CollabNext: A Person-Focused Metafabric for Open Knowledge Networks

Lew Lefton, Georgia Institute of Technology; Didier Contis, Kexin Rong

Use case description and societal challenge being addressed

The CollabNext tool originated as a partnership between Georgia Tech and the Atlanta University Center, and is now being developed jointly by Fisk University, Georgia Tech, Morehouse College, Texas Southern University, and University at Buffalo. Our goal is to develop a knowledge graph based on people, organizations, and research topics. We are adopting an intentional design approach which initially prioritizes HBCUs and emerging researchers in a deliberate effort to counterbalance the Matthew effect, a naturally accumulated advantage of well-resourced research organizations. Our goal is to utilize open science data sources with well-established persistent identifiers, maintain ethical data management (e.g. FAIR and CARE principles), follow human-centered design principles, and leverage state-of-the-art algorithms. Our current proof of concept version of the CollabNext tool is very rough and we know there are bugs, and data issues, which are to be expected at this early stage. We welcome feedback! Over the course of the Proto-OKN project we will move from alpha to beta to production versions with guidance from our Advisory Group.

We will build a person-focused tool that will help identify existing and potentially new research partnerships and help users answer basic questions such as “who is working on what and where?”. Our design process is driven by dozens of user stories, which include end users who have a variety of roles, including researchers (current and prospective PIs), campus leadership, research administration professionals, graduate and undergraduate students, conference organizers, research sponsors, and industry partners. By bringing greater visibility to what and who is often rendered invisible in the current science system, CollabNext will facilitate research collaborations and illuminate the broader research landscape.

Knowledge graph source datasets

We currently use data from OpenAlex (formerly Microsoft Academic Graph) as our initial schema for people, topics, and organizations. We will also include the Center for Measuring University Performance, and have plans to incorporate other open data sources (eg Patents, Grants, Dissertations). We will prioritize interoperability of our graph database with other Proto-OKN projects. Based on initial estimates (using SemOpenAlex, which is a project that generates RDF triple stores from OpenAlex data), we anticipate that our knowledge graph will be on the order of 30-40 billion triples.

User queries / competency queries for the use case

- As a principal investigator (or industry partner), I want to identify and contact colleagues (*either at HBCUs/MSIs/emerging research institutions or at well-resourced institutions*) with interest and expertise in specific research areas, so that I can build a stronger research team.
- As a sponsoring agency program officer (or journal editor), I want to identify researchers who specialize in certain areas to serve as reviewers so I can expose more researchers to what successful submissions look like.
- As a program or conference organizer, I want to curate a diverse panel of knowledgeable experts so that my event will engage a broader audience and have wider representation.
- As a student applying to a grad program, I want to identify researchers with whom my interests and values align, so that I can find potential advisors at an institution.
- As a graduate student or postdoc, I want to look for potential research collaborators within and outside my institution, so that I can strengthen my research network beyond my advisor.
- As an administrator, I want to understand the current research capabilities and focus areas of faculty in my unit, so that I can visualize existing research networks, and see potential new collaborations, so that I can facilitate partnerships and advocate for additional resources.