Prototype Data Resource Registry for the GHG Community

NIST MML Office of Data and Informatics

Gretchen Greene – Group Lead Data Science

gretchen.greene@nist.gov

Collaboration with: Tamae Wong, Anna Karion, James Whetstone, Raymond Plante, Kim Tryka, Logan Mitchell, Conor Gately, Francesca Hopkins, Lorraine Sugar, Thomas Lauvaux, Kevin Gurney, Marcus Newrock

Building Data Resource Registries

Goal:

Facilitate Scholarly Discovery and Navigation to key data resources and information to support synthesis and validation of GHG science in a ***federated*** community

Data Resources provide metadata, e.g. :

- Datasets – public, research, and private restricted
- Services – programmatic, custom, etc.
- Portals – Data Archives, Community, Research Project
- Organizations – Gov, Industry, Academic, Groups, Project
- Documentation - publication resources, key
- Tools/Software – Web Tools, Services, Applications

Proven Model in operations:

- IVOA (International Virtual Observatory Alliance)
- NIST IMMR (Global Metrology network)
- Materials Genome Initiative (MGI)

Focused Domain Expertise - reliability of the information – data quality, scholarship

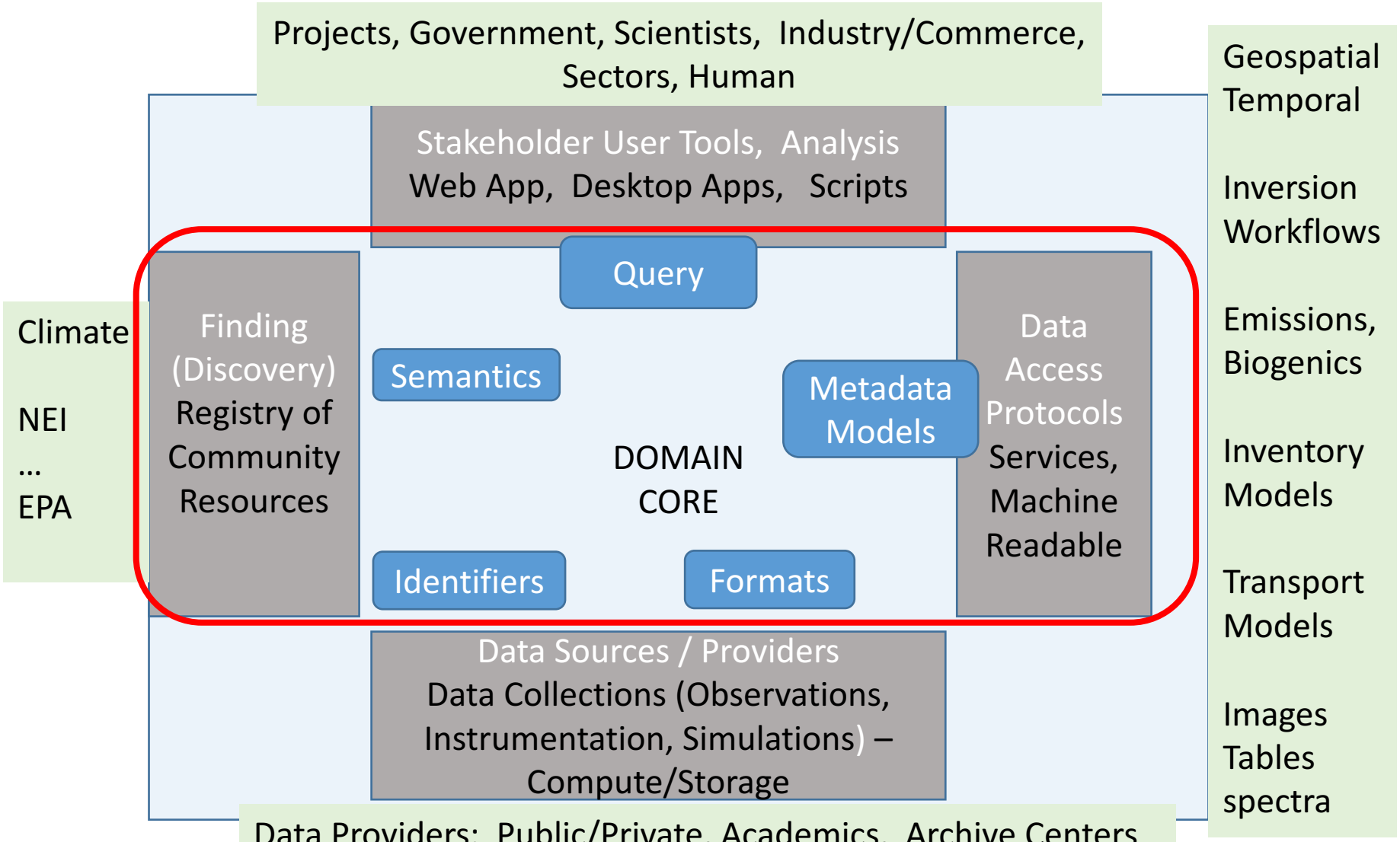
Opportunity for FAIR Data Science

- FAIR – relatively new US legislative efforts to extend the *open data* initiatives to improve upon scientific investment and validation
  - **Findable** – Discoverable resources which represent the domain space information
  - **Accessible** – Services for data access (tools can access data) , protocols for machine use
  - **Interoperable** – agreed upon standards/protocols for describing (metadata) datasets
  - **Re-usable** – tools and analysis recognize the data formats, standards and deliver results to query

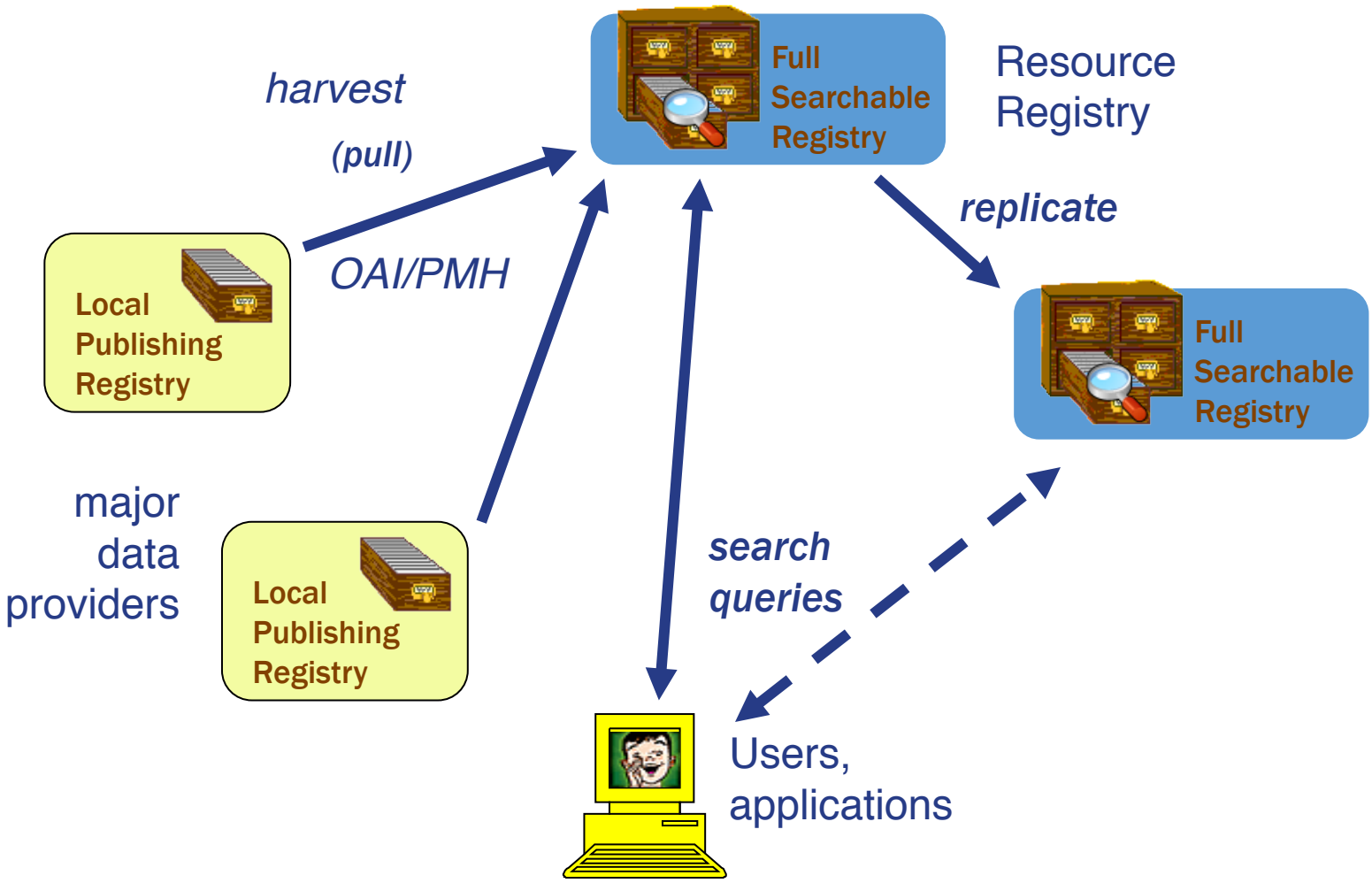
NOTE – FAIR does not preclude flexibility in use of methods or standards implementation

https://www.nature.com/articles/sdata201618

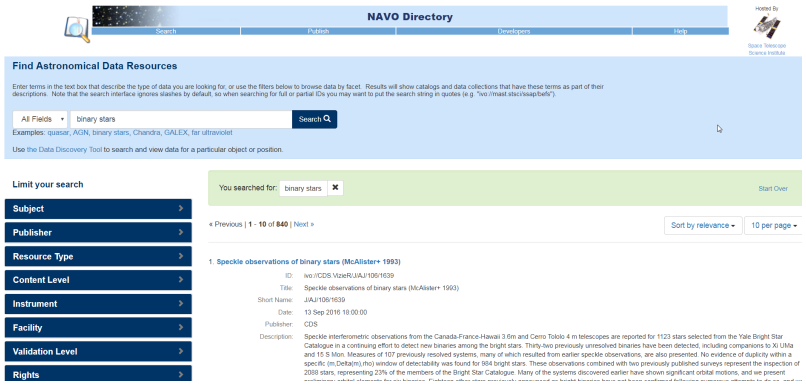
GHG Registry Context -example FAIR Framework



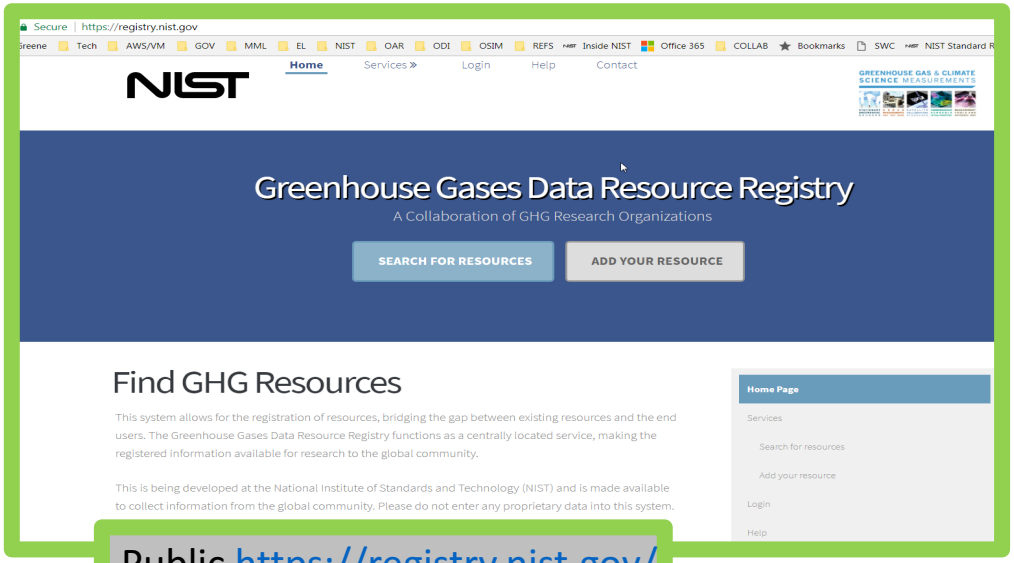
Federated Architecture



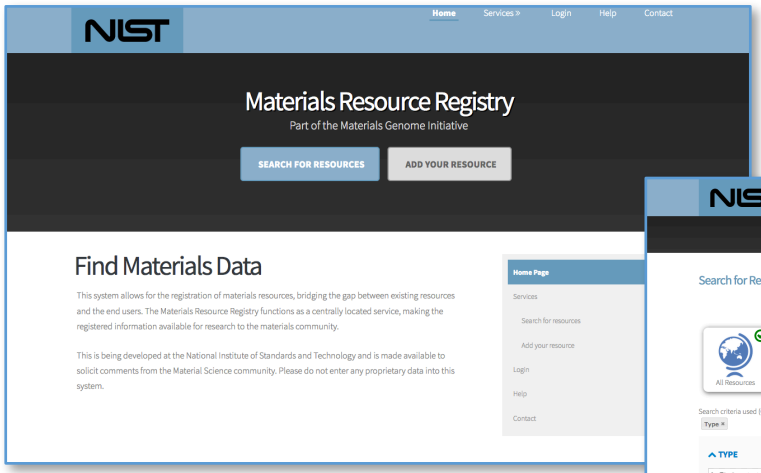
Resource Registry in Practice

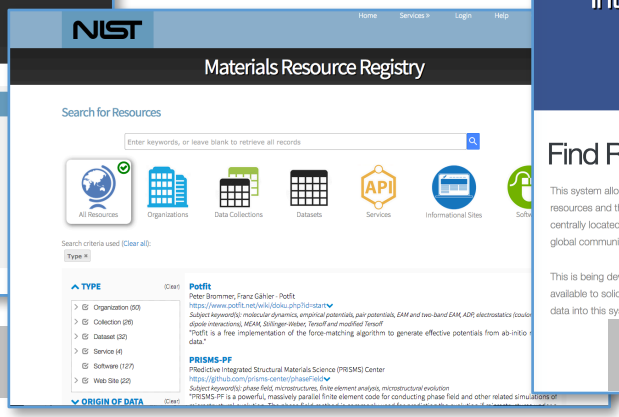


http://vao.stsci.edu/keyword-search/



Public https://registry.nist.gov/





https://materials.registry.nist.gov/

http://imrr.bipm.org/

Internal only

Next Steps

- Collect Feedback on Registry System value
  - Maintain public NIST registry and work with external deployments where possible. Encourage collaboration and feedback.
  - Identify metadata, granularity and usefulness to the GHG groups. Features of use, e.g. programmatic access
- Continue to Collaborate and support GHG Community for Data Interoperability
  - Share best practices and standards for interoperability
- NIST is participating in open source advancement
- Facilitate development of a federated architecture of scientific information in support of the GHG community
  - Distributed and shared expertise across data science space