Usage-based Discovery

Chris Lynnes
with contributions by
ESIP Discovery cluster members: many
GSFC Interns: Maggie Zhu, Megan Mehta, Parth Darji, Sophia Xia

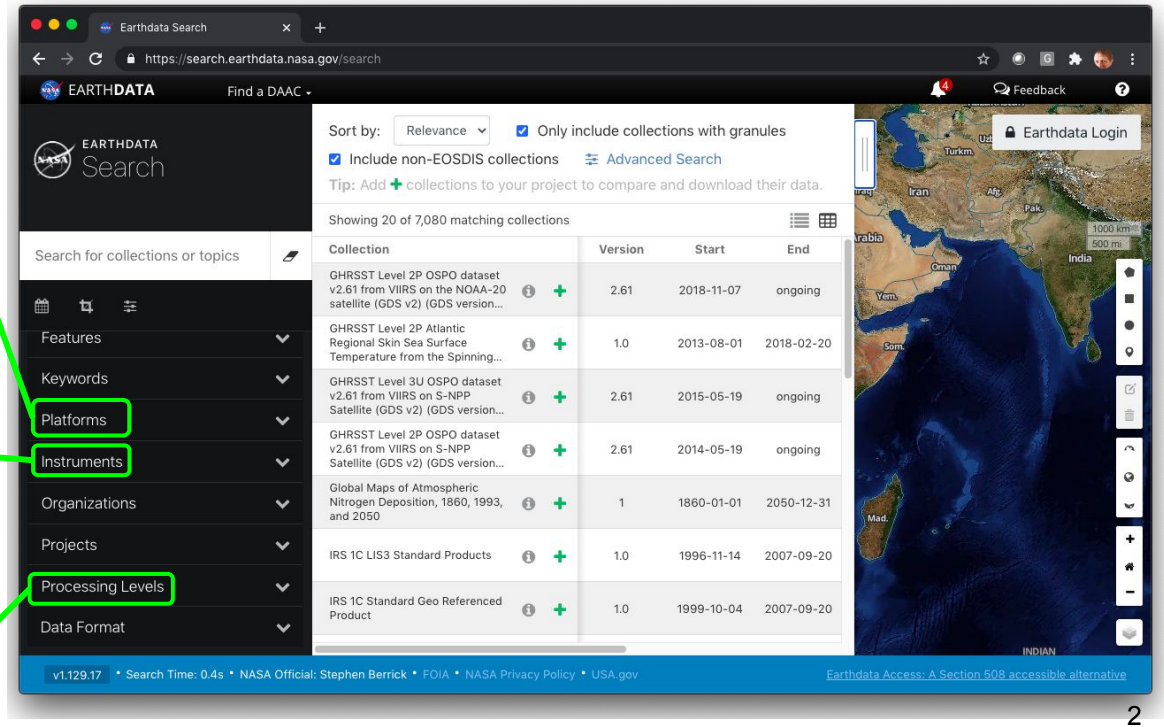
Current discovery interfaces present dataset characteristics

Platforms

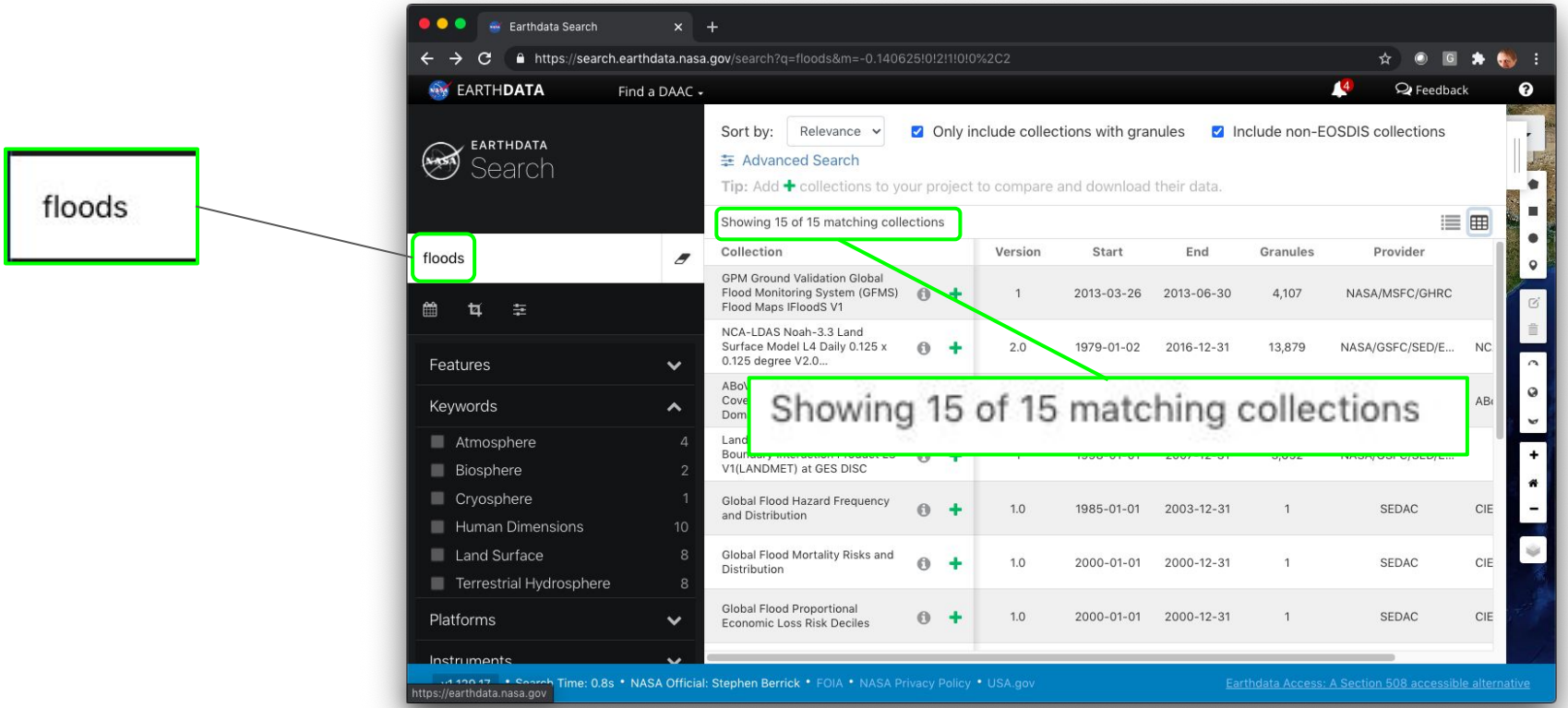
Instruments

Processing Levels

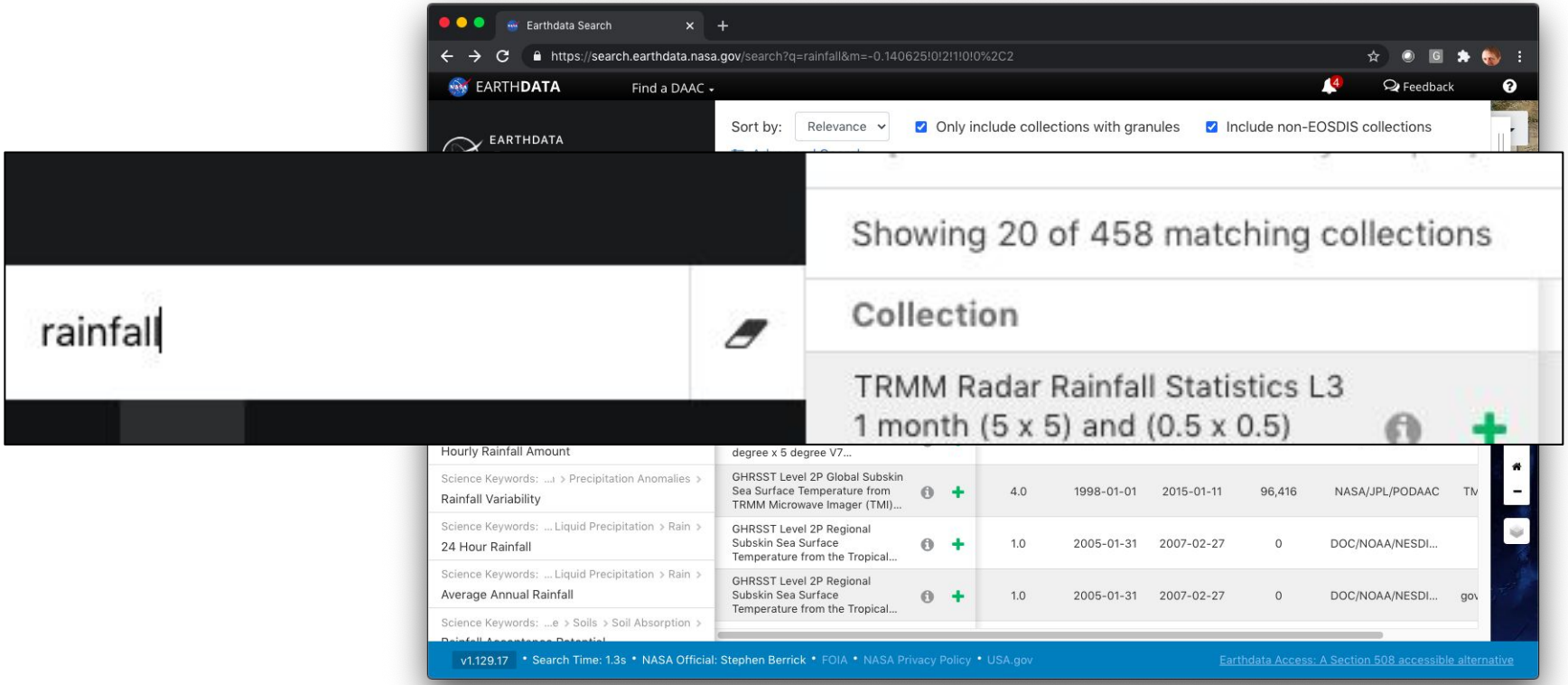
Earthdata Search



Current discovery interfaces present dataset characteristics



What about Rainfall?



400+ Results: too much of a good thing?

OK, what *kind* of rainfall data do you want for your flood monitoring?

- Satellite or rain gauge or merged?
- Spatial resolution?
- Temporal resolution?
- Spatial coverage?
- Temporal coverage?
- Quality?
- Latency?
- Reliability?

Let's try a new paradigm: Usage-based Discovery

"What datasets does <*favorite flood application*> use?"

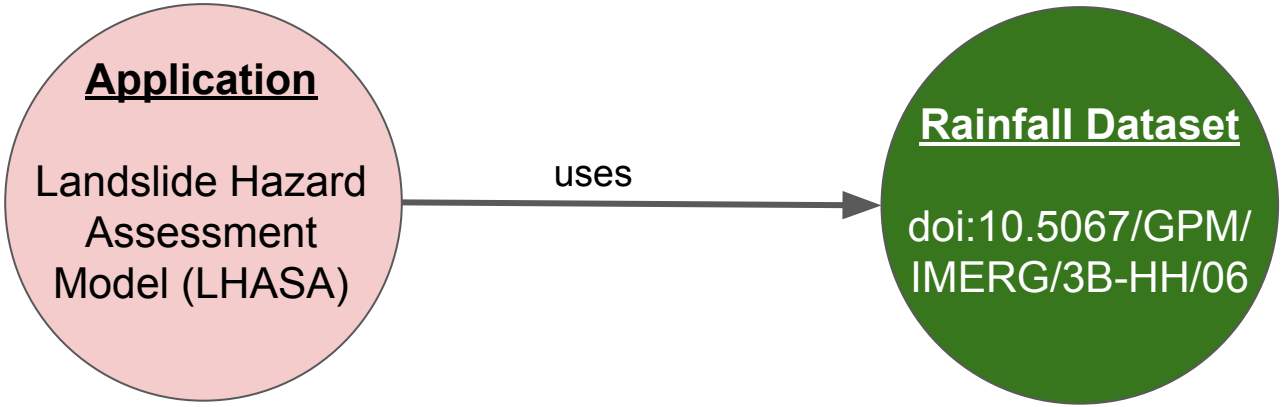
aka the

"*I'll have what they're having*" paradigm

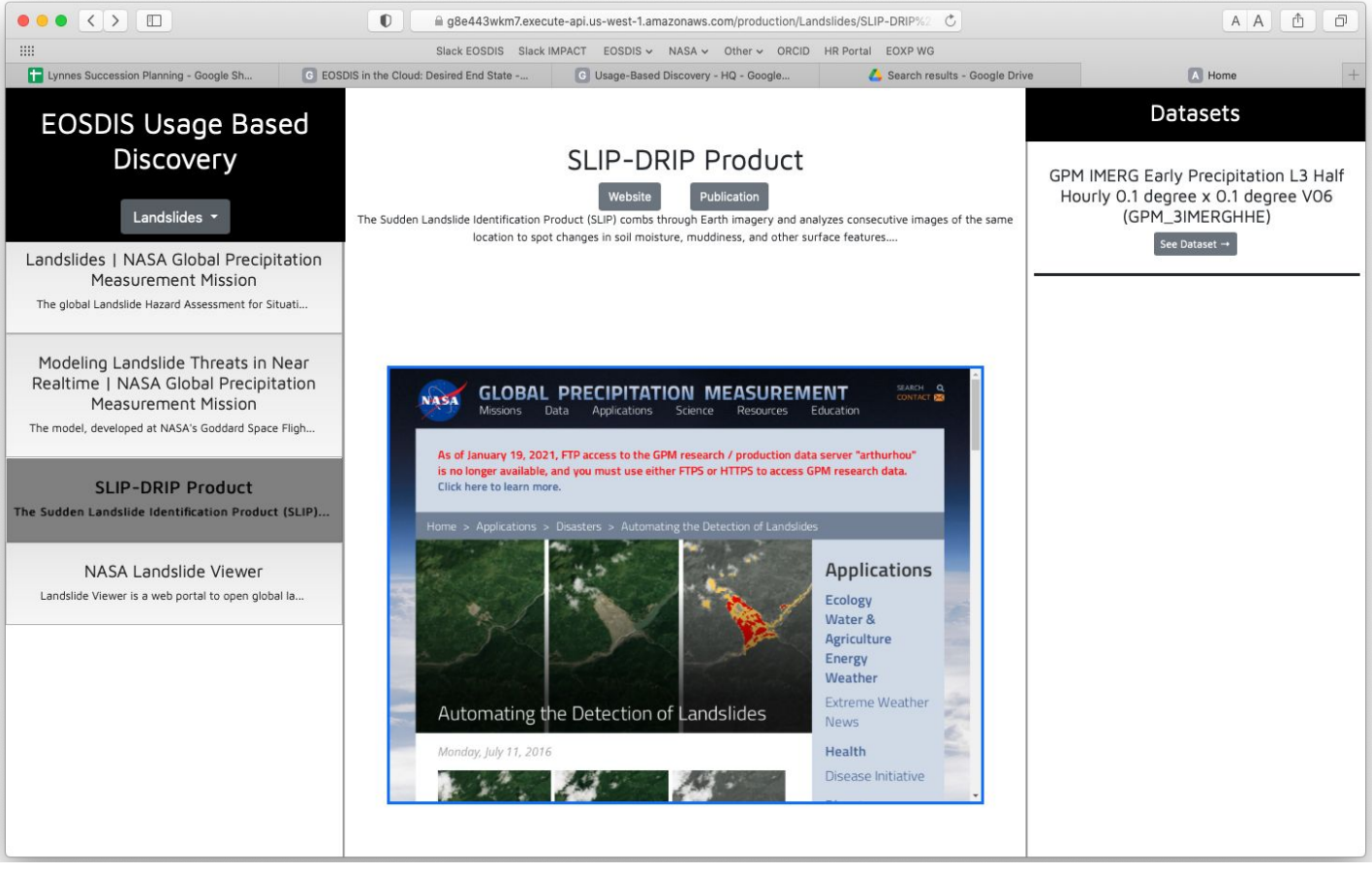
aka

Social Data Discovery

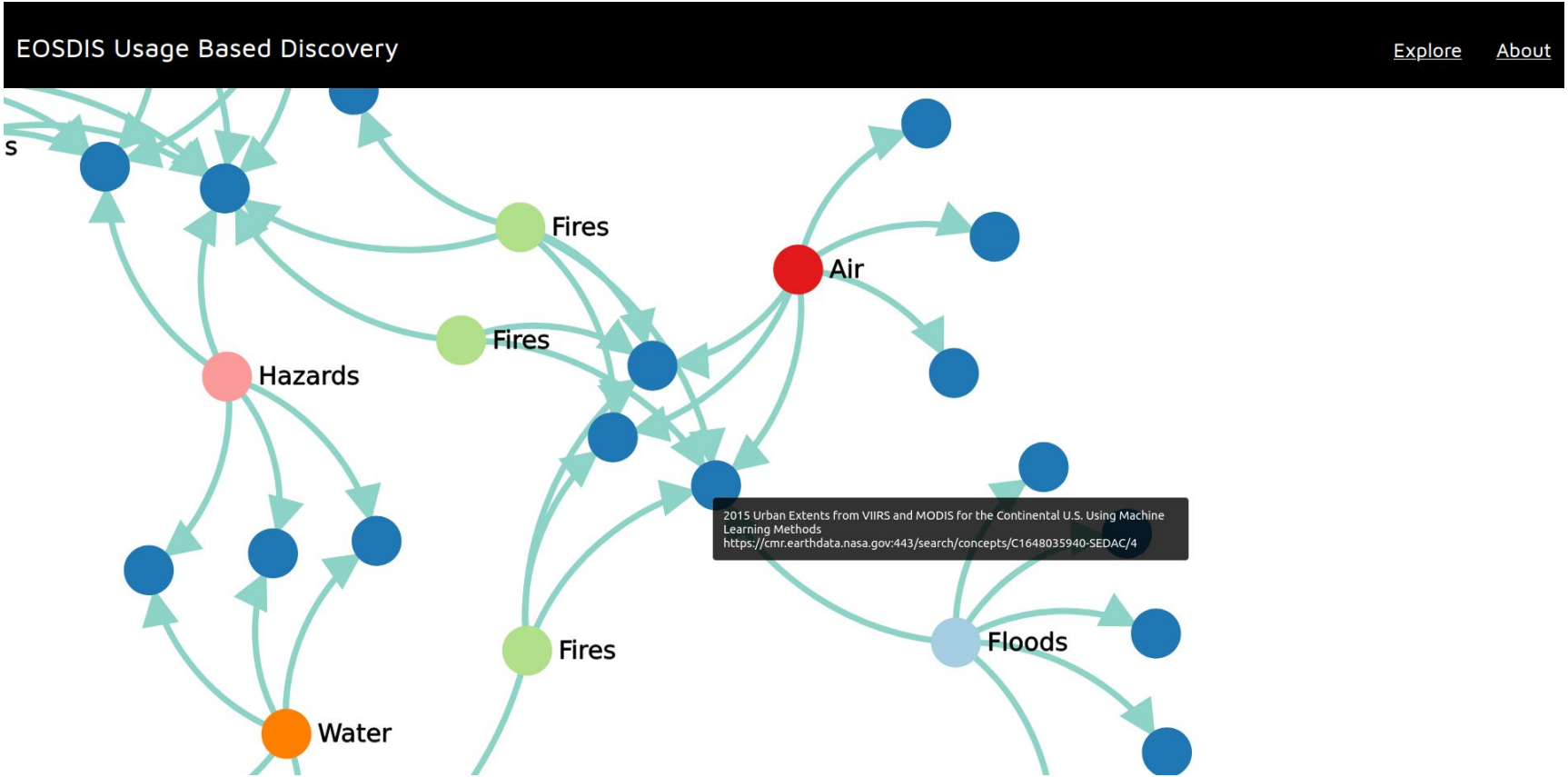
Graph relationship from Application to Dataset



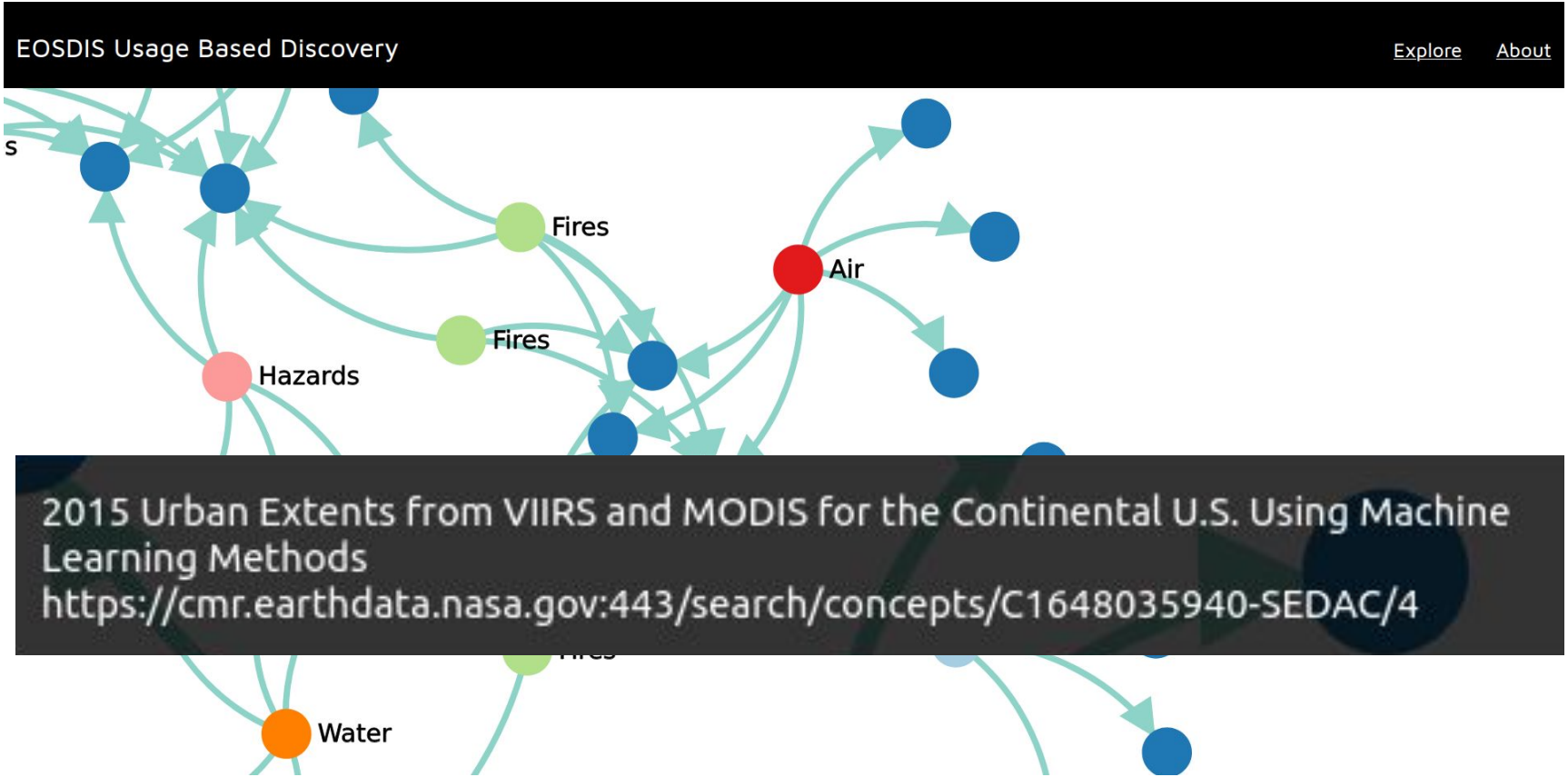
Usage-based Discovery User Interface



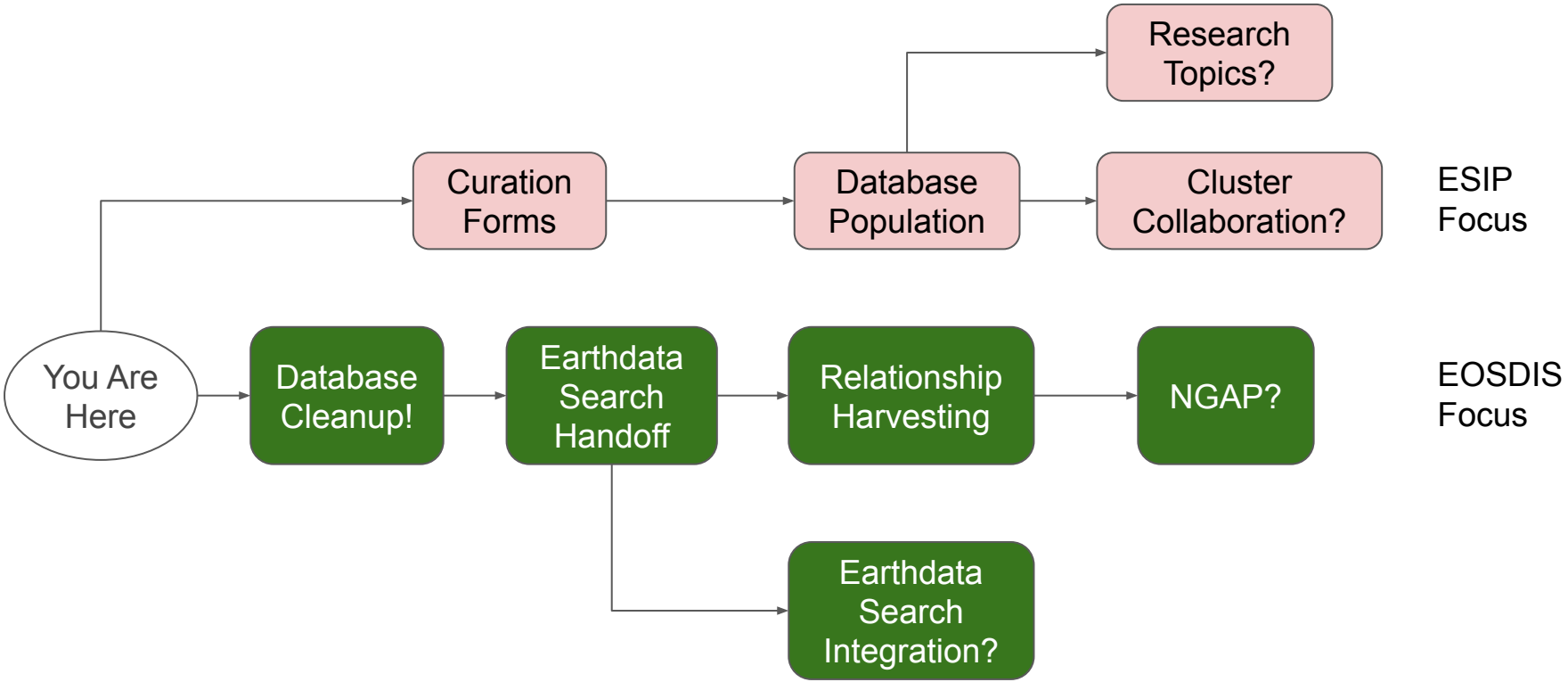
Hmmm...


EOSDIS Usage Based Discovery
Explore
About
s
Fires
Air
Fires
Hazards
2015 Urban Extents from VIIRS and MODIS for the Continental U.S. Using Machine Learning Methods
https://cmr.earthdata.nasa.gov:443/search/concepts/C1648035940-SEDAC/4
Fires
Floods
Water

Hmmm...


EOSDIS Usage Based Discovery
Explore
About
Fires
Air
Fires
Hazards
2015 Urban Extents from VIIRS and MODIS for the Continental U.S. Using Machine Learning Methods
https://cmr.earthdata.nasa.gov:443/search/concepts/C1648035940-SEDAC/4
Water

Current Plan



Speculation

- Minimal dependence on metadata consistency could help federate heterogeneous resource repositories
  - CMR, Earth Science Data Pathfinders, EONET, ESCCOR, IMPACT KG...
  - NASA: SDMWG?

- Usage ⇔ Dataset relationships have policy benefits
  - Impact metrics: what/how many applications/articles use dataset X?
  - Identification of collaboration opportunities