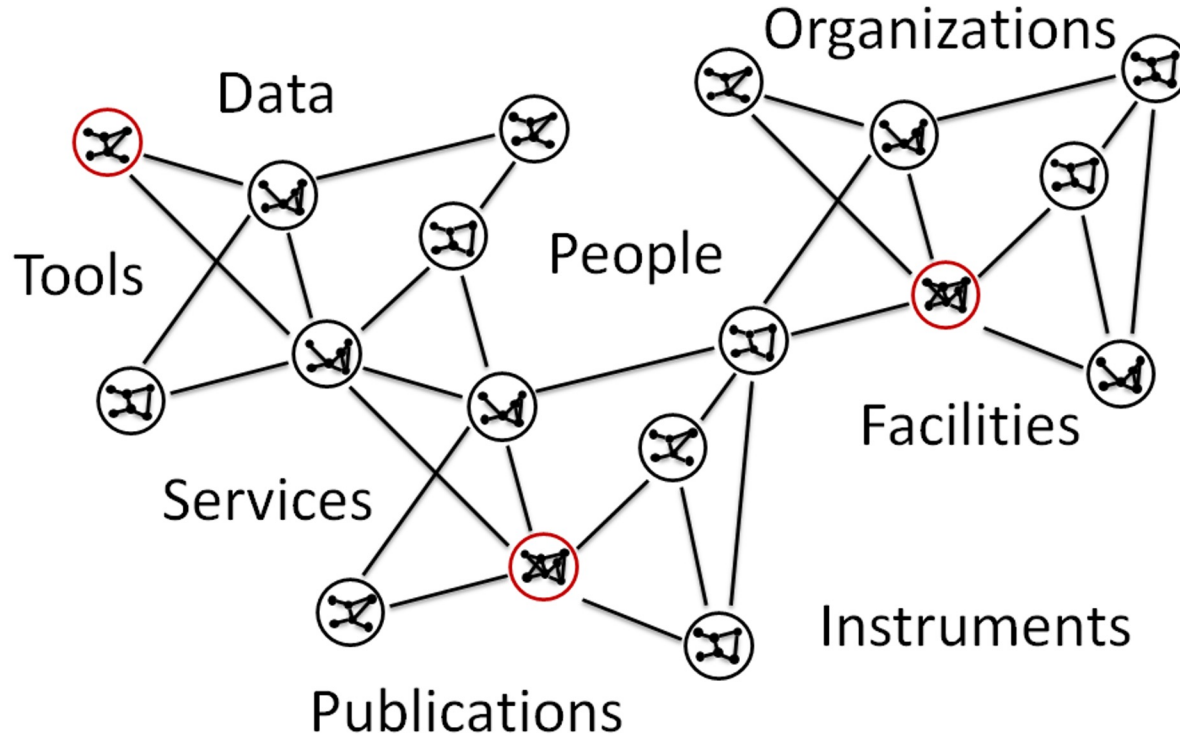
Two Cases

1. Linking scientific papers to underlying data, software, and other resources
   a. Goal - Enable discovery and access of related resources

1. Gathering impact metrics
   a. Goal - Assessing the impact of individual resources, identifying the diverse contributions to scientific work

Case 1 - Linking scientific papers to supporting data sets, software, and other resources

- Can we collect and display linkages to other resources or information?

- How best to make this information useful and understandable for users?

- How to do this in a tractable and sustainable way, e.g. automated as much as possible without much maintenance?

PID Linking - Workflow

- Parse PDFs of papers in Library institutional repository for DOIs

- Query DataCite metadata store to determine which DOIs are for data sets, software, facilities and services

- Add metadata for data sets and/or software to OpenSky records as "Related" information

- Display on paper landing pages

NCAR
UCAR

Description

An estimate of the exchange of kinetic energy between the large-scale atmospheric eddies and the zonal flow for January 1963 for the troposphere north of 20°N latitude has been made by means of a quasi-geostrophic method. This more complete result is compared with the anamolous one obtained... Show more

DOWNLOAD PDF

In collections
- SIParCS Collection

Supporting Datasets
- East African Great Lake Environments (EAGLE) Climate Downscaling Output

Supporting Software
- NCAR Command Language

Details

Author(s)	Min Huang
	James H. Crawford
	Gregory R. Carmichael
	Joseph A. Santanello
	Sujay V. Kumar
Date	2020-01-27
Volume	125
Issue	2
Title	Impact of Aerosols From Urban and Shipping Emission Sources on Terrestrial Carbon Uptake and Evapotranspiration: A Case Study in East Asia?
Date Published	2020-01-27
Resource Type	article
Subject	Oral history
	more Scripps Institution of Oceanography, University of California, San Diego

Case 2 - Using PIDs for Research Impact Metrics

NSF and other sponsors want quantifiable measures of the merit and broader impact of NCAR science (facilities & research)

Bibliometrics and facility usage statistics (e.g. HPC) are useful, but insufficient to represent value and impact of the Center

Multi-year Research Impacts Initiative to define new metrics and instrument a Research Information Management System for NCAR|UCAR

Case 2 - Using PIDs for Research Impact Metrics

RIM system (selection in process)
- Builds on
  - Established practice of managing citation record for NCAR-authored peer-reviewed publications
  - Custom database for staff activity reporting (conferences, panels, field campaigns)
  - Practice of assigning DOIs for outputs other than publications, and for instruments/facilities

Current challenges include
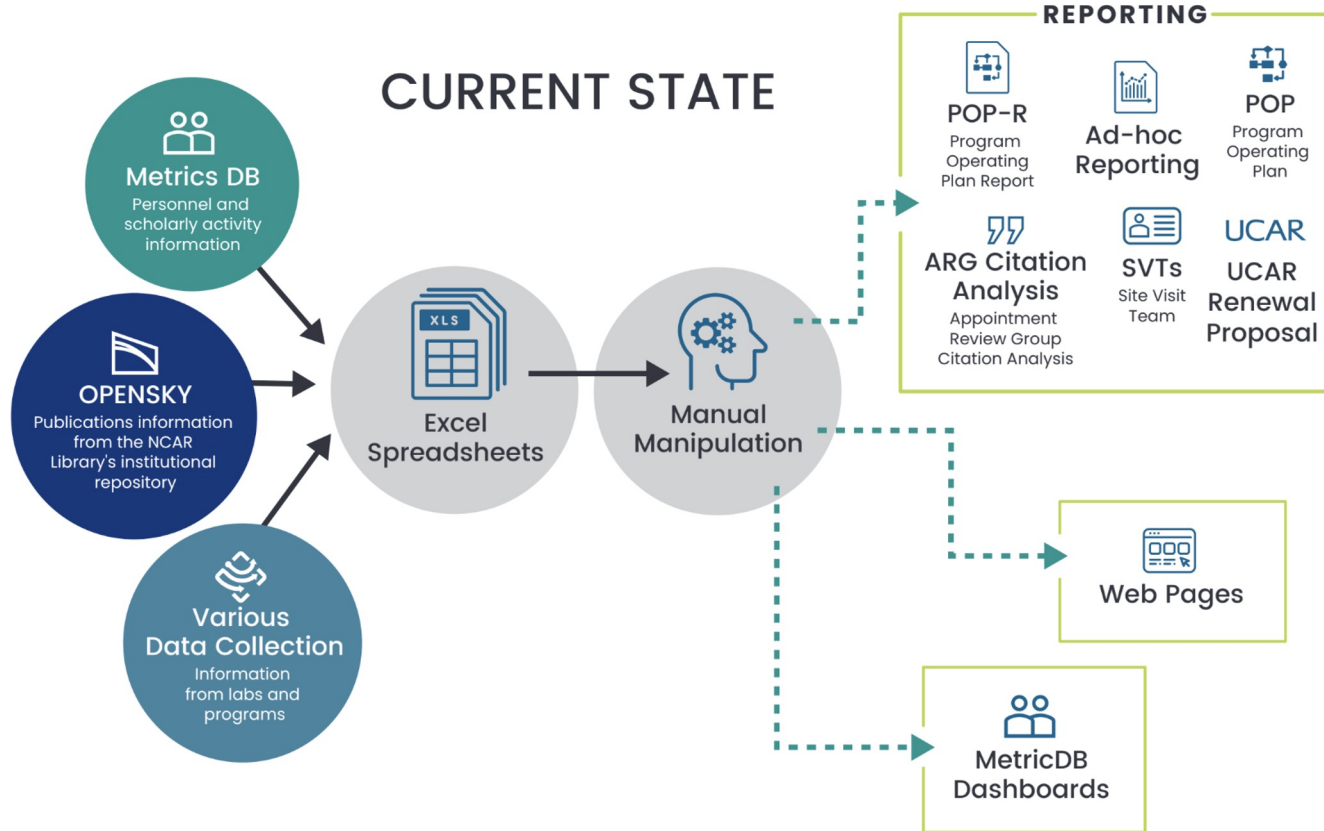
- No common controlled vocabularies

- Lack of standard identifiers for personnel, collaborators, and organizations

- Lots of customization and standalone systems

Goals for RIM system implementation
- Connectivity of UCAR systems (e.g. publications data with grants & awards system)
- "Interoperability" with broader Research Analytics ecosystem (e.g. CrossRef, Funder Registry)

NCAR
UCAR

CURRENT STATE

Metrics DB
Personnel and scholarly activity information

OPENSKY
Publications information from the NCAR Library's institutional repository

Various Data Collection
Information from labs and programs

Excel Spreadsheets

Manual Manipulation

REPORTING

POP-R
Program Operating Plan Report

Ad-hoc Reporting

POP
Program Operating Plan

ARG Citation Analysis
Appointment Review Group Citation Analysis

SVTs
Site Visit Team

UCAR
UCAR Renewal Proposal

Web Pages

MetricDB Dashboards

NCAR
UCAR

FUTURE STATE

Workday — Personnel Information from HR systems

ORCID

OPENSKY — Publications information from the NCAR Library's institutional repository

Kuali — Information on grants and sponsored programs

RIM — Research Information Management System

OUTPUTS
- Public Profiles
- Future/Novel Use Cases (Equipment pages/ Project pages etc)
- Ad-hoc Reporting
- OPENSKY
- Dash-boards
- Lab/Program Webpages

General Considerations

- PIDs alone do not provide much value - The value comes through embedding PIDs in other systems, connecting to them, visualizing connections, and leveraging relationships

- Inconsistent use of PIDs is still a limiting factor - e.g. unpopulated ORCIDs, inconsistent data citations

- Inconsistent use of PIDs can be actively deceiving - undercounts of data citations, incomplete collaboration networks based on ORCID profiles with gaps

- External services are improving, but inconsistent - e.g. Web of Science, Scopus, Dimensions, CrossRef

- PIDs require management and maintenance