Investigating Data Use Diversity with Usage-Based Discovery

Dr. Chris Lynnes, ESIP Discovery Cluster
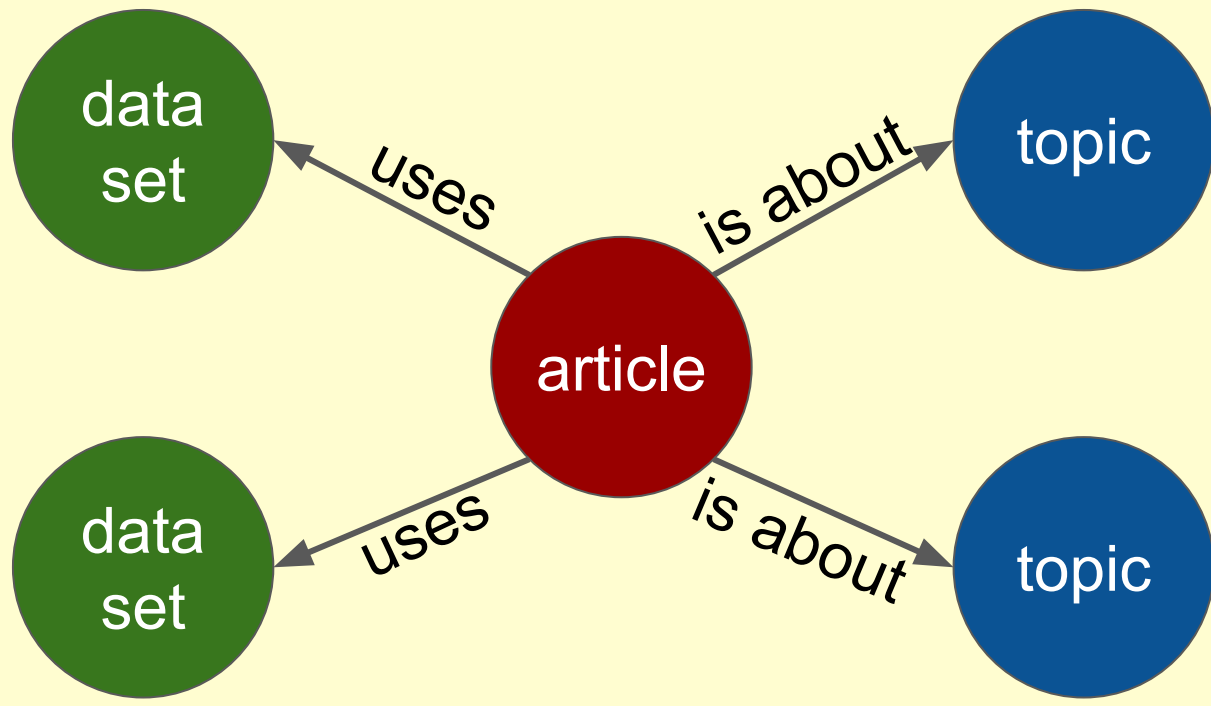
ESIP — MAKING DATA MATTER

One of the spinoffs of Usage-Based Discovery is the ability to investigate the myriad ways Earth Observation data are used. The knowledge graph powering the Usage-Based Discovery system (via the ESIP Discovery Cluster) includes relationships between dataset and article, and article and topic area. This graph allows us to examine and even measure the diversity of uses for a particular dataset.

Usage-Based Discovery (UBD) relies on connections between research articles and the datasets they use. Those articles are classified into one or more topics for to aid user navigation

These connections can serve other purposes, such as investigating how data products are being used.

Totals

Totals			
Articles	8475		
Dataset Uses	14261	Datasets Used per Article*	1.7
Datasets (unique)	2022		
Articles with Topic	1518	Topics per Article	2.4
Topic Assignments	3602		

*Undercount because:
1. Articles provided by DAACs (Distributed Active Archive Centers may also use datasets from elsewhere
2. Many datasets are neither cited nor described precisely enough to identify them

Multi-purpose Datasets
Datasets whose articles addressed 16 or more topics in aggregate

Topics	Dataset Name
23	ASTER Global **Digital Elevation Model** V003
20	MODIS/Terra **Vegetation Indices** 16-Day L3 Global **250m** SIN Grid V006
20	TRMM (TMPA/3B43) **Rainfall** Estimate L3 1 month 0.25 degree x 0.25 degree V7
18	MODIS/Terra+Aqua **Land Cover** Type Yearly L3 Global 500m SIN Grid V006
18	Daymet: **Daily Surface Weather** Data on a 1-km Grid for North America, V. 2
18	Daymet: **Daily Surface Weather** Data on a 1-km Grid for North America, V. 3
17	MODIS/Terra **Land Surface Temperature**/Emissivity **8-Day** L3 Global 1km SIN Grid V006
17	MODIS/Terra **Vegetation Indices** 16-Day L3 Global **1km** SIN Grid V006
16	Global Gridded Surfaces of Selected **Soil Characteristics** (IGBP-DIS)
16	GRACE Monthly Land **Water Mass** Grids netCDF Release 5.0

CAVEAT! Most of the dataset usage information so far comes from NASA DAACs: Land Processes, Oak Ridge Natonal Laboratory (Biogeosciences), Physical Oceanography, and:Goddard Earth Sciences Data and Information Services Center (Atmospheric Science and Hydrology). This will affect the prevalence of topics.

Journal Topics

Distribution of topics for journals whose articles in the UBD database address 17 or more topics in aggregate. Numbers in the cells are articles from UBD tagged with that topic.

Journal	topic count	Aerosols	Air Quality	Atmos Chem	Weather	Clouds	Rad Budget	Cyclones	Precipitation	Ocean Phys	Ocean Biol	Ocean Chem	Sea Level	Climate	Water Res	Floods	Snow + Ice	Volcanoes	Fire	Solid Earth	Geodesy	Earthquakes	Landslides	Soil	Land Surface	Vegetation	Biodiversity	Agriculture	Human Activ	Data Sci
Geophys Res Lett	26	4	2	8	12	3	7	7	11	2	1	3	2	15	9	0	3	0	7	0	3	1	1	5	1	12	1	2	8	1
Nature Sci. Reports	25	3	1	2	3	2	1	3	5	5	1	1	2	5	0	2	2	0	2	0	1	1	2	1	1	7	2	2	6	0
Remote Sensing	24	3	1	3	9	0	4	5	2	0	1	6	10	3	5	4	6	1	8	0	2	3	14	20	0	9	9	12
Sci of The Total Env	23	4	6	2	5	1	0	5	0	0	1	4	7	1	2	1	3	1	2	0	1	4	2	9	0	1	12	3
Rem Sens of Env	22	3	0	1	2	0	1	3	1	1	0	5	3	3	4	2	3	0	0	3	0	2	4	6	16	0	2	7	6
JGR Atmospheres	21	8	20	22	8	2	6	19	4	0	5	0	12	3	0	4	0	3	1	0	11	8	11	0	1	7	1
Water Resrc Res	18	0	0	0	1	0	1	1	9	0	0	1	0	1	15	3	2	0	0	1	3	0	2	8	3	5	0	4	5	3
Atmos Chem & Phys	18	21	6	20	10	13	8	1	2	1	0	6	0	0	1	0	7	10	0	0	0	0	1	0	5	1	1	8	0
Scientific Data	17	1	0	2	1	0	1	2	0	0	0	1	4	2	0	1	0	1	1	0	1	1	0	5	0	4	1
Env Res Lett	17	1	2	1	3	0	3	0	2	0	0	0	10	6	2	2	0	3	0	0	0	5	4	6	0	4	9	1

Diversity on the Long Tail: sampling of journals with only one article in UBD

Journal	Article
Fire Safety Journal	Self-ignition of natural fuels: Can wildfires of carbon-rich soil start by self-heating?
Australian Journal of Grape and Wine Research	Review of smoke taint in wine: smoke-derived volatile phenols and their glycosidic metabolites in grapes and vines as biomarkers for smoke exposure and their role in the sensory perception of smoke taint
Preventive Medicine	Hot weather and risk of drowning in children: Opportunity for prevention
Learned Publishing	Examination of data citation guidelines in style manuals and data repositories

What's Next?
- Classify more articles by topic? (~6K currently unclassified)
- Obtain more diverse, balanced population of articles and datasets?
- See ESIP Discovery Cluster session, Wed. at 11:00