Convergence Accelerator (CA) - Open Knowledge Network (OKN) - EarthCube Workshop Report
ESIP 2020 Winter Meeting | North Bethesda, MD

January 6, 2020

Image design: Trinka Kensill, NSF

Authors

Melissa H. Cragin, *San Diego Supercomputer Center, UC San Diego*
Ryan McGranaghan, *Atmospheric and Space Technology Research Associates, LLC*
Ilya Zaslavsky, *San Diego Supercomputer Center*
Wenwen Li, *Arizona State University*
Ibrahim Demir, *University of Iowa*
Alyssa Arce, *San Diego Supercomputer Center*
Christine Kirkpatrick, *San Diego Supercomputer Center, UC San Diego*

Organizers

Christine Kirkpatrick, *San Diego Supercomputer Center, UC San Diego*
Melissa Cragin, *San Diego Supercomputer Center*
Ilya Zaslavsky, *San Diego Supercomputer Center*
WenWen Li, *Arizona State University*

TABLE OF CONTENTS

Executive Summary

A cross-disciplinary workshop was organized by the EarthCube Coordination Office (ECO) to bring together researchers, technologists, and federal funders to share expertise on current science and data challenges in the Earth, Space, and Atmospheric Sciences, to consider opportunities and new collaborations and to utilize semantic technologies and computation to address these challenges. Invited participants included geoscience projects funded by the NSF EarthCube program, and Open Knowledge Network projects funded through the NSF Convergence Accelerator (CA); geoinformatics and data experts funded by NOAA, NASA, USGS, and NSF; experts in semantic technologies interested in the solving domain-oriented science problems; and, representatives from U.S. research funding agencies.

EarthCube and OKN projects were invited to submit position statements ahead of the meeting, which helped to provide context and to inform participants on topics related to science drivers for development and use of open knowledge networks, especially in the Geosciences. Additional topics included opportunities and challenges related to Geoscience data, how the EarthCube community might collaborate on CA projects, and the benefits of including a Geosciences track in future CA solicitations. The day consisted of two keynote talks, lightning talks and break-out sessions aimed at building relationships and considering new projects across these research areas. Forty-three people registered for the meeting, another 7-8 people joined the event throughout the day, several from the ESIP GeoSemantics Symposium.

Enthusiasm for the meeting and topic was evident, based on engagement and interactions throughout the day, several follow-on conversations, and responses to an evaluation questionnaire. Participants helped to generate an extensive shared notes document, and several reported they planned to undertake new activities or develop projects. Recommendations can be found in the final section of the report. Key meeting take-aways include:

- A high interest in the EarthCube (EC) and Open Knowledge communities to collaborate on earth, space, and atmospheric science-related problems.
- The opportunity for EarthCube - and the broader Geosciences community - to learn more about services and opportunities for the development and use of “AI ready” data.
- Investment in organizing and labeling data would help produce high-value datasets enabling application of AI and related computational methods.
- Enhanced capacity building to achieve the promise of Open Knowledge Networks, knowledge graphs, and utilization of related methods; in addition to domain researchers, this must also extend to the cyberinfrastructure and data communities.

It is important for the Geosciences communities to be part of - or connected to - the Convergence Science efforts, and for several reasons: There are existing and emergent research areas that require that domain and interdisciplinary knowledge be brought to bear for building and using OKNs. This includes understanding regional-to-national scale systemic effects of climate change; impacts of space weather on built and natural environments; and geo-health, which extends well beyond NIH-focused activities. The breadth of complex data types produced and analyzed in geosciences has served to develop deep disciplinary expertise in data and problems related to collecting, processing and labeling, integration and re-use. In addition, EC project representatives learned about the CA process, and found that several aspects of this would be of value and benefit the broader EC community.

Introduction

A one-day workshop funded by the NSF was held on January 6th, 2020, to cross-pollinate communities with ideas and expertise among Earth Scientists, Semantic (or Knowledge) engineers, and other geoscience enthusiasts to identify shared scientific challenges and motivate new collaborations. This report provides an overview of the topics and problem space that emerged during the event, and related concerns identified by the authors of the report.

The appendices of this report include: A. The workshop agenda; B. The participant list; C. Position Statements submitted ahead of the event; and, D. A list of presentations with links to that content.

Colocated with the Earth Science Information Partners (ESIP) Winter meeting in Bethesda, Maryland, the workshop brought together EarthCube projects and PIs, governance leaders, and others, together with relevant Convergence Accelerator pilot projects. Therefore, balance was sought to have participation across the geoscience community and emergent OKN communities to include projects that concern GIS (geographic information system), spatial data, or otherwise connect to the geosciences.

The aims of this meeting were to: 1) identify possible EC resources and partners that may enhance CA Track A pilot project plans as teams prepare for Phase 2 proposals; 2) facilitate a potential new Open Knowledge Network project to join the CA Track A program in the 2020 NSF solicitation; and, 3) identify a potential Track C opportunity for a 2021 proposal related to geospatial and earth-observation data (or otherwise overlapping with the geosciences). The workshop was organized by EarthCube Coordination Office, in partnership with the Council of Data Facilities (CDF).

This NSF-funded workshop was supported by the GEO Directorate's EarthCube program and the IIA's Convergence Accelerator program, to bridge cognate communities interested in development of shared infrastructure that will support interdisciplinary inquiry at the frontiers of earth science, applied computation, semantics, and data management.

Selected participants were asked to provide in advance a short whitepaper on research areas needing convergence. The workshop format included grouped lightning talks, and guided breakouts on topics that fed into the workshop report outline. A quick-turnaround report was intended to include input for the Convergence Accelerator 2020 solicitation; resulting information was shared and is now part of this community report. The workshop was successful in facilitating new connections between EarthCube members and Convergence Accelerator (CA) teams for potential proposals for future CA solicitations.

Participant scientists and data experts included those engaged in data-intensive research across many disciplines, in need of knowledge representation tools and services (and underlying data resources) necessary to analyze structure and function of geological phenomena at different scales of analysis; forge new understanding of multi-system interactions; and, to facilitate inferential computation.

The NSF Convergence Accelerator (CA) pilot program is intended to address these very opportunities, and “identify areas of research where investment in convergent approaches – those bringing together people from across disciplines, united to solve problems – have the potential to translate to high-benefit results and advance ideas from concept to deliverables… and, to support fundamental scientific exploration by creating partnerships that could potentially include stakeholders from industry, foundations, government, nonprofits and other sectors."[1]

The EarthCube program at NSF is designed to support research and training for the geosciences, and the wider research community through development of tools, methods, standards, architectures, and community connections. The EarthCube Coordination Office (ECO) exists to benefit geoscientists and the wider Earth, Space, and Atmospheric Sciences community through production, support and improved access to software tools, methods, standards, architectures. The ECO serves to identify synergies among geoscience and related community activities, and to facilitate connections and project opportunities.

The CA Track A program addresses Harnessing the Data Revolution (HDR) and emphasizes the use of Open Knowledge Networks (OKN). Open Knowledge Networks[2] are intended to generate "non-proprietary shared knowledge infrastructure, allowing stored data" to be located along with their relationship to real-world concepts or objects, as understood and described at a semantic level[3]. Earth and related Space Science communities are seeking opportunities to generate new resources that will advance the shared understanding of the state of knowledge in their respective disciplines. Further, development of new, related semantic tools will support application of data science and machine learning methods throughout the field.

Convergence Acceleration for the GeoSciences

The NSF Big Ideas initiative was launched in 2017 that included "Growing Convergent Research," and "Harnessing the Data Revolution." For the EarthCube community, and the geosciences broadly, an opportunity lies at the intersection of these, in the Open Knowledge Network area of the Convergent Accelerator (CA) program. The CA aims to "facilitate the development of "[a] new organizational structure to accelerate the transition of convergence research into practice, in areas of national importance."[4] As noted on the program webpage, "NSF's Convergence Accelerator effort is a new capability within NSF to accelerate use-inspired, convergence research in areas of national importance via partnerships between academic and non-academic stakeholders." Indeed, Open Knowledge

---

[1] National Science Foundation. https://www.nsf.gov/od/oia/convergence-accelerator/index.jsp Accessed 4 November 2019.
[2] National Science and Technology Council (2018). Open Knowledge Network: Summary of the Big Data Working Group Workshop, Oct. 4-5, 2017, Bethesda, MD. Available: https://www.nitrd.gov/news/Open-Knowledge-Network-Workshop-Report-2018.aspx
[3] National Science Foundation. (2019). Dear Colleague Letter: Request for Information on Future Topics for the NSF Convergence Accelerator. NSF 19-065, May 6, 2019. Available: https://www.nsf.gov/pubs/2019/nsf19065/nsf19065.jsp?WT.mc_id=USNSF_179
[4] C. Baru, NSF Convergence Accelerator workshop presentation. January 6, 2020. https://drive.google.com/open?id=1q2kX8e26KT7ZUX42TuAI_AChb0PCEjEh

Networks (OKNs) are one aspect of new knowledge systems needed to support convergence research problems in the earth and space sciences[5].

Open Knowledge Networks

The federal Open Knowledge Network programs emerged from collaboration launched among industry, government, and academic leaders to address the need for non-proprietary information infrastructure and data stores that will support AI applications and innovation across the national R+D system. From the initial workshop in 2015[6], it was clear that advances and products from the Big Tech companies were based on knowledge graphs and networks among them, and that this sort of vast data resource is needed for academia and open science, as well. The aim is to develop a network connecting high quality, annotated (or labeled) datasets that indicate explicit entity relationships or features, such that the networks support computational approaches that generate hypotheses or illuminate patterns or relationships not “visible” before. In addition to more traditional statistical approaches, researchers can leverage the scale of these data resources for training, and emerging Deep Learning techniques. For the Geosciences, potential impacts include opportunities to address grand challenges in geohealth, guiding theories for understanding geo-timescales, and building and testing new models for regional-to-national scale environmental change.

Challenges and Opportunities

Solutions to grand challenges for the Earth, Atmospheric, and Space Sciences require multi- or interdisciplinary research, and methods that can leverage integrated data that might span spatial or temporal resolution; utilize data generated by vastly different modes; or, that have varying bias types and error rates. AI and related approaches are maturing, yet to take advantage of these methods, researchers interested in developing and utilizing OKNs face fundamental social and technical barriers, and limitations of capacity in research-related services. These barriers are situated in practice, cyberinfrastructure, and data resources, and each must be addressed to realize the full potential of evolving knowledge systems. Opportunities and challenges were recently described in depth (Gil et al, 2019[7]), and several aspects of these challenges were evident, if not explicit, in this workshop, and point

[5] Yolanda Gil, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, and Shashi Shekhar. 2018. Intelligent systems for geosciences: an essential research agenda. Communications of the ACM, 62(1), 76–84. DOI:https://doi.org/10.1145/3192335
[6] C. Baru, NSF Convergence Accelerator workshop presentation. January 6, 2020. https://drive.google.com/open?id=1q2kX8e26KT7ZUX42TuAl_AChb0PCEjEh
[7] Gil, Y., Pierce, S.A., Babaie, H., Banerjee, A., Borne, K., Bust, G., Cheatham, M., Ebert-Uphoff, I., Gomes, C., Hill, M. and Horel, J., 2018. Intelligent systems for geosciences: an essential research agenda. Communications of the ACM, 62(1), pp.76-84. DOI:https://doi.org/10.1145/3192335

to the need for a range of activities to broaden (and even democratize) production, availability, and use of OKNs and related tools and services.

Some challenges at the intersection of EarthCube and the OKN activities will require funding to address foundational concerns, including capacity building for analyzing data that spans temporal scales; and, to develop solutions to complex issues such as quantifying geological process rates, which would improve organization of field data. These sorts of standards work will facilitate the use of new kinds of queries.

Just as fundamental is the need to express geoscience domain knowledge in OKNs. The community must address basic questions such as:

- How can geoscience data discovery and integration challenges be supported by a knowledge network?
- What formalisms are best suited for querying geo-focused OKNs (GraphQL, GeoSparql, or others)?
- How will such queries account for spatiotemporal relationships?

A related challenge concerns the processes, for and coordination of, selecting areas of the earth, space, and atmospheric sciences that should be prioritized for knowledge extraction and application of semantics. Relevant questions that must be addressed include:

- Which geoscience vocabularies shall be used to describe nodes and edges in a geo-aware knowledge network?
- Which metadata elements or data summaries will be most important for extracting knowledge nuggets from datasets, to be added to knowledge graphs?
- How can we maintain consistency in geospatial knowledge graphs as they are populated with statements derived at different spatial scales?

Assessment and decisions on how best to represent a variety of geoscience data types for use in knowledge graphs, especially across different spatial and temporal resolutions and with different data and metadata quality, is another challenge. For example, statistical summaries for well-defined administrative units such as states or counties, appear to be well-handled in existing knowledge graphs. However, knowledge representation, querying and reasoning over other types of geoscience information, including floating point and integer grids, or arbitrary user-defined areas, would require novel approaches.

GEO Domain Cases - Contexts for Opportunity

Space Weather

Geoscience is inextricably linked to space science, and the connection is strengthened by humanity's ease of access to and utilization of both near-Earth and deeper space. Understanding the physics of our space environment is one of the grand challenges of our time. Those physics are determined by stellar energy and its interaction with interplanetary space and each of the bodies in the stellar system. Nowhere are these interactions closer than in our own planet's stellar-terrestrial interactions, the manifestations of which are colloquially known as space weather (e.g., Shrijver et al., 2015)[8]. Our understanding and specification of space weather require a holistic systems-level knowledge cutting across the near-Earth space environment, through the magnetosphere and interplanetary space, to the Sun[9]. The traditional paradigm has been to compartmentalize these regions, which has produced remarkable knowledge yet has also led to three outstanding barriers to new progress: (1) the lack of a cohesive community, due to the wide variety of subject matter experts working on topics related in various ways to space weather, (2) the lack of effective data sharing, coordination, and analysis (e.g., data science) to leverage existing resources and knowledge efficiently; and (3) the diversity of physically dominant processes in each section of the space weather environment, making it difficult to relate various models and observations. In short, space weather and the space sciences face both an exciting opportunity and an important imperative to create a new frontier of linked link data, communities, and disciplines through the Open Knowledge Network, and the "Convergence Hub for the Exploration of Space Sciences" (CHESS[10]) project serves as an example to address this opportunity.

Disaster Resilience

It is estimated that natural disasters are affecting more than 250 million people worldwide every year and are causing billions of dollars damage to economies. The National Research Council[11] highlights the importance of data for a resilient community and reports the challenges and limitations of conventional sensing technologies and modeling efforts and describes the need for next-generation solutions for monitoring, prediction, and communication of data and information. Open Knowledge

---

[8] Schrijver, C.J., Kauristie, K., Aylward, A.D., Denardini, C.M., Gibson, S.E., Glover, A., Gopalswamy, N., Grande, M., Hapgood, M., Heynderickx, D., Jakowski, N., Kalegaev, V.V., Lapenta, G., Linker, J.A., Liu, S., Mandrini, C.H., Mann, I.R., Nagatsuma, T., Nandy, D., Obara, T., O'Brien, T.P., Onsager, T., Opgenoorth, H.J., Terkildsen, M., Valladares, C.E., Vilmer, N. (2015). Understanding space weather to shield society: A global road map for 2015–2025 commissioned by COSPAR and ILWS, Advances in Space Research, 55(12), 2745-2807. https://doi.org/10.1016/j.asr.2015.03.023.

[9] McGranaghan, R. M., Bhatt, A., Matsuo, T., Mannucci, A. J., Semeter, J. L., & Datta-Barua, S. (2017). Ushering in a new frontier in geospace through data science. Journal of Geophysical Research: Space Physics, 122, 12,586– 12,590. https://doi.org/10.1002/2017JA024835

[10] CHESS project: https://www.chessscience.com/

[11] National Research Council. 2012. Disaster Resilience: A National Imperative. Washington, DC: The National Academies Press. https://doi.org/10.17226/13457

Networks have potential to drive the next generation breakthroughs in data management, integration, computation and intelligence. Powered by Artificial Intelligence, Deep Learning, Internet of Things, Intelligent Systems, and Virtual and Augmented Reality, OKNs can significantly transform scientific research, as well as decision-making and operational applications in disaster preparedness and response.

Geospatial data and science

Geospatial data, or location-based data on or near the Earth's surface, is central to many geoscience applications, ranging from regional to global climate analytics, disaster management and response, to hydrological modeling, among others. Geographic Information Science (or GIScience), the science discipline that studies the theory, methods, techniques and applications to store, manage, discover, analyze and visualize geospatial data, has become the nexus connecting multiple geoscience domains to support collaborative spatial decision making and knowledge discovery. To this end, GIScience research and applications demonstrate the spirit of convergence research. Besides providing a problem-solving environment, GIScience also offers a unique perspective for addressing regional to global challenges spatially. In particular, the two notable spatial principles, spatial autocorrelation and spatial heterogeneity provide important guidance for building location-aware (machine learning) models to understand physical processes underlying different geoscience phenomena. In addition, the spatial principles can be extended to the spatiotemporal realm to advance understanding of global change and facilitating event prediction. Space and time are also key dimensions to index geoscience data, models and tools in a knowledge graph toward building a robust open knowledge network.

These domain examples illuminate the kinds of opportunities that are emerging to organize and coordinate engagement with stakeholders from across the breadth of the community. To fully realize these opportunities and grow the utility of OKNs for research and development, new capacity is required, including:

1. **OKN readiness** to address how best to prepare geoscience datasets for use in knowledge networks for applications of AI and other disruptive technologies. This will require added activities, such as
    a. Dataset validation
    b. Metadata augmentation/enrichment
    c. Ontology harmonization (cross-walking, or similar processes)
    d. Semantic tagging
    e. Deriving spatiotemporal extents
    f. When is a knowledge graph “ready for reasoning”?
    g. Fusing and updating OKNs - how will this work?

2. Consideration of the place and role(s) for “**human in the loop**”
3. Understanding limitations: **governance, training, sustainability**

a. What types of OKN queries raise concerns or violate ethics, privacy, regulatory, or IP constraints, and therefore should be blocked, set aside, or marked, “not answerable?”
b. What training is demanded specifically for geoscientists?
c. What would OKN governance look like, and who will be responsible for what and how conflicts/mismatches would be handled?
d. How will OKNs be developed and sustained? How are roles and responsibilities defined and distributed across or among organizations, institutions, the civic sector, or others?
e. How would OKN governance interact with other data governance (e.g., for metadata, PIDs, vocabularies, catalogs, tools, DOIs, protocols, etc.)

Summary and Recommendations

Earth, Atmospheric, and related Space Science communities are seeking opportunities to generate new resources that will advance the shared understanding about the state of knowledge within and across their respective disciplines, and that will benefit from OKN efforts that support the cross-disciplinary inquiry. Investments in data organization and “labelling” would contribute to production of high-value datasets that enable reuse in support of new inquiry, and particularly for application of AI and related computational methods. **Capacity building is needed to achieve the promise of Open Knowledge Networks, knowledge graphs, and utilization of related methods**; in addition to domain researchers, this must also extend to the cyberinfrastructure and data communities.

Development and application of new semantic tools will support the generation of new data (and metadata) resources, and in turn, facilitate the application of data science and machine learning methods throughout the field. Further, the early lessons drawn from the OKN program (noted above) on the benefits of “bootstrapping” activities for interdisciplinary and multi-sector collaborations would seem to be quite valuable for the wider EarthCube community.

- **Researchers from all domains need assistance and resources on getting started and creating OKNs.** Unless someone is working on a current OKN project, there is confusion about what that is, how it differs from an open knowledge graph, the audience for it, as well as how to organize it technically and in the context of a project.
    - There is also confusion about the term “convergence accelerator”.
    - The CA program could benefit from some basic OKN and other supporting services designed for researchers and supported akin to XSEDE or CloudBank[12], where grants get credits to use in their project and are saved the time of evaluating and procuring services.
    - Regular training is needed to acquire these skills by the research teams. It’s not sustainable to have all the skills in one senior PhD student who will leave eventually.
    - CA could benefit from an even small, backbone organization function.

[12] https://www.cloudbank.org/ National Science Foundation Award #1925001.

- **Convergence research requires overcoming the "Catch-22" of researchers and practitioners in dissimilar fields finding each other and speaking long enough to find commonalities**, and this is a major challenge to achieve convergence across disciplinary boundaries. But because these groups don't use the same terms, they are not likely to find each other via a self-service query tools using their familiar terminologies.
    - Workshops and matchmaking are needed to get people with complimentary backgrounds talking long enough so they see the commonality and the opportunity in their different approaches.
    - The more data is mapped across domains (ontology mapping), the more likely dissimilar but compatible domains can find each other using data resources.
    - Researchers find it invigorating to look at their problem through the eyes of someone who studies it in a totally different or new way. The ones at our workshop wanted more of it, not less.
    - Favorite example was the mobility people realizing they could study ship and port traffic alongside fisheries and plankton 'traffic'.
- **The data science and computational/tool building skills honed by the geoscience community are in demand by CA teams**. At the same time, it can be difficult for teams to welcome new or unknown members to an already risky area of research, especially those that are breaking new ground.
    - One approach would be to develop focused multi-program PI (or project) meetings to facilitate technology transfer. Bringing together representatives from projects funded by NSF's CA, HDR, and EarthCube and other GEO programs would serve to cross-pollinate awareness and uptake of technical tools and resources.

1. Need to develop services that assist with passing back new information, "and corrections we learn from cross-disciplinary work into the original knowledge network that was used." See, for example, the Magnetics Information Consortium (MagIC) Position Statement included here (page 26).

2. It was noted that there is a need for additional training sessions (face-to-face or virtual) for domain scientists on the selection and application of computational methods (e.g., Machine Learning, generating knowledge graphs and networks, etc.).

3. Calls for increasing awareness and capacity in analytical approaches for Open Knowledge resources. Training on the distinctions between graphs and networks, making data 'knowledge graph-ready,' and value for the different kinds of things we can learn through analysis and interpretation.

4. Addition of "Demo Days" (or booths) at relevant domain and CS conferences and events.

5. Provide travel funding for CA (geoscience related) OKN projects that were not selected for awards to participate in EarthCube events and activities.

6. Outcome from 'Convergence Accelerator: lessons learned 'bootstrapping OKNs' which is unique across domains, but with commonalities. How could EarthCube benefit or be affected by these lessons learned?

7. Outcome from Convergence Accelerator: design 'best practices' for collaboration between domain scientists and knowledge representation (e.g. ontology) engineers, much like the domain sciences and machine learning communities are beginning to interact more closely.

Acknowledgements

We thank all the participants for their many substantial contributions to the content and conversation that anchored this meeting. We thank NSF's Convergence Accelerator team, especially Lara Campbell and Chaitan Baru (NSF) for the workshop support, and also thank Chaitan Baru for presenting background and providing context for the meeting. Thank you to Christine Colvis (NIH) for lending her expertise and to the EarthCube Office's NSF Program Officer, Eva Zanzerkia, for her help in making this collaboration possible. This work was supported through the National Science Foundation award #1928208.

References / Sources

National Academies of Sciences, Engineering, and Medicine. 2019. Fostering the Culture of Convergence in Research: Proceedings of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/25271. (Available: https://www.nap.edu/catalog/25271/fostering-the-culture-of-convergence-in-research-proceedings-of-a)

National Science and Technology Council (2018). Open Knowledge Network: Summary of the Big Data Working Group Workshop, Oct. 4-5, 2017, Bethesda, MD. Available: https://www.nitrd.gov/news/Open-Knowledge-Network-Workshop-Report-2018.aspx

National Science Foundation. NSF Convergence Accelerator (C-Accel) Overview. Available: https://www.nsf.gov/od/oia/convergence-accelerator/index.jsp Accessed 4 November 2019.

National Science Foundation. (2019). Dear Colleague Letter: Request for Information on Future Topics for the NSF Convergence Accelerator. NSF 19-065, May 6, 2019. Available: https://www.nsf.gov/pubs/2019/nsf19065/nsf19065.jsp?WT.mc_id=USNSF_179

Appendix A: Agenda

Agenda: Convergence Accelerator-Open Knowledge Network/EarthCube Workshop

Workshop Goals

1. Identify resources and partners that may enhance CA-OKN Phase 2 project proposals.
2. Gather community input for the 2020 NSF CA-OKN solicitation related to geospatial and earth-observation data or otherwise overlapping with the geosciences.
3. Identify thematic areas for a potential geosciences-related Track C in 2021.
4. Facilitate interest in a working group(s) to generate a new geo-related CA proposal.

Time	Item	Goal	Speaker	Mins
8:30	Light breakfast available	Networking and fellowship	-	30
9	Welcome, agenda, workshop goals	- Start us with a shared understanding of the day's purpose and goals - Show breadth of projects, use cases, and data - Familiarize participants with CDF, CFP, office projects / capabilities	Christine Kirkpatrick	15
9:15	CA/OKN Overview	Level set knowledge about CA project, OKN track, where we are in the cycle and how EarthCube community is/isn't represented yet	Chaitan Baru / Lara Campbell	15
9:30	Information science perspective on semantics and knowledge networks	Bring perspective on underlying methodology used by knowledge networks	Jane Greenberg	30
10	Lessons from implementing knowledge graphs in open science	- Common understanding and level setting of knowledge networks - Computer science perspective on OKN technologies and maturity	Peter Fox	30
10:30	Intros and Ice breaker	Provide opportunity for projects to find affinities, and to look for common goals	Christine K.	20
10:50	Coffee Break	Networking		20

Time	Item	Goal	Speaker	Mins
11:10	OKN Lightning talks - SPOKE, Sergio Baranzini - SDS, Mike Flaxman / Xinyue Ye - KONQUER, Ilya Zaslavsky	Gain understanding of OKN example projects and understand potential places for EarthCube community to add value/partner	Moderator: Melissa Cragin	25
11:35	EarthCube Lightning Talks - Strabospot, Doug Walker - OpenTopo, Chris Crosby - CDF, Danie Kincade	Show breadth of EarthCube activity, especially those with potential to partner with OKN projects	Moderator: Melissa Cragin	25
12	Lunch	Networking and fellowship		30
12:30	Biomedical Data Translator	NIH perspective on OKN research	Christine Colvis	30
1	Knowledge Graph Demonstration	Introduction to knowledge graph tools	Anirudh Prabhu	30
1:30	Affinity group breakouts: Opportunities and gaps for projects	Find commonality between represented OKN and EC projects	Christine K.	60
2:30	Report Out & Discussion	Start to see potential collaborations	Christine K.	30
3	Break	Refocus		10
3:10	Small group breakouts: Feedback to NSF	Explore questions for NSF solicitation feedback	Christine K.	30
3:40	Report Out & Discussion	Highlight novel insights and epiphanies. Give NSF a chance to ask clarifying questions	Christine K.	15
3:55	Closing thoughts, next steps on report activity, thanks	Set expectations for follow up and opportunities to give additional report input	Christine K. / Lara / Chaitan	5
4	Adjourn			

Appendix B: Participant List

Registered:

Sibel Adali, *Rensselaer Polytechnic Institutes* sibel@cs.rpi.edu
Alyssa Arce, *SDSC / UCSD* alyssa@sdsc.edu
Sergio Baranzini, *UCSF* sergio.baranzini@ucsf.edu
Chaitan Baru, *National Science Foundation* cbaru@nsf.gov
Andong Chen, *Univ. of Michigan* andongch@umich.edu
Christine Colvis, *NIH* christine.colvis@nih.gov
Melissa Cragin, *SDSC / UCSD* mcragin@sdsc.edu
Christopher Crosby, *UNAVCO / OpenTopography* crosby@unavco.org
Enrique Curchitser, *Rutgers Univ.* enrique@marine.rutgers.edu
İbrahim Demir, *University of Iowa* ibrahim-demir@uiowa.edu
Doug Fils, *Ocean Leadership* dfils@oceanleadership.org
Michael Flaxman, *Geodesign Technologies* mflaxman@geodesigntech.com
Peter Fox, *Rensselaer Polytechnic Institute* pfox@cs.rpi.edu
Johannes Friedrich, *World Resources Institute* johannes.friedrich@wri.org
Jane Greenberg, *Drexel Univ.* jg3243@drexel.edu
Sharat Israni, *UCSF* sharat.israni@ucsf.edu
Nicholas Jarboe, *SIO / UCSD* njarboe@ucsd.edu
Alexander Kerney, *Gulf of Maine Research Institute* akerney@gmri.org
Danie Kinkade, *Woods Hole Oceanographic Institution* dkinkade@whoi.edu
Christine Kirkpatrick, *SDSC UCSD* christine@sdsc.edu
Rebecca Koskela, *EarthCube Office (ECO)* rkoskela@gmail.com
Chengkai Li, *Univ. of Texas at Arlington* cli@uta.edu
Lewis McGibbney, *NASA Jet Propulsion Laboratory* lewis.j.mcgibbney@jpl.nasa.gov
Ryan McGranaghan, *Atmosphere and Space Tech. Research* rmcgranaghan@astraspace.net
Kenton McHenry, *Univ. of Illinois Urbana-Champaign* mchenry@illinois.edu
Charles Meertens, *UNAVCO* meertens@unavco.org
Aditi Misra, *Univ. of Michigan* aditimis@umich.edu
Tom Narock, *Goucher College* thomas.narock@goucher.edu
Mark Parsons, *RPI* parsom3@rpi.edu
Mark Schildhauer, *Arctic Data Center* schild@nceas.ucsb.edu
Martin Seul, *CUAHSI* mseul@cuahsi.org
Adam Shepherd, *BCO-DMO / WHOI* ashepherd@whoi.edu
Noel Southall, *NIH* southalln@mail.nih.gov
Pingbo Tang, *Carnegie Mellon Univ.* ptang@andrew.cmu.edu
Chad Trabant, *IRIS Data Services* chad@iris.washington.edu
Doug Walker, *Univ. of Kansas* jdwalker@ku.edu

Shaohua Wang, *New Jersey Institute of Technology* davidsw@gmail.com
Yinghui Wu, *Case Western Reserve Univ.* yxw1650@case.edu
Xinyue Ye, *New Jersey Institute of Technology* xye@njit.edu
Riley Young Morse, *Gulf of Maine Research Institute* rmorse@gmri.org
Ilya Zaslavsky, *UCSD* zaslavsk@sdsc.edu

Joined from ESIP Winter Meeting:
Rahul Ramachandran, NASA

Joined from the ESIP GeoSemantics Symposium:
Ruth Duerr, *Ronin Institute*
Mohan Ramamurthy, *Unidata*
Gary Berg-Cross, *Ontolog Forum*
Ted Habermann, *Metadata Game Changers*
Steve Richard, *US Geoscience Information Network*
Howard Burrows, *Library Innovation Center*

APPENDIX C: Position Papers

Abstracts begin on the following page. (8) position statements were submitted ahead of the workshop, contributing to our understanding on the breadth of research targets in Earth, Space, & Atmospheric sciences, as well as current approaches to developing Open Knowledge Network resources for several grand challenge topics. These papers are arranged alphabetically by last name of the first author.

Intelligent Disaster Analytics and Communication Cyberinfrastructure

PI: İbrahim Demir, University of Iowa
Collaborators: Yusuf Sermet, Muhammed Sit, Enes Yildirim, Zhongrun Xiang

https://hydroinformatics.uiowa.edu/ - ibrahim-demir@uiowa.edu

Team Expertise

I am an assistant professor at the Civil and Environmental Engineering, and Electrical and Computer Engineering departments at the University of Iowa (UI). I am the director of the UI Hydroinformatics Lab with over 15 researchers and students. Our research focuses on hydroinformatics, environmental information systems, scientific visualization, big data analytics, intelligent systems, scientific computing, and information communication. Our research projects include applications of intelligent systems, smart assistants for disasters, crowdsourced augmented reality applications for environmental monitoring, virtual reality, and cyberlearning systems for hydrological simulations, and holographic applications for emergency management. We have developed and actively managed many end-to-end research and operational information and cyber systems for water resources and disaster preparedness and response including Iowa Flood Information System, Iowa Water Quality Information System, Upper Mississippi Information Systems, and many others.

Thematic Area for Track C 2021

Hurricanes Harvey, Irma and Maria, and other natural disasters caused more than $306 billion in 2017 and marked the costliest year for natural disasters in the US. It is estimated that more than 250 million people are affected by natural disasters worldwide every year. The National Research Council highlights the importance of data for a resilient community and reports the challenges and limitations of conventional sensing technologies and modeling efforts[13] and describes the need for next-generation solutions for monitoring, prediction, and communication of data and information.

We like to propose a thematic area on 'Intelligent Disaster Analytics and Communication Cyberinfrastructure'. The proposed cyber frameworks will provide multi-level (country,

[13] Duffy, C., Gil, Y., Deelman, E., Marru, S., Pierce, M., Demir, I. and Wiener, G., 2012. Designing a road map for geoscience workflows. Eos, Transactions American Geophysical Union, 93(24), pp.225-226.

watershed, property), personalized risk analysis[14] and predictions[15] to support real-time (within hours) decision making and long term (30 years) planning to reduce disaster-related costs and save lives. The end-to-end cyber systems will require tackling many big data challenges for crowdsourced Internet of Thing (IoT) sensor networks[16], processing terabytes of data per day, creating novel data structures[17] and algorithms, implementation of data-driven modeling[18] and predictions using machine learning[19], communicating risk[20] and insights using intelligent systems[21] (NLP, OKN), and educating the data science workforce. Novel artificial intelligence techniques can be used to improve the spatial and temporal resolution of the environmental datasets, enhance risk analysis and prediction, and communicate disaster risk and information using intelligent systems. Some of the preliminary work from our projects that can support this focus area include a) an intelligent system for flooding using natural language processing (Siri for Flooding), b) holographic flood risk communication using HoloLens, c) serious gaming for participatory decision making, d) crowdsourcing environmental measurements using smartphone sensors, and e) distributed computing and GPU acceleration on web-based system for large scale data analysis and visualization.

---

[14] Yildirim, E. and Demir, I., 2019. An integrated web framework for HAZUS-MH flood loss estimation analysis. Natural Hazards, 99(1), pp.275-286.

[15] Xiang, Z., Yan, J. and Demir, I., 2020. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water resources research, 56(1), p.e2019WR025326.

[16] Sermet, Y., Villanueva, P., Sit, M.A. and Demir, I., 2020. Crowdsourced approaches for stage measurements at ungauged locations using smartphones. Hydrological Sciences Journal, 65(5), pp.813-822.

[17] Sermet, Y. and Demir, I., 2019. Towards an information centric flood ontology for information management and communication. Earth Science Informatics, 12(4), pp.541-551.

[18] Ebert-Uphoff, I., Thompson, D.R., Demir, I., Gel, Y.R., Karpatne, A., Guereque, M., Kumar, V., Cabral-Cano, E. and Smyth, P., 2017, September. A vision for the development of benchmarks to bridge geoscience and data science. In 17th International Workshop on Climate Informatics.

[19] Sit, M. and Demir, I., 2019. Decentralized flood forecasting using deep neural networks. arXiv preprint arXiv:1902.02308.

[20] Sermet, Y. and Demir, I., 2019. Flood action VR: a virtual reality framework for disaster awareness and emergency response training. In ACM SIGGRAPH 2019 Posters (pp. 1-2).

[21] Sermet, Y. and Demir, I., 2018. An intelligent system on knowledge generation and communication about flooding. Environmental modelling & software, 108, pp.51-60.

Potential Synergies between the Civil Infrastructure Systems Open Knowledge Network (CIS-OKN) and the EarthCube

NSF C-Accel Team#: A7115
PI: Nora El-Gohary
Co-PIs: Arnab Chakraborty, Lucio Soibelman, ChengXiang Zhai
SP: Ximing Cai, Jong Lee, Yung-Hsiang Lu, Sami Masri, Yanfeng Ouyang, Pingbo Tang, Jiansong Zhang, Yanchao Zhang
Attendee/Presenter: Pingbo Tang

The nation's aging civil infrastructure is deteriorating at an alarming rate, with overwhelming transportation, power, water, and pipeline system failures. Currently, "over 240,000 water main breaks occur each year in the U.S.", "over 188 million trips are taken daily across deficient bridges", and "21% of the nation's highways are in poor condition"[122]. Overall, America's infrastructure received an overall grade of D+ (poor), across 16 infrastructure categories, in the American Society of Civil Engineers (ASCE)'s report card[1], with a need for a $2 Trillion to close a 10-year investment gap in the nation's infrastructure. This infrastructure condition "has a cascading impact on our nation's economy, impacting business productivity, gross domestic product (GDP), employment, personal income, and international competitiveness"[1]. Improving the civil infrastructure is a grand societal challenge that can no longer be ignored.

Recent advances in data analytics and machine learning have created a unique opportunity to solve this grand societal challenge by learning from past and current conditions (captured in various data sources) to better assess and predict the conditions and resilience of the transportation infrastructure systems, derive insights about maintenance and investment decision making, and better assess and predict the socioeconomic impacts of civil infrastructure operation and maintenance decisions on a community's or region's accessibility, economic health, and vulnerability. The CIS-OKN will provide the foundation necessary to advance the understanding of these interdependencies and to help develop shared knowledge among the different types of stakeholders. The CIS-OKN will be a shared cyberinfrastructure that can (1) extract information from both structured and unstructured sources, (2) connect the scattered data, information, and knowledge from disparate sources and in heterogeneous formats, and (3) enable deeper interdisciplinary analysis of connected data/information/knowledge. It will enable access to critical data/information/knowledge and analytics of infrastructure transportation, and socioeconomic datasets. The CIS-OKN will also include pertinent

---

[22] American Society of Civil Engineers (ASCE). Report Card for American's Infrastructure. http://www.infrastructurereportcard.org/.

spatially-referenced data given that the structural and transportation analyses and the related socioeconomic impacts are geospatial by nature.

The potential synergies between the CIS-OKN and the EarthCube efforts are therefore numerous, which could be summarized along two dimensions:

- Vertical synergy: There is an important opportunity to collaborate on geoscience data that are very relevant to the CIS-OKN analytics, such as GIS, meteorological, structural geology, and hydrological data. The CIS-OKN project could support the linking, querying, and reasoning about geoscience data; and could also serve as an important application area for datasets developed under the EarthCube program. For example, the deterioration trends of bridges and roads in different environments can vary significantly. The meteorological structural geology, and hydrological data can help capture the spatiotemporal correlations between the conditions of transportation infrastructure and environmental conditions, such as weather, subsurface geological conditions, and water flows that influence the erosion and scouring of physical structures. These spatiotemporal correlations between civil infrastructure conditions and environmental conditions have the potential of augmenting deterioration prediction methods for more reliable maintenance planning of bridges and roads. Also, the socioeconomic data collected and organized for the CIS-OKN project has the potential to enrich the GIS platform that supports the EarthCube program and could bring unique research opportunities for discovering how natural, technical, and civil infrastructure systems influence each other to understand and predict a complex and evolving solid Earth. On the other hand, the GIS platform can help achieve more reliable sharing and visualization of socioeconomic data for comprehending the socioeconomic impacts of the operation and maintenance decisions of a network of transportation infrastructure.
- Horizontal synergy: The CIS-OKN project could also collaborate on solving important horizontal challenges. Like the CIS-OKN project, the research supported through the EarthCube will require public data infrastructures that address important issues such as information extraction, data linking and fusion, data security, data equality and trust, and scalability. For example, the knowledge graph, as well as the intelligent data linking, fusion, and retrieval methods developed for the CIS-OKN project has the potential to enable automatic data linking and information retrieval from large amounts of geoscience data. The CIS-OKN project will examine how raw data quality, defective data transmission and information processes, and data sharing protocols influence the reliability of the retrieved information and related decisions. Similar data quality and trust issues could also exist in integrated geoscience data analysis. The EarthCube program has accumulated long-term experiences of data sharing and has identified

challenges related to data sharing and cybersecurity, which could cross-fertilize the cybersecurity studies for ensuring civil infrastructure security that has both the physical safety element and the cybersecurity element.

ESSO Project 418/419

Doug Fils | Ocean Leadership

Semantic technologies have continued to grow and impact day to day consumer searches with many interactions on the web augmented if not driven by knowledge graphs. The advent of the Google Data Set Search is an example of a large commercial entity leveraging web and semantic standards to develop a knowledge graph to improve search for data in more research focus.

The funding of Project 418/419, in which this author was involved, is a tiny example of where this can impact the research community. To be clear this work was not heavily science driven but rather focused on the data facilities and other providers of data to the scientific community. Driven more by FAIR data, Belmont Forum, CODATA and RDA developments than by direct science goals. Not that science goals don't exist that could be integrated, just that these were not our initial source for drivers of this work.

Leveraging the EarthCube Council of Data Facilities and the ESIP semantic community has resulted in developments in vocabularies, tools and validation methods. Examples of these include;

[1] https://gleaner.io/

[2] https://github.com/earthcubearchitecture-project418/CDFSemanticNetwork

[3] https://github.com/ESIPFed/science-on-schema.org

The goal of these is to help providers with FAIR data practices, semantic methods and web architecture to improve the discovery and use of data resources across a broad range of geoscience communities. These are basic and simple products from the EarthCube work but are still very much active and developing interest that the project members would enjoy continuing.

This work did highlight a few challenges such as;

- Need for advocating more use of persistent identifiers
- Guidance with vocabulary selection and providing implementation examples
- Approaches to validation such as SHACL
- Identifying gaps in provider and consumer tooling
- Automated methods to identify and link named entities / keywords to existing resolvable identifiers for terms
- Need for better spatial and temporal implementation patterns for providers (to help address more specific science drivers for example)

There is a strong foundation in this and other projects of EarthCube and the broader geoinformatics community for further developments. An environment such as the Convergence Accelerators has the potential to foster creative connections and tools from that foundation. Existing entities like EarthCube and ESIP (and ESIP Labs) provide structure for governance and community interactions that would be useful in such work.

Implementation of Open-World, Integrative, Transparent, Collaborative Science Data Platforms Using Open Knowledge Networks

Peter Fox

Rensselaer Polytechnic Institute, Tetherless World Constellation, Troy, NY, United States and Deep Carbon Observatory (DCO) Data Science Team

Abstract Text

Ending in December 2019, for ten years the Deep Carbon Observatory (DCO) has revolutionized scientific insights into the quantities, movements, forms and origins of carbon in deep Earth. Participants in DCO come from over 40 countries, 400 institutions, representing close to 1800 researchers, perhaps 15% of which are considered early career scientists. In amongst all DCO's achievements and legacy products, monographs, special journal issues, films, blogs, field sites, instrumentation, databases, when surveyed the DCO community rated the DCO science network itself to be the biggest, impactful and likely the most lasting legacy of DCO. In 2011, with some foresight and a little good fortune, the DCO Data Science team envisioned a way to both represent the fledgling DCO science network and begin to implement an open data science platform that both captured essential linkages in the DCO network and facilitated collaboration, auto-generation of web page content, project reporting and more. Also around 2010 the first indications that semantic technologies, were maturing enough that semantic application integration (cf. data integration) was possible using conceptual information model integration and ontology engineering. The DCO ontology was born, as was the DCO knowledge graph. When rendered in various forms that graph was the DCO knowledge (science) network. This presentation briefly chronicles the origins, design concepts, modeling, and technical infrastructure that led to an evolving knowledge network (that resides within a publicly available SPARQL endpoint). Conclusions and perspectives are offered for emerging open knowledge network development, implementation and maintenance based on our 8+ years of DCO experience.

Abstract Citation

Fox, P. (2019), Implementation of Open-World, Integrative, Transparent, Collaborative Science Data Platforms Using Open Knowledge Networks, Abstract IN22C-09 presented at 2019 Fall Meeting, AGU, San Francisco, CA, 9-13 Dec.

Scalable Precision Medicine Oriented Knowledge Engine (SPOKE): an OKN for Precision Medicine

Sharat Israni, Sergio Baranzini | University of California, San Francisco

Position Statement_A7160 (SPOKE)

Our team hypothesizes that the massive amounts of data and information related to biomedicine generated over the last century are siloed by artificial disciplinary boundaries (e.g. cellular, physiological, molecular, etc.). We further hypothesize that computers will be more efficient than even trained humans (e.g. doctors) at integrating all digitally available information to assist in decision-making and generation of new knowledge. Specifically, this proposal is concerned with the development of a multi-domain, multi-scale heterogeneous network composed of biomedical data and information that will enable the execution of complex queries and the emergence of new knowledge by convergence of seemingly disparate datasets. A key component of our lives is geolocation data, and even our phones and watches are tracking us 24/7. We propose to integrate geospatial information into SPOKE, such as UV radiation maps, ozone, particulate matter, air quality, etc. This information, together with clinical and demographics already available to us, will result in a comprehensive and realistic assessment of health throughout the country allowing more specific recommendations, treatments and personalized care.

Our Core team at UCSF has developed SPOKE, a heterogeneous network composed of a few million nodes (biomedical concepts) and edges (well-established relationships among them) that is proving useful to prioritize drugs for repurposing and prediction of yet undiscovered genetic associations. Once a full team has been assembled, we propose to group SPOKE from millions to billions of nodes (and trillions of edges) to increase the reach of topics it currently harbors, and therefore, its potential impact. Additional teams at UCSF will provide expertise in drug development, cancer biology and personalized approaches to practicing medicine. At the same time, an advanced development team at Google Inc. has created DataCommons, an open platform to identify and integrate datasets all over the world wide web. Further capitalizing on the biomedical expertise and advanced multi-scale phenotyping capabilities of the Institute for Systems Biology (Seattle) and the computational and analytical prowess of the Lawrence Livermore National Laboratory, we propose to apply a truly convergent approach to accelerate discoveries and potentially transform the way in which modern medicine is practiced.

We anticipate that massive public (both expert and non-expert) acceptance and utilization of this OKN will result in significant societal gains. First, we anticipate that the scientific community will benefit from seamless access to (de-identified) multi-domain integrated biomedical data and information in the same way that access to single-domain information (e.g. genomes, drug targets, disease associations) has fueled discoveries in genetics, new therapeutic drugs, and basic biology. But we also envision “citizen scientists” taking advantage of such resources to test new hypotheses, disprove current paradigms, and even contribute new knowledge. This role could be assumed by literally any member of society, from STEM-oriented teenagers, to educated persons from other disciplines, to the general public.

Magnetics Information Consortium (MagIC) Position Statement

Nicholas Jarboe | Scripps Institution of Oceanography, University of California, San Diego

The Magnetics Information Consortium (MagIC) data repository (earthref.org/MagIC) supports the rock magnetic, geomagnetic, and paleomagnetic communities by providing data archiving, data retrieval, searching, data analysis, and plotting tools. We implement FAIR data practices and work closely with publishers and authors to archive and expose data to the public in a timely and complete manner. We have been working with the EarthCube P419/GeoCodes project to make data in our repository more easily discoverable and interoperable with the inclusion of schema.org/JSON-LD meta data and the development of schema.org extensions to include geoscience specific terms.

One high impact use of open knowledge graphs/networks in the Geosciences would be the establishment of a system that can inform other nodes when the geologic age of a sample or geologic zone/horizon has been changed. Age changes can be due to a large number of factors and having a robust way for systems to query back/receive updates on the geologic age of a piece of data would greatly improve most deep time geoscience projects: climate change investigations, plate tectonic reconstructions, oil and gas source modeling, geologic mapping, etc. The paleomagnetic data in our repository all have geologic ages associated with them and we are actively working with and seeking out groups to help develop systems where those ages can be both exposed to other systems and updated from the original age sources. The EarthCube Sparrow project seems to be especially promising as a start to a system where updated age data could be available. We would be interested in working with other groups that are interested in this problem and feel that a geoscience Convergence Accelerator solicitation could be a successful one. Many private companies and the USGS are likely to be interested in collaborating with a project on automatically updating the accuracy and precision of geologic ages.

Biological and Chemical Oceanography Data Management Office (BCO-DMO)

Adam Shepherd | Woods Hole Oceanographic Institution

Position Statement

I serve as the Technical Director of the Biological and Chemical Oceanography Data Management Office (BCO-DMO), an NSF-funded data repository and Council of Data Facilities (CDF) member. BCO-DMO's current NSF grant proposed to implement knowledge representation techniques such as knowledge graphs as a mechanism for improved sustainability and extensibility of data repositories. Through this effort, all BCO-DMO dataset metadata are described using RDF resources which could be re-used by other knowledge graphs initiatives in the geosciences. These resources become relevant to existing C-Accel funded projects such as the spatial and urban flooding open knowledge networks. BCO-DMO publishes spatial representations for dataset collection points and research cruise trajectories. The repository also includes coastal flood inundation data that could provide environmental contexts to the urban flooding initiative.

As co-lead on the EarthCube Project 418 and 419, I developed the guidelines for CDF members to publish their dataset metadata following graph techniques, in this case using the schema.org vocabulary. Through this project, our team harvested graph data from over a dozen NSF data repositories and projects providing insight into the challenges of harvesting, storing and linking graph data in the geosciences. This work also led to identifying the needs within the geosciences to accurately describe science datasets within their domain of practice - including the use of domain vocabularies, challenges in representation of temporal and spatial coverages, and access to data services and APIs.

As a project member on the Urban Flooding Open Knowledge Network C-Accel project, I lead the prototype development effort that combines persona identification, persona-driven use case development, development of conceptual and logical models, and the harmonizing of multiple datasets including flooding model results, roadways, power grids, elevations, urban developments and flood predictions into a comprehensive knowledge graph.

Open Knowledge Network on Telecoupling Earth System (OKN-TES)

Xinyue Ye, Associate Professor | College of Computing, New Jersey Institute of Technology

The Earth system is increasingly connected through **distant interactions**, such as international trade, foreign direct investment, tourism, species invasion, migration, information dissemination, technology transfer, teleconnection, and water transfer. Previous research mainly focused on a **single** distant interaction and its socioeconomic or environmental impacts **separately**. However, distant interactions may be **interrelated** (e.g., enhance or offset each other) and have profound implications for global and national challenges such as climate change, environment, land use change, biodiversity, and human well-being.

To understand various distant interactions and their interrelationships across spatial and temporal scales, the concept of telecoupling and its award-winning framework have been developed to encompass all kinds of distant interactions. To quantify the effects of telecouplings, there is an urgent need for big data and for innovative analytic tools. Telecoupled systems are very complex and highly dynamic, as telecouplings can affect and be affected by a wide array of earth system factors, e.g., socioeconomic, demographic, cultural, climatic, geographic, ecological, and environmental dimensions. At present, the data sources are diverse, ranging from population census, interviews, surveys, field collection, to remote sensing. To gain deeper insights and uncover hidden patterns in the earth system, we need a platform to facilitate the creation, verification, sharing, and reusing of telecoupling research and knowledge convergence.

We propose to develop an innovative **Open Knowledge Network on Telecoupling Earth System (OKN-TES)** to transform telecoupling research and create novel solutions that address nationally important socioeconomic and environmental challenges and opportunities associated with telecouplings. As a common technological infrastructure, it will be **horizontal** to **address all telecouplings** such as international trade, tourism, and species invasion. Complex telecouplings and feedbacks across multiple systems and scales pose barriers to the understanding of telecoupling impacts. Through merging and empowering ideas, data, information, approaches and technologies from diverse fields of knowledge, it will provide a powerful pathway to lower/remove those barriers while changing research/education paradigms and broadening participation in telecoupling science and engineering.

Significance: (1) **OKN-TES** will be an **open platform** that links heterogeneous datasets, tools, and services as a nonproprietary shared knowledge network infrastructure for addressing various telecouplings (e.g., trade, tourism, and species invasion). Every resource will be openly accessible, easily edited, reused, shared, verified by others; and can be combined with other resources to make new resources. More importantly, (2) **OKN-TES** will advocate **reusable and reproducible** telecoupling research, including data, analytical procedures, modeling outcomes, data visualization and analytics tools, and many other resources. (3) Through constant information updates, the platform will be **dynamic** in nature to reflect dynamics of real-world complex systems. (4) The **applications** of the platform will uncover socioeconomic and environmental impacts of various telecouplings and provide innovative telecoupling solutions to help make progress towards many of the 17 UN Sustainable Development Goals that the US and 192 other countries have adopted.

APPENDIX D: Presentations

Baru, Chaitan. (2020). The NSF Convergence Accelerator Pilot: Track A – Open Knowledge Network. Talk for the EarthCube Convergence Accelerator-Open Knowledge Networks meeting. January 6, 2020. N. Bethesda, MD. Available: https://bit.ly/2Wktah9

Fox, Peter. (2020). [Open-World, Integrative, Transparent, Collaborative Science Data Platforms] Using Open Knowledge Networks. Talk for the EarthCube Convergence Accelerator-Open Knowledge Networks meeting. January 6, 2020. N. Bethesda, MD. Available: https://bit.ly/3etBdyJ

Greenberg, Jane. (2020). Information Science: Perspective on Underlying Methodology of Knowledge Networks. Talk for the EarthCube Convergence Accelerator-Open Knowledge Networks meeting. January 6, 2020. N. Bethesda, MD. Available: https://bit.ly/2DMr7fD

Kirkpatrick, Christine. (2020). CA-OKN EarthCube Workshop. Talk for the EarthCube Convergence Accelerator-Open Knowledge Networks meeting. January 6, 2020. N. Bethesda, MD. Available: https://bit.ly/2Wpc81v

Prabhad, Anirudh. (2020). Insights from Knowledge Graphs. Talk for the EarthCube Convergence Accelerator-Open Knowledge Networks meeting. January 6, 2020. N. Bethesda, MD. Available: https://www.slideshare.net/AnirudhPrabhu/insights-from-knowledge-graphs

Image credit: James Kurose, NSF

Earthcube.org

National Science Foundation Award #1928208