Discovery Cluster Session: Federated Knowledge Graphs

Email Address: **douglas.j.newman@nasa.gov**

Session Lead Organizer Name **Doug Newman**

Session Lead Organizer Email **douglas.j.newman@nasa.gov**

Other Session Organizer(s): **Jonathan Blythe**

Session Title: Federated Knowledge Graphs

# Session Description

**Theme for conference!**

We’d like to pursue and advance the federation of knowledge graphs. Currently, knowledge graphs represent knowledge from specific domains and often describe specialized knowledge for specific applications. This practice presents a problem for the task of knowledge graph federation, because harmonizing knowledge concepts from disparate knowledge graphs remains a monumental task. However, if we break down the problem and solve it in part, or at least solve it for a particular use case, then we may have a more manageable and achievable problem that we can begin to make more rapid progress on part of the solution.

For this session, we propose that knowledge graph federation can help solve the following use case: Under the Evidence Act, Federal agencies must report on the usage of data that they have collected. In evaluating how to do this, the discovery cluster has come to realize that data usage is not an absolute metric, but it is a metric biased by the tools and publication systems that we employ and subscribe to. Even efforts that leverage machine learning and natural language processes that go out and mine dataset usage information across a variety of publishing platforms, like the Coolridge Initiative that Julia Lane presented at the Discovery Cluster’s ESIP 2022 Winter Meeting Session, may ultimately fall short of standards of systematic and reproducible processes that science agencies like NASA will accept. When we frame this problem as a symptom of the greater challenge of knowledge graph federation, this presents an opportunity to evaluate the problem further and identify opportunities for incremental progress.

So, a large part of the problem of knowledge graph federation are the diverse knowledge systems and domains that must be negotiated and traversed in the federation solution. However, if we ground the knowledge graph in a particular domain, and stick to accepted concepts from this coherent perspective, then we may have a better foundation to begin building the federation solution. As implied through the evidence act and OMB’s evidence building agenda, that shared foundation is provided by the social sciences, and the statistical approaches originating from economics, in particular. OMB defines evidence building as the application of scientific data to questions important to constituencies and demographics through a lens of statistical analysis. For example, a temperature measurement is not just a physical property of an Earth system, but it is contextualized in terms of the social, political, and sectoral constituencies that encounter and have to respond and react to this physical property of the Earth system. Therefore, framing Earth systems dynamics within this social domain may be a way to constrain the knowledge graph federation problem.

We have invited Bob Downs to present work from Center for International Earth Science Information Network (<http://www.ciesin.columbia.edu/>) to understand the social science applications of earth science data and information from the NASA Socioeconomic Data and Applications Center (SEDAC). Bob will describe research using satellite data that cited one or more of SEDAC’s statistical data products, for example, SEDAC’s Global Rural-Urban Mapping Project (GRUMP) collection. From this specific perspective, we will generalize challenges and opportunities for Knowledge Graph federation, using SEDAC citation metrics. This also follows up on the presentation on usage based discovery and breakouts that followed at the ESIP 2022 Winter meeting session “Unearthing semantic web resources for ESIP communities,” and in addition to this federation objective, it appears to also address objectives that the semantics committee has identified for better managing the semantic resources.

If we want to treat knowledge in the same way we treat data, we need to federate the stewardship of it. Discipline-specific knowledge needs to be stewarded by experts in that discipline. However, we recognize that knowledge can be leveraged across disciplines and, as such, needs to be connected. We also need to provide ways to navigate those connections. This is what we mean by federated Knowledge Graphs, it is a means to,

* Connect knowledge across discipline-specific repositories or islands
* Provide a means of navigating between these islands in the same way we allow navigation within an island via graph traversal languages.

# Session Purpose

To demonstrate the need for federated knowledge base techniques by use cases.

To establish a vehicle for guidelines, standards and best practices for the federation of knowledge graphs particularly in the following areas,

* Linking data between knowledge graph ‘islands’
* Maintaining integrity between those links
* Supporting the ability to execute graph traversal querying across knowledge graph island boundaries
* Establish use cases that could be solved by employing the above techniques.

In much the same way that ESIP successfully tackled the federated discovery problem with OpenSearch standards and best practices.

# Outcomes/Goals

* An agreement on whether or not this is a problem worthy of solving
* Is the discovery cluster the right construct to pursue this solution, or should we leverage capabilities across ESIP clusters to investigate solutions to our use cases using a federated knowledge base technique
* Long term - establish standards, technologies and best practices to achieve federated knowledge graphs

# ESIP logistics

Length of Time/Duration: **90 minutes**

Speakers: **Bob Downs, Doug Newman, Vinnie Inverso**

Target Audience:

* **Knowledge graph producers and consumers**
* **Multi-discipline collaborators**

Session Level: **Intermediate/ Advanced**

Are you looking for additional speakers? **Yes**

If you are looking for additional speakers, please describe what kind of speakers/talks you are looking for.:

* **Implementers/owners of knowledge graphs**
* **Experts in Neo4J, AWS Neptune, Gremlin, Sparql**

Recommended Ways for Session Attendees to Prepare:

* [**https://www.oracle.com/autonomous-database/what-is-graph-database/**](https://www.oracle.com/autonomous-database/what-is-graph-database/)
* [**https://kelvinlawrence.net/book/Gremlin-Graph-Guide.html#walk**](https://kelvinlawrence.net/book/Gremlin-Graph-Guide.html#walk)
* [**https://en.wikipedia.org/wiki/Knowledge\_graph**](https://en.wikipedia.org/wiki/Knowledge_graph)

Session Tags: **federated, discovery, graph, knowledge-base, Common Metadata Repository**

Collaboration Area Tags: **discovery, semantic-web**

Additional Session Tags: **none**

Will this session feature or discuss any ESIP-supported outputs? If yes, please list. **No**

Do you have any suggestions of potential plenary speakers in line with the meeting theme of Leading Innovation in Earth Science Data Frontiers? **No.**

Please tell us which session organizers are currently planning to travel to the meeting and who, if any, may desire to participate virtually.

| **Speaker** | **In person/Remote** |
| --- | --- |
| Doug Newman | In person |
| Jonathan Blythe |  |
| Vincent Inverso | In person |

Please tell us which speakers are currently planning to travel to the meeting and who, if any, may desire to participate virtually:

| **Speaker** | **In person/Remote** |
| --- | --- |
| Robert Downs | Unknown at session proposal time |