NSF

CONVERGENCE

ACCELERATOR

2019
COHORT
GUIDE

The NSF Convergence Accelerator (CA) is a new organizational structure to accelerate the transition of use-inspired convergence research into practice, and build team capacity around exploratory, high-risk projects, initially in three convergence tracks. NSF has created the Convergence Accelerator to leverage all fields of science and engineering to produce practical deliverables in an accelerated timeframe, with an agile project management approach that is well suited for accelerating promising solutions. NSF and the scientific community are using convergence to define a future where researchers from across disciplines work integratively to solve grand challenges of national importance.

The goal of the Convergence Accelerator is to accelerate scientific discovery and innovation, including use-inspired fundamental research, as well as to transition to practice important research results. In particular, the CA will focus on use-inspired, convergence research, with directed deliverables, using an approach that rewards innovation, risk-taking, and transition to practice. NSF will use a team-based approach, with cohorts of teams working together and in competition with one another in a form of coopetition, and with distinct phases of the accelerated research that allow for timely adjustments of resource allocation and direction based on progress.

The Convergence Accelerator is initially focused on two of the NSF Big Ideas— Harnessing the Data Revolution and the Future of Work at the Human-Technology Frontier. What these two ideas have in common is that their research communities have displayed readiness to take some of their best ideas and put them into practice. These tracks also align with Administration R&D Priorities including leadership in artificial intelligence, the President's Management Agenda, and the U.S. 5-Year STEM Education Strategic Plan. NSF performers will partner with industry, non-profits, and others to get access to their practical requirements, challenges, and resources. The accelerator projects will have clear goals, milestones, and deliverables to ensure consistent progress. Program Directors will work with the individual teams within the cohort so that they are cooperating within their track, but also competing to produce the best results possible. While all the teams in a cohort will be focused on a common research goal, they may be pursuing different approaches to achieve it. This tactic will help us stimulate innovation by encouraging as many ideas as possible in pursuit of practical deliverables.

INNOVATION
BEGINS HERE

TABLE OF

CONTENTS

MAIN | Table of Contents

TRACK A
TEAMS

NSF | Convergence Accelerator

TRACK A TEAMS | Table of Contents

TRACK A
VERTICAL
TEAMS

NSF seeks to create an Open Knowledge Network (OKN), which has the potential to drive innovation across all areas of science and engineering and unleash the power of data and artificial intelligence to achieve scientific discovery and economic growth. The overall goal of the Open Knowledge Network track is to **enable the creation of a nonproprietary shared knowledge infrastructure that allows stored data (both structured and unstructured) to be located and its attributes and relationship to other data and to real-world objects and concepts to be understood at a semantic level.** This track will have a particular focus on exploiting publicly available U.S. Government data and similar public datasets. Teams may choose to address "vertical" challenges in specific topic domains such as geosciences, smart health, finance, and manufacturing.

THE CONVERGENCE HUB FOR THE EXPLORATION OF SPACE SCIENCE (CHESS)

Ryan M. McGranaghan
rmcgranaghan@astraspace.net

Bob Arritt	**Jacob Bortnik**
Morris Cohen	**Chris Green**
Joe Hughes	**Adam Kellerman**
Jackson McCormick	**Chigo Ngwira**
Charles Perry	**Karthik Venkataramani**

OVERVIEW

Our power grid is a complex system with many vulnerabilities, one of which is the threat posed by space weather. This is due, in large part, to the lack of data sharing across the multiple disciplines involved—from space scientists, to earth scientists, to power system owners and operators, and power consumers. The Convergence Hub for the Exploration of Space Science (CHESS) is the next-generation approach to bring resilience to the power grid. Our product converges data, individuals, and knowledge, across the space weather-to-power grid divide, providing our utilities with a new tool for situational awareness, prediction, and decision support for their operations.

DESCRIPTION

Three barriers hold back power grid resilience to space weather: (1) the **lack of a cohesive community**, owing to the wide variety of subject matter experts required, (2) the **low-level of effective data sharing**, coordination, and analysis (e.g., data science); and (3) the **diversity of physically dominant processes** in each part of the space weather environment. The CHESS platform curates, links, and provides the diverse data, layers on machine learning predictive models, and communicates at the appropriate level depending on the individual at the power utility. CHESS is like Google Maps for the power grid, putting the power of awareness and prediction in the hands of those who need to keep the lights on for society. The Phase II objective is to move from a prototype to implementation, working directly with >5 utilities across North America. Gaining adoption within utilities' operations is the pathway to sustain and amplify the CHESS product throughout and beyond Phase II. CHESS has the potential to change how we monitor, protect, and make decisions about the power grid, reaching utilities and governmental energy agencies across the US that

have limited information about space weather and are thus unprepared to proactively respond.

DIFFERENTIATORS

The problem utilities face now is connection to the communities that have the knowledge and knowhow about the threats to the power grid, specifically those from space weather. The communication of space-weather hazard has traditionally been through a global index, whereas the impact of space weather on the grid is critically regional and local - this would be like using one number (say global cloud cover) to specify chance of rain in Washington DC. The convergence challenge is that data and individuals from space weather, geosciences, and power grid communities are siloed. CHESS provides the structure, a knowledge network, that unifies the knowledge from the data and communities. Our platform is a data integrator, prediction tool, and social network, enabling radically convergent interactions required to keep the power on for society and the predictive capability to create a more resilient power grid. There have been significant barriers to meaningful interaction between the space weather and the power grid communities, including language, shared evaluation of the value of the tools, and restrictions to changes on the side of utilities (e.g., resources to implement the change). The CHESS team is uniquely composed of scientists, power grid engineers, the public and private sectors, and data scientists with a focus on building trust with utilities that make us unique with respect to previous efforts. We are partnered with the Electric Power Research Institute (EPRI), an institution that has pioneered transdisciplinary interactions for power grid resilience since the 1970s and that coordinates our dedicated user group, the electric utilities participating in the SUNBURST project, and have an expansive stakeholder network structured by explicit connections and trust that offers a direct path to utility

commercialization and operationalization. Through our knowledge network we will build the trust in the CHESS product and across the new community around it to transform power grid resilience at a more rapid pace and on a broader scale than ever before.

ROAD MAP

We organize our approach along three stages (with corresponding deliverables and metrics):

1. Foundation-building (November 2020 - May 2021): Scale our Phase I approach to stakeholder discovery to Phase II. Each new stakeholder and collaborator will be added to the scalable CHESS knowledge network. Deliver all data sets, linked and made *usable* through our solar-to-power grid ontology, to our open prototype platform.

2. From prototype to implementation (May 2021 - July 2022): Transition the prototype into a deployed platform. We will work with 'beta testers' to improve our prototype and increase the maturity to transition to implementation. During this stage we will also improve the predictive models.

3. Positioning the product to bring insight and resiliency to the power grid (November 2021 - November 2022): Link our product to procedures and tools already in use by the power utility community. We will measure success via adoption by power grid utilities and the use of our product to guide future investment.

PARTNERSHIPS

Our core team includes the Atmosphere and Space Technology Research Associates, NASA, the Electric Power Research Institute, Georgia Institute of Technology, and University of California Los Angeles. Our convergent community covers each link in the space weather-to-power grid system, including the NSF, NASA, the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC), the United States Geological Survey (USGS), the Department of Energy (DOE), North American Electric Reliability Corporation, the Federal Energy Regulatory Commission, and numerous utilities and power transmission operators. We benefit from regular and rapid feedback from each domain, each contributing their knowledge to improve the product. From the Convergence Accelerator teams we benefit from rich interactions and expertise with data integration, human-natural systems analysis, knowledge network creation, and interoperable knowledge

networks. The Convergence Accelerator cohort is the portal to broader application of each of our knowledge networks. Our partners are fluidly growing.

INTELLECTUAL PROPERTY

We will protect pre-existing IP, as well as new IP that will be generated, which includes existing power grid analysis tools from EPRI and the Natural Resources Canada and the CHESS product.

BUILDING THE GLOBAL ENERGY DATA COMMONS FOR ASSESSING ENERGY SYSTEM RISKS THROUGH OPEN DATA

Kyle Bradbury
kyle.bradbury@duke.edu

Johannes Friedrich   **Roshanak Nateghi**
**Dylan Harrison-Atlas**   **Steven Rose**

OVERVIEW

Recent wildfires in California and Hurricane Maria in Puerto Rico caused major power losses and other disruptions, highlighting the diverse risks that energy systems face from natural hazards. To help address these risks we are creating the Global Energy Data Commons (GEDC), an interactive online platform and open knowledge network that enables stakeholders to assess exposure of energy infrastructure to natural hazards and informs planning for resilient energy systems.

Through the GEDC, an investor or power company could discover with a few clicks on a map the location of power plants to see how they may be exposed to wildfire, sea level rise, water shortages, and other natural hazards.

DESCRIPTION

Natural hazards affect energy systems around the world, causing disruptions with economic consequences. Power companies, communities, investors, and emergency planners need information to understand and manage these risks but are held back by data barriers. In Phase I, we interviewed energy data users and experts who identified a key issue: the lack of integrated, high resolution, credible, and up-to-date data. For example, data on power plants and grids are not readily available, fragmented across various sources, and are not integrated with natural hazard information.

The primary goal of the GEDC is to give core users (e.g. utilities, investors, asset holders and insurers, system operators, and disaster risk planners) the data to better understand the potential impacts of natural hazards on energy systems such as power plants and power grids, and inform risk assessments and planning for resilient energy systems. Nonprofits, researchers, regional governments, and the public will also be users, able to explore, visualize, and map geographic exposure to hazards.
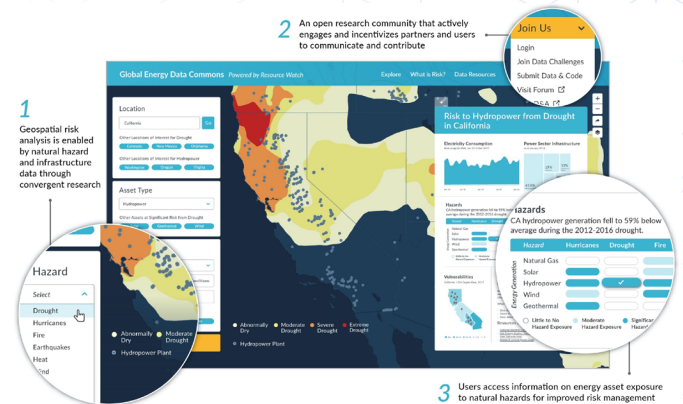
The GEDC platform will collect, integrate, and standardize historical data on natural hazards, such as droughts, ice storms, wildfires, hurricanes, sea level rise, and other extreme events, with energy system data. When data are not openly available, we will create new energy datasets by applying innovative methods like computer vision and natural language processing to extract data from satellite imagery, press releases and other text sources. We will develop a web data visualization platform and provide guidance to apply the data in private sector risk planning processes.

If successful, the GEDC will inform and align academic research priorities to the needs of risk planners and will ultimately improve risk management and strengthen the resilience of our existing and future energy systems.

DIFFERENTIATORS

The GEDC offers three unique value propositions: (1) a user-centered approach informed by interviews with 70 organizations, (2) innovative AI-based data collection methods, cutting edge visualization platform, and guidance for applications of data, and (3) a partnership that brings together leading organizations and methods from computer science, data science, and artificial

intelligence with expertise in energy system operations and planning, engineering, economics, earth science, and climate science.

In Phase I, we surveyed the energy data landscape, conducting interviews with end-users, data providers, and energy and risk experts, and inventoried over 400 existing datasets, data collection methods, and open data platforms. In Phase II, we will draw on insights from these efforts, as well as our collective expertise in energy, data collection, risk, community-building, and platform development, to launch the GEDC. Our aim is to advance understanding of hazard risks to the energy system, beginning with the US electric power sector and expanding to additional countries and globally where possible.

We will build on existing efforts wherever possible, partnering with public and private entities to leverage open resources. Our focus on delivering open source and integrated data will facilitate assessment of risks, incentivize researchers to contribute to a common open knowledge pool, and bring the latest science into the hands of end-users. We will also host public Data Challenges to catalyze new research and data creation.

ROAD MAP

- Quarterly working-groups (start Nov. 2020)

- Build community, digital communications, user outreach (start Nov. 2020)

- Form working groups on risk framing and energy data integration (Nov. 2020)

- Launch beta version of GEDC platform with initial set of natural hazards and energy data and application guidance (April 2021)

- Host Data Challenges (May 2021 - Aug. 2022)

- Publish all computational tools for data collection on GitHub (July 2022)

- Launch production version of the GEDC platform and application guidance (Aug. 2022)

- Enable post-Phase II persistence of the GEDC by leveraging partnerships and commercial licensing models and philanthropic funding

PARTNERSHIPS

Key to the success of the GEDC is the partnership of more than two dozen organizations that have already submitted Letters of Collaboration. In addition to Duke, EPRI, NREL,

Purdue and WRI, we are bringing together big data companies (Google, Amazon, IBM), data standardizers (DBpedia, GEOSS), data providers and risk experts (EPA, NASA, PNNL, UKERC, Battelle), and potential end-users (Goldman Sachs, GE, World Bank, IADB and others). Based on the knowledge network started in Phase I, we plan to expand the network to hundreds of organizations that will inform a global research agenda that facilitates improved risk analysis and risk management.

INTELLECTUAL PROPERTY

The GEDC will provide data with permissive licenses while enabling the continued sustainability of the platform and respecting the licenses of data included in the platform. We anticipate the curation and creation of new datasets on energy infrastructure and their exposure to natural hazards.

AN OPEN KNOWLEDGE NETWORK TO MEET OCEAN DECISION CHALLENGES (OCEANOKN)

Lisa A. Kerr
lkerr@gmri.org

Jon Hare   **Carla Gomes**
Andrew Pershing   **Kevin St. Martin**

Gulf of Maine
Research Institute

OVERVIEW

Throughout history, people have been able to rely on their past experience to inform their decisions about the future. We are now entering a period of rapid change where the past is no longer a reliable guide. There is a critical need for tools to help individuals, communities, government agencies, and businesses make decisions that anticipate changes. The Open Knowledge Network to Meet Ocean Decision Challenges (OceanOKN) will enable a transformative approach to forward-looking, climate-informed decision making in fisheries.

DESCRIPTION

Fisheries are a cornerstone of the U.S. Blue Economy, contributing to economic growth, as well the nation's food security and social fabric of our coastal communities. Fisheries are directly threatened by rapid changes in ocean ecosystems, and this is further exacerbated by the limited use of climate data in fisheries decision making.

The OceanOKN project will accelerate the transformation of ocean data into information products and knowledge that improves ocean resource decision-making in a rapidly changing world. By bringing together climate, ocean, ecosystem, fisheries, and socioeconomic data with cutting edge Artificial Intelligence (AI) and machine learning (ML) techniques, future projections and decision-making for fisheries can be dramatically improved. Through iterative interactions with stakeholders, the OceanOKN team will design products that address key questions and help stakeholders integrate climate considerations into their strategic planning and risk management. We aim to make global climate information actionable at the regional and local scale to support: 1) sustainable fisheries management, 2) seafood business strategy, and 3) coastal community adaptation.

The OceanOKN team will demonstrate the utility of this approach in the Northeast U.S. and build functionality for expansion to other regions and eventually, beyond fisheries. The Northeast U.S. has experienced unprecedented changes in ocean conditions that are already impacting fish populations and challenging the decision-making process in fisheries. We will 1) build a knowledge graph that integrates downscaled climate projections with diverse ocean, fish, and fisheries data, 2) advance application of AI and ML models that use the Ocean OKN for projecting future fish distribution and abundance, and 3) apply the OceanOKN and advanced models to address fisheries decision challenges. This effort will enable us to address critical challenges: *How do we sustainably harvest resources as they move outside historic ranges? How will we maintain profitable fisheries when fish productivity and distribution change? How do we support the resilience of our coastal communities in the face of rapid changes in the ocean and the economy?* The need to shift from backward to forward-looking decision-making involving forecasts and projections transcends the marine and fisheries space. Understanding how individuals and groups of humans make this transformation and the kinds of data, information products, and processes that enable these behavioral changes is a critical question facing society.

DIFFERENTIATORS

Forward-looking decisions require quantitative estimates of future conditions and the ability to project the performance of possible decisions. Thus, models, including AI and deep-learning models, are central to our project and the goal of helping fisheries stakeholders plan for the future. Existing knowledge networks for ocean data, such as NSF's BCO-DMO focus on detailed descriptions of data collected by ships and sensors.

However, these networks do not describe the process of creating data using models. Existing repositories for ocean and climate models record the model output and some documentation of provenance, but they are not connected with observations. Missing are systems the support the act of modeling and integrate across the domains of information required to support complex decisions about resources or economics. Our OceanOKN will implement a knowledge graph that centers on the process and interpretation of models. Our network will be linked with a cloud environment that will facilitate rapid development of interdisciplinary models that integrate climate, ocean, fisheries, and economic data.

ROAD MAP

Timeline	Key Milestone
Nov 2020	Data access and readiness work
Nov 2020	Knowledge graph (KG) and AI-ML model development initiated
Mar 2021	Prototype KG available
May 2021	Environ. Competency Groups (EGCs) launched
May 2021	User-focused design initiated
June 2021	Prototype AI-ML models avail.
June 2021	Prototype use case applications
Nov 2022	KG available for application
Nov 2022	Applied modeling initiated to support decision making
May 2022	User applications avail.
Aug 2022	ECGs and user design completed
Oct 2022	Final products released and applied for decision making

PARTNERSHIPS

Our key partner on this project is NOAA, which is responsible for the largest inventory of ocean data in the nation. NOAA has committed significant staff time (~ 6 months per year) to support this effort. Furthermore, we are engaging a range of decision makers including fisheries managers and advisors, fishing industry, and representatives of the seafood supply chain. These stakeholders routinely make decisions and investments that have multi-million dollar consequences.

INTELLECTUAL PROPERTY

Our project team is committed to open source software development. We will be working in part with confidential data collected from the fishing industry by NOAA and seafood businesses. Models built from these data will be released, but the data will remain confidential. All other data and models will be publicly available. Commercial applications of our OKN by our external partners and others will be encouraged.

DESIGN AND MANUFACTURING GRAPH-AS-A-SERVICE (PROJECT MOKN)

Binil Starly
bstarly@ncsu.edu

Daniele Panozzo
Shivakumar Raman
Farhad Ameri

OVERVIEW

More than 75% of all US manufacturing is carried out by small manufacturing companies with fewer than 20 employees. With razor thin profit margins, the question arises - how can technology better assist the discovery of these services by their clients? Currently, these small manufacturers rely on human network relationships to find new business, which in this digital era limits access to diversified clients. This project aims to leverage the power of knowledge networks to enable the search and discovery of these small manufacturing firms. To promote discovery of services, the open manufacturing knowledge network will democratize access to design knowledge, enabling students, startup entrepreneurs, amateurs and non-technical professionals to find such manufacturing services through the product designs they have produced. Our mission - Build an Open Knowledge Graph in Manufacturing to structure the world's public information on product design and manufacturing.

DESCRIPTION

Consider this very relevant scenario, a doctor in Minnesota designs a custom ventilator comprised of standard and custom designed parts. She might manually go through product catalogs, find related 3D models from model repositories, and utilize manual web forms to solicit companies willing to make the custom parts. But what if there was a faster and simpler way? What if, she could simply take her Bill of Materials (BOM) and automatically search for fabrication companies who are capable and willing to make it? Alternatively, what if she could say — "MOKN, list sheet metal fabrication companies who are within a 100 mile radius and serve the medical device industry?" Project MOKN's broader impact is to make information available on sourcing critical part components, instantaneous gathering of specific manufacturing capabilities, location of those services and availability of resources. The global pandemic crisis serves as a contextual example as to the value of this knowledge — with implications to public health and national security. Accessibility also empowers rural and sub-urban communities dependent on manufacturing services.

The merit of our work is the network graph linking 3D engineering model information combined with the visual relationship between parts, underlying semantic meta-data, and manufacturing services. The team will build algorithms to make design datasets to be findable, accessible and reusable and to connect them with manufacturing information available to shorten the product development life-cycle. Geometric processing algorithms, geometric deep neural networks and natural language processing with novel interfaces will be developed to analyze 3D model information and domain specific engineering text information that will incrementally link diverse pieces of product design, manufacturing process and manufacturing service information. The knowledge graph will result in new natural language query interfaces and search tools for users to interact with the manufacturing graph.

The initial target market is Fabricated Metal Product Industry (NAICS 332, ~40,000 US companies). In Phase-III, technical work will expand to broader sectors of the manufacturing industry. Potential revenue streams through licensing include 1) Manufacturing specific search engine optimization (SEO) premium tools (starting at $99 per year); and 2) Search within 3D model Galleries (@10c per unique model).

DIFFERENTIATORS

Our key differentiators are: 1) Superior information recommending part designs and services, while current solutions only offer text based keyword search for 3D

models or manufacturing services. 2) Partnerships with 3 of the 5 major 3D model product galleries resulting in a head-start to delivering solutions to these user groups (more than 100M parts). 3) No SEO plugins offer manufacturing specific vocabularies and data markup. In a potential Phase-III, Project MOKN will rely on rapid solution roll-out and scaling to connect the thousands of manufacturing services and the millions of 3D engineering product models available publicly.

ROAD MAP

4Q '20	Release Mfg Service Data Markup Tool Enroll 50 Mfg Companies to use tool
2Q '21	Process and link 1M+ part 3D models
3Q '21	Enroll 450 Mfg Companies to use markup Release of MOKN Portal and Plugins
4Q '21	Release of MOKN Portal + CAD Plugin
1Q '22	Portal Usage 100 unique users per month
2Q '22	Process and Link 5M+ part models Dataset downloads: 10,000
3Q '22	Enroll 1000 manufacturing companies Process and link 10M+ part models Portal user base grows to 1000 per month Commercialization Phase

PARTNERSHIPS

Academic partners NC State and OU bring deep domain specific manufacturing knowledge. NYU brings geometric processing and natural language processing expertise, while TX State brings knowledge management expertise. Industrial partners bring: 1) 3D engineering model datasets (OnShape, GrabCAD, Autodesk); 2) Small Manufacturers and non-profit manufacturing alliances; 3) Manufacturing software services (nTopology, Siemens). The 3D model datasets is a resource for scalable testing of geometric processing, search algorithms and recommendation engines. Feedback loops from their users and query terms will help improve algorithms. Small manufacturers provide an early user group and manufacturing alliances provide us with access to additional manufacturers. Saint Gobain and AWS NLP team will contribute engineering time to the natural language processing of manufacturing research papers. We are seeking new large partners who have interest in plugging into our knowledge network for supplier sourcing and discovery, plus an interest in keeping internal databases updated with near real-time information.

INTELLECTUAL PROPERTY

Software tools and datasets will be made available for use through open license. Manufacturing annotated data will be made publicly available under a CC-BY-SA 4.0 license. 3D Model datasets from our partners will be made openly available for use based on their original license terms of usage (MIT License). Geometric processing algorithms and NLP processing pipelines will be released under MPL2 license terms.

TRANSFORM THE TRANSPARENCY AND ACCESSIBILITY OF COURT RECORDS

Luis A. Nunes Amaral
amaral@northwestern.edu

David L. Schwartz   **Rachel D. Mersey**
Adam R. Pah   **Charlotte S. Alexander**
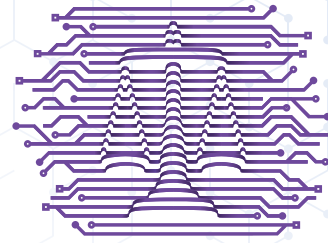
OVERVIEW

The U.S. court system produces millions of records per year. These records are supposed to be open to the public, but in practice they are trapped behind paywalls and dysfunctional, outdated software. Our project's goal is to enable a broad spectrum of public stakeholders to efficiently access, evaluate, engage with, and understand the work of the courts. Our mission is to create an Open Knowledge Network (OKN) that will serve as a foundation for advocates and researchers to analyze court data systematically. Our platform will also enable data gathering and integration that, in turn, supports intelligent analysis and meaning extraction so all citizens, entrepreneurs, journalists, lawyers, potential litigants, policy makers, scholars, and even the judiciary itself, can better understand and evaluate how the courts function.

DESCRIPTION

Data drives information and insight. Government agencies, central banks, health organizations, and law enforcement agencies all gather data in order to better understand and communicate the events, trends, and relationships between them that define our world. All of this is in service of the twin goals of understanding and transparency.

While the US court system collects similar data sets, those data are rarely used in support of the goals of understanding and transparency. This shortfall is the result of three features: availability of data, lack of data integration, and limited tools that support intelligent information analysis for non-technical users. Availability is hampered by the fact that much of the relevant court data resides behind a federal pay-for-use firewall. While individual case information is affordable, the data needed to do system level analysis would cost tens of millions of dollars to access. Integration is limited by data availability

as well as the tools and skills required to support this work at scale. Understanding and insight, even if one had full access to the documents, is thus hindered by the lack of available tools for the journalists, legal scholars, and decision-makers who most need it.

With these issues in mind, we have brought together a team of computer and data scientists, legal scholars, journalists and policy experts, to develop a suite of tools to enable access to court records and analytics. The Systematic Content Analysis of Litigation EventS Open Knowledge Network (SCALES OKN) aims to provide access to both the data and the insights contained within them to a broad range of users with diverse technical skills. Our project encompasses five complementary elements. Data Access: We will make all of the data we incorporate into SCALES and the integration of such data freely accessible to the public. Data Integration: We will provide tools to upload relevant data sets (e.g., judicial, firm, and corporate profiles) and support data cleaning, normalization, and integration. These tools will include intelligent data interpretation supported by Natural Language Processing, Machine Leaning and crowd-sourcing of instance tagging. The aim of this integration is the development of a *dynamic* knowledge graph that supports information access, analysis, and inference. Extraction of Insight: We will develop a public facing information access system that allows users to explore the data to answer questions about trends, comparisons, and correlations by simply asking questions. Expanding on work in language processing, information goals, and intent guided analytics, the system provides access to not just the data, but the information and insight contained within. Education: We will develop materials aimed at providing users with an understanding of both the data, the integration, and analytics methods. Community: We will foster the nucleation and organic

growth of a community of stakeholders that both add to and use the data and the system supporting it.

DIFFERENTIATORS

There are limited alternative sources of judicial records and none that enable systematic access and analysis for the public. Commercial legal services such as *Westlaw, LexisNexis,* and others have purchased many judicial records, but they limit access through their own fees and prohibit bulk downloads—foreclosing systematic analysis even for those who pay the hefty fees. A pioneering open alternative, The Free Law Project, maintains a user-generated free repository of court records, but its coverage is spotty and its interface is not designed to support systematic analysis.

We are unique in that we are providing users with not just access to raw data but also the meaning that it supports through access to intelligent analytics learned from across the legal system.

ROAD MAP

During year one of Phase II we will further develop our current alpha prototype and launch a pilot of SCALES with extended capabilities, such as intelligent identification of litigation events and geographic comparisons. We will also integrate information from other OKNs, further enriching the court-record data with links to business entities, judge characteristics, law firms, and statutes— contextualizing the entities within these records.

In the second year, we will expand the core technology and facilitate organic growth by helping external contributors to add data and annotations to SCALES. We will also expand on query/analytics capabilities following user needs.

PARTNERSHIPS

During Phase II, we will engage partners in academia, law, industry, and journalism. We are forming an advisory board comprising prominent experts to evaluate technical, legal, and ethical aspects of the SCALES OKN. The Free Law Project will continue to contribute data and will serve on the SCALES advisory board. Our other partners (including the MacArthur Justice Center; American Bar Foundation; Jenner & Block; The Center on Wrongful Convictions; NYU Law, Technology, and Policy Clinic; and more) have contributed their time to help us define SCALES as a product and will continue to provide their time in testing SCALES as it is developed and expanded. We

will also work with industry partners to identify corporate filings and collaborate on knowledge graphs that connect litigation with existing laws.

INTELLECTUAL PROPERTY

We plan to license for free all intellectual property we create under a standard open source license.

ADDITIONAL INFORMATION

With other Track A teams, we are creating Open Knowledge Network Solutions, a 501(c)(3) organization tasked with project management and code development for the creation and maintenance of the OKN platform. The entity will also hire a business-development professional to fund-raise and identify partners for continued development and long-term sustainability beyond the two-year NSF award. Northwestern University will provide administrative and event-coordination support.

TEXTBOOK OPEN KNOWLEDGE NETWORK

Michael Genesereth
genesereth@stanford.edu

Richard Baraniuk   **Craig Heller**
Mark Musen   **Vinay K. Chaudhri**

OVERVIEW

College students today face the challenge of mastering concepts in the new subject areas, relating those concepts across multiple disciplines, and one size fits all nature of textbooks. Intelligent Textbooks (ITB) using Artificial Intelligence (AI) and knowledge graphs (KG) solve these problems by allowing students to dynamically interact with the textbook content, increasing their ability to understand concepts, increasing engagement, and thereby, improving academic performance. ITBs offer students easy access to definitions and descriptions of concepts, make connections across different sections of the syllabus, and allow students to pose their own questions.

DESCRIPTION

Initial trials of ITBs that utilize KGs have been found to improve student grade outcomes by a full letter grade over the control group that was using a conventional textbook. ITBs have been found especially helpful for underperforming students,thus, broadening participation.

The potential of ITBs to facilitate better learning has been extremely difficult to realize without major investments of time, money, and expertise. The reason is that KGs are currently constructed using human subject-matter experts in a process that is extremely expensive and time consuming. Due to the large investment required, publishers and ed tech providers keep their KGs proprietary, eliminating their utility outside of the scope of the project for which they were created.

The primary outcome of our project is an open source Textbook Open Knowledge Network (TOKN) that can be freely used for creating ITBs and a variety of education technology applications. A secondary outcome is a novel process and tools for creating the KGs that combine automatic construction of a KG with validation by humans to ensure high accuracy. We envision a community of educators who would co-create TOKN, and eventually take the ownership for its future development and evolution.

We will use the new approach for KG construction developed in our project for Biology and Psychology undergraduate textbooks that are published by the open source publisher OpenStax. We will eventually integrate the ITB technology into the full OpenStax library of 41 textbooks, which has the potential to influence millions of students. OpenStax has further developed a robust ecosystem of 54 commercial partners (including every major publisher save one) who use OpenStax materials within their educational technologies and a further 30 major college and university systems who are using OpenStax to institutionalize open educational resources (OER) at their campuses.

Our KG construction tools will be integrated with Stanford's Protégé environment. Protégé is one of the most widely used knowledge authoring environments in the world today. In 2019 alone, Protégé was downloaded 145,000 times, and its web version hosts over 70,000 projects with over 50,000 user accounts.

DIFFERENTIATORS

An ITB relies on an explicit representation of knowledge in a textbook that matches human understanding and enables precise reasoning with it. The current narrative in AI is dominated by machine learning and natural language processing that achieve scale by sacrificing either accuracy or expressiveness. Such a compromise is acceptable in applications such as search, recommendation systems, machine translation, etc. In education, our domain models must be accurate and expressive. The textbooks need to be nearly 100% accurate. Any computer-based tools that will eventually be as good as human tutors have

to use an explicit model of the knowledge of the domain. These characteristics are also shared by many other applications, for example, income tax calculations and automated enforcements of a contract.

Our team is a world leader in research, application and education of formal knowledge representation methods. We have created strong partnerships with non-profit and for-profit organizations in the field of textbook publishing, educational technology, and academic search, thus, substantially enhancing the probability of success of our Phase II effort.

ROAD MAP

Activity	Milestone
KG Construction Tool	9 months
KG Tool Evaluation for Biology	12 months
KG Tool Evaluation for Psychology	15 months
ITB incorporating the KG	18 months
ITB Evaluation	24 months

There will be two primary products of our project: a KG construction tool, and TOKN that contains KGs for a few chapters from Biology and Psychology textbooks from OpenStax. We will evaluate the KG construction tool for OpenStax Biology and Psychology textbooks, and measure the time and effort required per chapter. We will evaluate the resulting KGs by incorporating them into an ITB and measure its impact on student learning and engagement.

PARTNERSHIPS

The foundation of the team is a partnership between Stanford University and OpenStax at Rice University. We bring together strong prior research on ITBs and KGs from Stanford with the vision to reimagine and reinvent textbooks from OpenStax.

To ensure that our innovations impact the commercial textbook industry, we have formed a collaborative relationship with Macmillan Learning. They will evaluate ITBs for two of their textbooks, gather user feedback, and help guide our future development.

To ensure that TOKN can be useful outside the context of an ITB, and for educational technology products, we have teamed with Educational Testing Service (ETS). ETS will specify the requirements of KGs for item generation, and once the KG is developed, evaluate it

for item generation and question scoring outside the context of an ITB.

To leverage and contribute to publicly available Big Data, we have formed a collaboration with Microsoft Academic Graph (MAG). In our collaboration with MAG team, we will investigate if their technology can help bootstrap the taxonomy for TOKN, and if TOKN could be incorporated into MAG ensuring its wide usage and contribution to the task of academic search.

INTELLECTUAL PROPERTY

TOKN, the KG construction platform, and new ITB applications will be released under an open source license. The existing ITB platform, called Inquire, is owned by Vulcan Inc, and Stanford has rights to sublicense it. Automatic Item Generation algorithms are owned by ETS and will be licensed to the Stanford team.

NETWORK SCIENCE FOR CENSUS AND STATE GEODATA

Moon Duchin
Moon.Duchin@tufts.edu

Charles Stewart

OVERVIEW

In 2020, public service administrators are struggling to modernize the flow of geo-electoral data between state and local officials. Our project will deliver an open-source integrated geodata management (IGM) tool to provide much-needed support in the delivery of critical services to the public. By structuring the resulting data in a multi-layer network, we can target optimal solutions to planning problems.

DESCRIPTION

The major deliverable of our Phase II project is an open-source software suite, aiming to build on the best features of existing election management tools, while expanding beyond elections to integrate with wider geospatial data sources and functionality. Our IGM tool will fuse three kinds of data that are often handled in separate and ad hoc forms: (1) shapefiles of precincts and administrative districts, (2) geocoded demographic data, and (3) election results for public office, ballot initiatives, and so on. State-specific data layers, from environment to public safety, will be incorporated. This tool will help meet the urgent needs of legislatures and state-level public administrators for more secure and current geo-electoral data, as well as lightening the technological burden on local-level civil administrators who have geodata requirements but no GIS staff. We plan for pilot adoption of the tool in two cohorts of five states each during the Phase II grant term.

IGM tool adoption will reduce data friction within states, producing a data stream suitable for network analysis and computational study. Our team is developing novel techniques in network science to leverage these new geospatial networks, with immediate applications in public service delivery, record maintenance, secure election administration, and disaster response. Efficient record linkage, partitioning, siting, and resource allocation will aid state officials with tasks as diverse as shoring up food pantries, de-duplicating voter files, and managing patient overflows from public hospitals.

DIFFERENTIATORS

We have assembled a unique and high-powered team that bridges from database design to accessible cartography to civic policy and administration to the research frontier in mathematics and computer science. The core team at Tufts houses mathematicians, developers, GIS specialists, and domain experts in the civic data space. Our MIT collaborators are trusted leaders in preparing quality-controlled civic, policy, and electoral data. We have secured nonprofit partners who are well-positioned to enlist state-level offices and stakeholders for the pilot cohort. Our partner civic organizations have well-established track records of effective public outreach. And we have high-level buy-in, including commitment of cloud computing resources, from tech leaders like Google and Microsoft.

We will build on the successes of the Geo-Enabled Elections project underway by our Phase II partner NSGIC while broadening the mandate significantly. This is a first-of-its-kind effort at a systematic solution to long-term, interlocking problems with the currency and security of state geodata.

ROAD MAP

Year 1

- Civics OKN holds Workshop 1 to share progress and prototypes with Track A peers in December 2020

- Pilot cohort of five states identified by January 2021

- Targeted geodata collection continues through May 2021

- Ongoing stakeholder input on usability to inform front-end interfaces

- Ongoing stakeholder input on essential features to inform back-end development

Year 2

- Civics OKN holds Workshop 1 to share progress and prototypes with Track A peers in December 2020

- Pilot cohort of five states identified by January 2021

- Targeted geodata collection continues through May 2021

- Ongoing stakeholder input on usability to inform front-end interfaces

- Ongoing stakeholder input on essential features to inform back-end development

PARTNERSHIPS

This complex project will be managed by the core team at Tufts University. MEDSL (MIT Election Data Science Lab) continues to be our major partner for Phase II in all aspects of planning and implementation. TAMU GeoServices and Mapbox are new partners with a civic tech and development focus. NSGIC (National States Geographic Information Council) will foster connections with their network of state GIOs (geographic information officers) and help build the pilot cohorts. NSCL (National Conference of State Legislators) offers access to a wide network of legislators and legislative staff. Common Cause and Lawyers Committee for Civil Rights are nonprofit partners from Phase I who will continue to ground our work in public outreach efforts for Census projects and products.

INTELLECTUAL PROPERTY

We are committed to open-source software development and plan to make all code and data publicly and freely available.

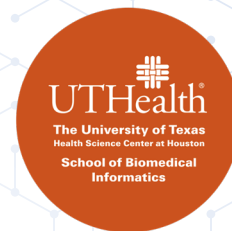# KNOWLEDGE OPEN NETWORK QUERIES FOR RESEARCH (KONQUER)

Lucila Ohno-Machado
lohnomachado@health.ucsd.edu

Peter Rose
Hua Xu
Ilya Zaslavsky

OVERVIEW

Good data lead to better decisions; however, if data are hard to find and not integrated across different fields of science, it cannot be analyzed to generate needed responses to worldwide challenges such as the COVID-19 pandemic. The KONQUER project aims to remove barriers to conducting more informed research by helping researchers find, integrate and analyze cross-disciplinary data in a seamless and intuitive way.
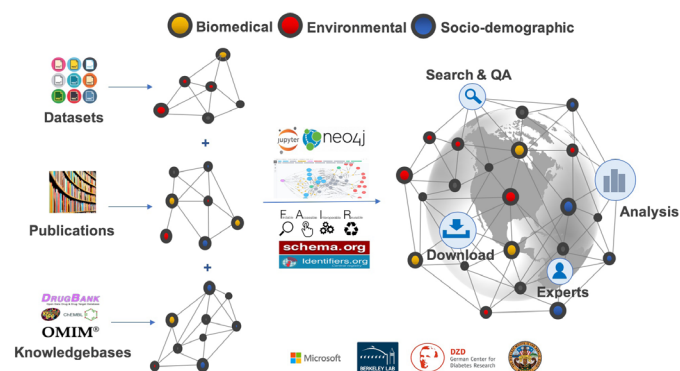
DESCRIPTION

KONQUER will demonstrate the capability of efficient cross-disciplinary research by integrating data from multiple sources: biomedical science (DataMed, funded by NIH) and geoscience (Data Discovery Studio, funded by NSF) data discovery indices. At its core, KONQUER will use the Open Knowledge Network (OKN) to help researchers and the general public get answers to complex questions by traversing a web of relationships between concepts from diverse scientific disciplines, and develop a comprehensive response to critical societal challenges such as the COVID-19 pandemic.

Use cases will be developed to answer timely scientific questions. We will develop a novel knowledge graph (KG) and natural language processing (NLP) methods to find answers to factual questions, e.g., "Are environmental factors such as temperature and humidity associated with COVID-19 transmission?". We will build a KG by integrating diverse data sources (curated knowledge bases, publications, and key datasets) and develop a KG-based question answering (QA) system. This requires the system to:

a) Locate and integrate datasets associated with cross-disciplinary questions, e.g., "What are the available datasets linking weather information with seasonal allergies at different locations?".

b) Use NLP to model spatial information to link across different types of geographies and spatial granularities (zip code, city, county, state).

c) Introspect, explore and analyze linked datasets, e.g., "Was the spread of COVID-19 worse when average temperatures were below 40F?".

d) Ensure that datasets and KGs are findable, accessible, inter-operable, and reusable (FAIR), to enable convergence of multiple research projects on a broader OKN infrastructure.

Additionally, we will collaborate with other C-Accel teams to develop inter-operable tools, thus affording the integration of various KGs to answer questions spanning additional disciplines.

DIFFERENTIATORS

KONQUER combines the expertise of team members from informatics, biomedicine, and geoscience domains. Compared with other efforts on data discovery (e.g., Google Dataset Search) and knowledge integration, KONQUER provides unique features: 1) cross-disciplinary data discovery via representative use cases involving

biomedical, socio-demographic, and environmental data; 2) advanced knowledge integration ranging from answering simple factual questions, finding relevant datasets, linking datasets across domains using geographic information, to interactively analyzing linked data; 3) leveraging of a sophisticated KG for data discovery and contribution of analysis results back to the OKN and; 4) NLP-enabled user interfaces to support broad types of users including citizen scientists, researchers, and policy makers.

ROAD MAP

Year 1 - We will construct a cross-discipline KG from diverse sources and develop a working system with intuitive user interfaces for querying, browsing, and analysis. Use cases from broad communities and other OKNs will be collected to refine the infrastructure and make it inter-operable. We anticipate introducing KONQUER to over 10,000 users.

Year 2 - We will further expand KONQUER with additional features such as an NLP-based QA. To achieve maximal impact, we will engage data creators, end users, and at least 10 academic institutions and industrial partners to objectively assess KONQUER's utility in advancing research.

Year 3 and beyond - With proof that our system can help a large user base, we will ensure sustainability by exploring an open source business model and building a foundation of strong user partners.

PARTNERSHIPS

We have assembled domain experts from academia and industry with decades of experience in creating and curating databases, integrating heterogeneous data sources, and deploying technologies to others.

Industry partners include *Microsoft* - will provide common queries, NLP tools, and expertise in data annotation/ harvesting using schema.org, *Neo4j* - will provide access to their enterprise graph database and tools and assist with KG integration across the NSF OKN ecosystem, and *Esri* - will provide expertise in creating analytical dashboards.

OKN teams including A-6950, A-6677, A-7099, A- 7152, A-7908 are committed to form a community of practice around the management of convergent, technology-focused research and work jointly on interoperability of KGs at the human-environment interface. Fifteen OKN teams plan to participate in a joint working group on data privacy and security. We are currently collaborating with

teams A-6677 and A-7160 in creating COVID-19 KGs that link infrastructure, population, environmental, pathogen, and biomedical data with disease outcomes.

The Biolink team at *Lawrence Berkeley National Laboratories* will provide expertise in property graph interoperability, the *German Center for Diabetes Research* and *EMBL-EBI Europe PMC*, will provide resources and expertise in linking literature data to datasets. The *San Diego County Health & Human Services Agency and Public Health Alliance of Southern California* will provide public health use cases and public health expertise.

INTELLECTUAL PROPERTY

We embrace the FAIR principles and will offer open data and open software using unrestricted licenses.

A MULTI-SCALE OPEN KNOWLEDGE NETWORK FOR PRECISION MEDICINE

Sergio Baranzini
sergio.baranzini@ucsf.edu

Sharat Israni
Sui Huang
Jim Brase

OVERVIEW

The human brain cannot possibly integrate the incredible amount of information modern societies have been able to amass. This hampers the generation of new knowledge, specifically in the biomedical sciences and its implications for human health, where the subject complexity is vast and stakes are high. Our knowledge network will incorporate billions of factual relationships among biomedical concepts, providing a discovery engine that will enable doctors, researchers, the pharmaceutical industry and the citizen scientist to explore biomedicine in its whole might.

DESCRIPTION

Human health has become so complex that even doctors turn to Google to understand difficult cases. Then, based on their extensive training, they can better interpret, diagnose and treat illnesses. Clearly, Google is not based purely on accepted science nor is it specialized enough to handle rare or complex conditions, find a cure for a given disease, or discover the root of a biological process. For those cases, specific information needs to be integrated by skilled researchers into formulating the right hypothesis (consistent with previous evidence and maximizing utilization of current knowledge), which then needs to be tested experimentally.

The response to each of the previous scenarios requires navigating a deluge of complex data and information and connecting the dots in a meaningful way. Our knowledge network (implemented both as *SPOKE* and *Google Biomedical DataCommons*) will integrate billions of biomedical concepts into a knowledge engine to enable doctors, drug developers and citizen scientists connect the dots and produce a biologically meaningful answer to these questions. Healthcare and related industries represent 1/5 of the entire US economy.

We anticipate extensive adoption of this platform will have a significant societal impact by reducing healthcare costs, health disparities and accelerating therapeutics, ultimately improving the quality of life for every American.

DIFFERENTIATORS

Even as high-throughput modern technology in biomedicine has facilitated the acquisition of vast amounts of data, it has only widened the chasm between its generation and its interpretation. Those approaching such a complex task based on individual strengths are bound to fail. While few efforts have been devoted to addressing this vacuum, we are pioneering the paradigm of Knowledge Networks in Biomedicine - a paradigm amply proven in Search, into a discipline that is inherently graph-theoretic. Our experience with Search (Google), graph theory (LLNL), PI participation in NCATS Biomedical Translator, and the long track record of creating this network is a testament to vision and commitment to transforming data into knowledge. The NIH Office of Data Science Strategy has even acknowledged this.

Our multidisciplinary program brings some of the most renowned experts in all aspects of biomedicine and data science, from doctors, researchers and epistemologists, to database specialists, computer scientists and statisticians. Our team also draws expertise from other programs in the C-Accelerator, thus capitalizing on existing talent to ensure overall track success.

ROAD MAP

As part of our Phase 1 program, we have already developed and made available a fully functional biomedical KN. Our Phase 2 proposal involves continuous development of the engine and 4 products powered by the knowledge network throughout the 2-year period with a commercialization plan for at least one of these

products to ensure sustainability. We will also demonstrate the validity and utility of the network via use-cases. Specifically, we plan to:

- Establish a governance structure and process for continuous expansion of the network by the biomedical community. (Q1, 2021)

- Create a business plan and incorporate a start-up company, NewCo, to support a subscription-based product for medical R&D. (Q1, 2021)

- Develop two additional products, a medical dashboard for the clinic and a biomedical knowledge interface for citizen scientists. (Q3, 2021)

- Develop and apply computationally intensive analyses to the network, enabling sophisticated validations of the network against real-world observations. (Q2, 2022)

PARTNERSHIPS

Phase 1 Partners (all will continue in Phase 2):

Google will continue being a key strategic ally through its DataCommons platform, including active participation in the governance of KN, and in-kind support for cloud computing and co-development.

Institute for Systems Biology (ISB) will continue to work with us both in the scientific development of the knowledge network and with resources to expand its utility and reach.

Lawrence Livermore National Lab (LLNL) will continue leveraging its extraordinary computing power and technical knowhow on graph theory and analytical approaches to guarantee an organic and balanced growth of the graph.

Additional Phase 2 Partners:

Indiana University: A new partner from the C-Accelerator (Team B6656/7036), Katy Borner's team will lend their world-class expertise in complex graph visualizations and analysis.

UC San Diego (UCSD): Also a new partner from another C-Accel Team, Peter Rose will bring expertise in protein domain and structure, a critical missing part of our KN at present.

Amazon: Will provide in-kind support for hosting SPOKE in the Neptune Graph Database environment on AWS.

UCSF Ventures: who have already provided seed funding for our NewCo, will provide support for the creation of a NewCo and handling of intellectual property matters.

National Center for Advancing Translational Science (NCATS): Our KN is an active participant of the NCATS as an Autonomous Relay Agent Team.

ELSI: We partnered with Erin Kenneally and Camille Nebeker (UCSD) on ethical legal and social implications of our project.

INTELLECTUAL PROPERTY

Inventions disclosures are being filed with UCSF Ventures. This will ensure proper documentation of existing IP and resulting IP from this project.

FLOOD INFORMATION FOR ANYONE & ANYTIME

Lilit Yeghiazarian
lilit.yeghiazarian@uc.edu

Sankar Arumugam   **Venkatesh Merwade**
Ximing Cai   **Torsten Hahmann**

OVERVIEW

On average flooding causes more than 100 billion dollars of economic loss and 500 deaths per years in the US. Some of this could be avoided if people have access to flood-related information such flood inundation depth at any location or possible evacuation route during a flood event. We will create a Google-like application that people can use to find out how flooding impacts them so they can take actions to mitigate it.

DESCRIPTION

Recent studies have shown that approximately 41 million people (~13% of the population) are at risk of severe flooding in the U.S. Despite such high risk of flood exposure, it is hard to find answers to simple questions such as "What is the total impact of flooding on a city?". This is because, while urban infrastructure is connected, data and models that describe them are not. We call this connected urban infrastructure the Urban Multiplex. It includes the power grid, transportation network, surface water and groundwater systems, storm water and sewage systems, drinking water systems, inland navigation and dams, all intertwined with the socioeconomic and public health sectors that form the fabric of modern cities. So, when one part of the Urban Multiplex fails from a flood, its cascading impacts across the city are unknown.

Our project aims to address the issue of quantifying flood impacts on an Urban Multiplex by developing a publicly accessible National-scale Urban Flooding Open Knowledge Network (UF-OKN). The proposed UF-OKN will connect multiple datasets, tools and models across the Urban Multiplex to ascertain and forecast the true impact of flooding.

Product design, implementation and delivery are guided by user needs and partnerships with local, state and federal agencies and private industries. The UF-OKN can be used to answer questions such as: "which route can I take to work during a storm?" or "will my house lose power during a storm?" At an organizational level, a decision maker can ask questions such as: "Which neighborhoods and when to evacuate to minimize human loss"? or "How flood risk will change in my city in the next 20-30 years"?

We expect that the UF-OKN will directly and indirectly serve millions of people impacted due to flooding - by providing the necessary tools and resources to enable real-time response and long-term planning and decision making.

DIFFERENTIATORS

Currently, flood related information is available from different sources - and in different forms. The two most common ones are flood insurance rate maps (FIRMS), produced by FEMA and available as static GIS files or paper maps; and dynamic flood forecasts and water levels provided by the U.S. Geological Survey and the National Weather Service. All these datasets are available only through the respective agencies and require some knowledge of how to navigate their systems. Additionally, these datasets cannot be easily integrated to create a holistic view of a flood impact on an Urban Multiplex at different temporal and spatial scales.

UF-OKN integrates flooding information with other related datasets in an Urban Multiplex so users ranging from an individual home owner to decision makers can get answers to their questions through a simple user interface or map on a mobile device or computer.

ROAD MAP

Phase II will deliver key technologic and products to serve the needs of two User Archetypes representing emergency responders and federal planners involved in planning and coordination to mitigate flood impacts). Our Strategic Framework consists of five Planes: User Relations & Product Development; Technology Development; Research & Development; Transfer to Practice & Sustainability Model Development; and Urban Flooding Scientific Community Development. Activities and milestones achieved in each Plane inform decisions throughout the Strategic Framework.

Month 8. *Milestones:* partner data, models & results ingested into UF-OKN; live demo for focus groups with key partners; knowledge models running; demos ready for beta testing; initiate branding. *Deliverable:* Minimally Viable Product (MVP)

Month 12. *Deliverables:* potential user/buyer feedback; operational infrastructure in place; flood impact socioeconomic assessment.

Month 18. *Milestones:* draft business canvas; refined UF-OKN ready for deployment; partner training, UF-OKN testing, add new features/functionalities; collect potential user/buyer feedback.

Month 24. *Milestones:* UF-OKN opens to general users; soft launch. *Deliverables:* Final business canvas.

Final *Deliverable:* Viable product/service, market evaluation.

PARTNERSHIPS

Phase I partners (Cities of Wilmington, Raleigh (NC), NC Dept of Public Safety; two FL counties) contributed personnel time, data, model results, prototype testing, evaluation and feedback. All Phase I partners will adopt UF-OKN for their operations, thus becoming our *first users*.

New Phase II partners will contribute at the same level as Phase I partners. They include USEPA (National repository of underground and above ground fuel storage tanks; National drinking water infrastructure data; Joint modeling of flood impacts in these critical facilities; real-time sensing data); USGS (real-time sensing data; will participate in sustaining UF-OKN past Phase II); NOAA (personnel time); six additional FL counties.

INTELLECTUAL PROPERTY

We anticipate that significant intellectual contributions will result from this project. Given the collaborative and multi-institution nature of our team, IP created by this project may be jointly owned. Participating project team members have agreed to work together on the protection, maintenance, and commercialization of any jointly owned IP according to applicable laws and policies.

CIVIL INFRASTRUCTURE SYSTEMS (CIS)-OKN

Nora El-Gohary
gohary@illinois.edu

ChengXiang Zhai
Arnab Chakraborty
Lucio Soibelman

OVERVIEW

This project will create a Civil Infrastructure Systems Open Knowledge Network (CIS-OKN). The CIS-OKN will empower stakeholders and decision makers such as State Departments of Transportation (DOTs) with data analytics and AI tools to access, integrate, and analyze CIS data — from multiple disparate sources and across disciplines and institutional boundaries — to improve construction, operation, and maintenance decision making. This will enable new modes of data-driven discovery, innovation, and decision making in the CIS domain and improve safety, accessibility, and economic opportunity *for all citizens*.

DESCRIPTION

The nation's *civil infrastructure* is deteriorating at an alarming rate. Over 21% of the nation's highways are in poor condition, over 188 million trips are taken daily across deficient bridges, and over 128 bridge failures occur every year. The American Society of Civil Engineers (ASCE) is warning Americans across the nation: "deteriorating infrastructure, long known to be a public safety issue has a cascading impact on our nation's economy, impacting business productivity, gross domestic product (GDP), employment, personal income, and international competitiveness". *Every citizen is affected.* According to ASCE, the resulting annual costs are over $2 trillion in business loss, $1 trillion GDP loss, 2.5 million lost jobs, and $3,400 disposable income loss *per household*.

Recent advances in data analytics and artificial intelligence (AI) have created a unique opportunity to address this grand societal challenge by linking, integrating, and learning from the existing wealth of data to better assess current conditions and predict future conditions of civil infrastructure systems, and derive insights about maintenance and investment decision making. *The data are big and rich, but cannot yet be fully harnessed because they are heavily siloed, isolated, and in heterogeneous formats — and numerous datasets (e.g., text and images) are even inaccessible because of their unstructured nature.*

The CIS-OKN will address these longstanding challenges by allowing integrated representation, access, and analysis of data from multiple sources, both structured and unstructured. It will disrupt the entire CIS domain by breaking the silos and unlocking the power of data to enable better funding, operation, and maintenance decisions. These data are coming from different disciplines and knowledge domains, and include design, construction, inspection, maintenance, and operation data of our nation's infrastructure, as well as social, economic, traffic, and environmental conditions.

In addition to benefiting a wide base of users from state and federal agencies and industry stakeholders, the CIS-OKN will improve the quality of life of *every citizen* by improving transportation safety and accessibility, as well as economic opportunity.

DIFFERENTIATORS

The CIS-OKN will establish a new paradigm in the CIS domain, where we can *for the first time* transform isolated CIS data, knowledge, and expertise into forms that enable integrative analytics and AI across disciplines and institutional boundaries. This *has long been needed, but not previously possible* due to the silos of data, the many disparate data schemas and systems, and the scattered knowledge and expertise. Breaking these silos is a challenge that cannot be solved by isolated teams, and instead needs a convergence approach that includes academia, industry, and government agencies as well as expertise from multiple disciplines working together. *The CIS-OKN is not simply a better choice among other*

possible choices because it is allowing these powerful types of analyses to happen for the first time.

ROAD MAP

In Phase I of this project, the research team and partners developed a thorough understanding of user needs and built an initial working prototype, laying a solid foundation for the CIS-OKN. *In Phase II (2 years) we will refine, scale, and implement the prototype with a primary focus on deterioration prediction and maintenance decision making for the nation's bridges and road infrastructure. Our main deliverable will be a fully tested and functional CIS-OKN ready for adoption by State DOTs. The CIS-OKN will include:*

- CIS datasets from multiple sources (federal and state agencies, contractors, public sources, etc.).

- CIS Data Schema, Ontology, and Knowledge Graph, with mapping to existing schema.

- Software tools for extracting information (e.g., bridge defects) from text sources and images (e.g., bridge inspection reports and images).

- Software tools for linking data from multiple sources into the CIS Knowledge Graph.

- Software tools for machine learning, AI, and analytics for data-driven decision making.

- Software tools for interactive access, search and retrieval, and visualization using BIM and GIS.

PARTNERSHIPS

Strategic partnerships with federal and state agencies and key industry players will enable us to make rapid advances and accelerate transition to practice. Their contributions in Phase I have been instrumental, providing us with data, collaborating in the development of the CIS-OKN using a participatory design approach, and helping us engage with additional partners and users. In Phase II, our partners will continue to share data and expertise, as well as support pilot implementation, facilitate rapid adoption, and advise on the sustainability of the CIS-OKN beyond Phase II.

The CIS-OKN partners and collaborators include:

- Federal Highway Administration (FHWA), the main federal body that provides stewardship.

- 10 State DOTs (AZ, CA, CT, FL, IL, IN, IA, SC, UT, WI), who will be the primary users.

- buildingSMART and the National Institute of Building Sciences, who will provide expertise in creating and mapping to data standards.

- 11 key industry leaders in the transportation domain, including technology providers and consultants (ESRI, WSP, HDR Engr., JobsiteTech, InfoTech, Leica, Fair Cape Consulting, Spectrum Technologies, Alta Vista Solution, Roadbotics, Global Quality Corp.).

- Google, Microsoft, and Amazon.

- Midwest Big Data Hub and 3 major transportation centers (IL, CA, WI).

INTELLECTUAL PROPERTY

We had two efforts that had patents or were in process. We are going to make new efforts so as not to require us to deal with licensing issues.

TRANSPORTATION EQUITY KNOWLEDGE NETWORK

Robert C. Hampshire
hamp@umich.edu

Carol Flannagan   **H. V. Jagadish**
Tayo Fabusuyi   **Margaret Levenstein**

Transportation | Data | Equity | Policy

OVERVIEW

How many children live in our cities' food and transportation deserts? How many people miss medical appointments due to lack of transportation? What strategies can we borrow from peer cities to improve job accessibility? How will we deploy driverless vehicles in our city in an equitable way?

This project enables transportation and economic development officials, private companies, and advocacy groups to make equitable planning and investment decisions by providing data-driven insights into the multifaceted role transportation plays in the lives of citizens and communities' functions.

DESCRIPTION

Ridesharing, scooters, and bike share options are just the beginning; driverless vehicles and drones are around the corner. As this sector continues to grow, so does the universe of information available to policymakers and practitioners. However, this information is often scattered across multiple sources and is not always easy to access or identify. Mobility companies maintain large stores of private travel data and proprietary knowledge graphs. Integrating such data—and making the resulting information accessible to the public—is critical for policy decision-making, but cannot be accomplished by any individual firm.

We all, as a society, are impoverished by the many people being left behind by inefficient and inequitable access to transportation. We lose $15B each year due to missed medical appointments from a lack of transportation; U.S. housing values are $1.68 trillion lower than they could be, due to lack of access to transit.

We are creating a transportation equity open knowledge network (TEOKN) to connect vast amounts of disparate public information about why, how, when, and where people travel; and resulting individual and societal outcomes. The TEOKN will be broadly available to users in the public and private sectors. A data safe haven will also facilitate the integration of private data. Partners will be able to easily and securely contribute data to the OKN, as well as extract safe (adequately protected) data from it. Innovative technical outputs will include an ontology for transportation and new mobility data, a transportation security index, and data source fusion techniques.

At scale, the TEOKN will apply the public health model for improving community health and wellbeing to transportation inequity: 1) *Measure the Problem* with transportation equity indicators delivered on a public website; 2) *Identify Factors* through knowledge-discovery tools and results made publicly available; 3) *Test Countermeasures* in cities with case studies; and 4) *Ensure Widespread Adoption* by sharing case studies in an open peer learning environment.

DIFFERENTIATORS

An ecosystem of data products takes into consideration the relationship of transportation to other aspects of life. However, the current data ecosystem is fragmented, and users have only a partial view of available information. Our TEOKN will link together these products with newspaper and academic articles, case studies, as well as GPS data, and novel measures of unmet demand. Our TEOKN stitches together this patchwork of isolated products to unlock insights revealed in the connections and patterns. For example, insights from our demo TEOKN developed in Phase I contributed to the City of Detroit's Office of Mobility Innovation's response to COVID-19. In the earliest days of the outbreak, the Detroit Department of Transportation suspended service due to a temporary bus driver strike. The City was able to leverage the TEOKN team's analysis of travel patterns and alternate transportation modes in Detroit to mitigate the effect of

decreased transit service on essential workers. During Phase II the team will expand our partnership with Detroit to include food security, and health outcomes for seniors.

ROAD MAP

• *Year 1:* Deploy and evaluate TEOKN-powered web tools with the Chicago Fed and Shared Use Mobility Center (SUMC) to enable 10,000+ policy makers to assess transportation equity impacts on the economy, and environmental impacts.

• *Year 2:* Launch a nationwide Transportation Equity TEOKN (TEOKN V1.0) to reach policy makers and advocates representing more than 10 million people. Use the TEOKN to lead community transportation equity assessments in two medium-sized cities. Disseminate TEOKN insights via peer learning tools, data-linked interactive case studies, workshops, and community engagement activities.

• *Beyond Year 2:* Expand TEOKN to support additional policy use cases and provide technical assistance to small and medium-sized communities. Foster a tiered-membership model and a private innovation ecosystem to build products and services based on TEOKN.

PARTNERSHIPS

We seek to work hand in hand with partners to accelerate their existing service offerings while simultaneously co-developing the TEOKN. Current partners' data products aim at regional economic development, sustainability, mobility knowledge discovery and learning, and accessibility, respectively:

• *Federal Reserve Bank of Chicago* will in Phase II integrate transportation equity and novel job accessibility metrics into their Peer Cities Identification Tool and host regional policy and research conference based on the TEOKN.

• *Shared Use Mobility Center* will in Phase II co-develop enhancements to their popular Mobility on Demand (MOD) learning center, opportunity analysis, metro profiles, and emissions calculator.

• *Urban Institute* will in Phase II lead community transportation equity assessments.

• *National Renewable Energy Laboratory* will in Phase II contribute their nationwide Mobility Energy Productivity (MEP) metric.

• *Deloitte* will in Phase II deliver a robust and scalable implementation of the TEOKN using their deep expertise building knowledge networks for government clients.
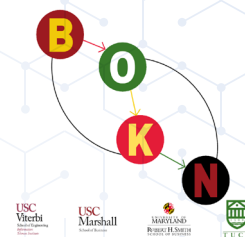
INTELLECTUAL PROPERTY

The TEOKN will be built using background IP developed by Deloitte and the University of Michigan. New partners will have access to the TEOKN platform and any new co-developed IP.

BUSINESS OKN AND ENTREPRENEURSHIP PORTAL

Jay Pujara
jpujara@isi.edu

Gerard Hoberg   **Gordon Phillips**
Craig Knoblock   **Louiqa Raschid**

OVERVIEW

Most entrepreneurs fail. One cause is data inequity: established companies have a wealth of data, while entrepreneurs struggle to find the resources they need to succeed. The Business Open Knowledge Network (BOKN) will integrate knowledge from millions of company webpages, social media, patents, and regulatory filings. The BOKN will fuel an Entrepreneurship Portal, an easy-to-use tool we are developing to help entrepreneurs find customers and partners, explore the competitive landscape, and commercialize ideas.

DESCRIPTION

Starting a business and succeeding is an essential component of what many consider the American Dream. Unfortunately, entrepreneurial activity in the US is declining. Beyond hard work and innovative ideas, entrepreneurs face a host of information-critical challenges: finding customers, differentiating from their competitors, securing suppliers and partners, and commercializing innovations. Hours on search engines can yield meager results, while commercial tools are costly, limiting entrepreneurs from their true goals of building a successful business.

The Business Open Knowledge Network (BOKN) is designed to fill these informational gaps. Our team combines domain expertise from business school professors at the Smith, Tuck, and Marshall schools of business and the latest AI techniques developed by computer science faculty at the University of Southern California's Information Sciences Institute. Using a vast and diverse set of data from websites, social media, regulatory filings, and patents, our team is creating a knowledge graph of nearly a million public and private US companies and modeling the relationships between them.

BOKN will provide a *hybrid* knowledge graph, where information and relationships delivered are a combination of traditional data queries to find factual information such as a firm's address or CEO, and results from state-of-the-art machine learning to predict the most likely competitors and customers. To build this knowledge graph, we will pioneer new approaches to learning entity representations in high-dimensional vector spaces, matching and resolving entities across data sources, and predictive analysis which will fuel insights into the growth of new businesses in the US.

The knowledge in BOKN provides the foundation for our entrepreneurship portal (E-Portal). The portal offers entrepreneurs an accessible interface for to describe their business plans and instantly receive guidance. Users of the E-Portal can find market intelligence to identify close competitors, helping entrepreneurs to differentiate their business plans. Analyzing historical patents, the portal will suggest potential commercialization strategies and licensors of intellectual property. Customer discovery features allow entrepreneurs to engage with other businesses who may wish to purchase their products and services, and supply chain analysis will help new businesses find the materials or components to manufacture their product.

During Phase I of the Convergence Accelerator, the BOKN team created a rich, hybrid knowledge graph, developed a prototype E-Portal, and engaged with various stakeholders in the government, financial technology firms, and entrepreneurship centers to focus our efforts. BOKN provides a vertically integrated solution to go from data to insights for entrepreneurs, while simultaneously providing horizontal services for matching entities, processing large amounts of noisy text data, and supporting both complex data-intensive queries and fuzzy search for business-specific concepts. BOKN resources can support a broad range of

applications, allowing researchers to track innovation and regulators to estimate economic and financial risk.

DIFFERENTIATORS

Resources for entrepreneurs abound, but our project provides a unique combination of data and AI-based analysis. Public programs are focused on general guidance for starting businesses, but do not tailor their recommendations to the entrepreneur's needs. Commercial tools tap into proprietary data curated by investment and finance firms, but come at a steep cost while still being limited in scope. BOKN uses state-of-the-art models trained on terabytes of data to analyze business webpages, patents, and regulatory filings. Using a spatial representation of business activities, we map each public and private company to a 200-dimensional space, allowing us to find tailor-fit matches. Our tools are free, designed to be easy to use, and allow entrepreneurs to iteratively refine ideas.

ROAD MAP

BOKN's Phase II roadmap involves parallel tracks contributing to creating a fully featured entrepreneurship portal and developing the core technologies necessary for constructing and maintaining Open Knowledge Networks. Early in the program, we will release an improved prototype E-Portal, and provide quarterly updates, each focused on a core capability such as market intelligence, commercialization, customer discovery, and supply chain analysis. Our partner organizations will help run regular E-Portal evaluations several times a year. In addition, BOKN will release a platform for OKN development, supporting knowledge graph construction, including learning about entities from large quantities of data, linking entities across knowledge graphs, and creating hybrid knowledge graphs, along with tutorials and examples to encourage adoption.

PARTNERSHIPS

The BOKN team will have strategic partnerships with other teams in the cohort, including the Manufacturing OKN, to help include richer metadata about manufacturing firms and enable more flexibility in manufacturing supply chains, and the Legal OKN, to match companies and key personnel with legal cases. The Greif Center for Entrepreneurial Studies at USC will assist in evaluating and disseminating our portal and improving its efficacy. The National Center for Women and Information Technology will also support engaging diverse populations to use the BOKN. IBM and

Open Corporates will provide data and analytic tools in addition to those developed. Data.world and the National Bureau of Economic Research will host BOKN resources to provide long-term continuity and improved access to data for researchers across multiple disciplines.

INTELLECTUAL PROPERTY

The entire BOKN knowledge graph will be available under a Creative Commons CC-BY license. The data will be completely open for both research or commercial use. This license allows us to maximize the interest and use of the knowledge graph, which, in turn, encourages the community to contribute new data, link, and extend BOKN.

CHEMICALS & MATERIALS OPEN AUTHORITATIVE KNOWLEDGE

James K. Ferri
jkferri@vcu.edu

D. Tyler McQuade
Monty Alger

OVERVIEW

The Chemicals & Materials Open Authoritative Knowledge (OAK) A7017 is partnering with leaders in chemical manufacturing and information technology to create an ecosystem of data, tools, and people to support chemical and materials industry and education from material procurement to end-of-life. The information architecture in this ecosystem will support academic and commercial partnerships through interactive cloud services and enable market-to-manufacturing connections for those who manufacture, formulate, and benefit from chemicals and materials.

DESCRIPTION

Chemistry touches more than 96% of all manufactured goods. American chemistry provides the materials to make life healthy, safe and comfortable – contributing to a higher standard of living for people in the U.S. and around the world. Solutions from this industry are needed to ensure a safe and plentiful food supply, clean air and water, efficient and affordable energy sources, and life-saving medical treatments.

OAK supports innovation and promotes general understanding of the science and technology that touches more than 25% of the U.S. gross domestic product. OAK has three main focuses:

1) Partnership with stakeholders to identify key data resources and expand existing datasets in the ecosystems. These data sets will be used to create models for both manufacturing transformations and materials performance at multiple length scales.

2) Development of an authoritative semantic ontology for describing chemical systems knowledge. It is centered on stakeholders, features stakeholders want, systems that delivery those features, and components that play the functional roles required by these systems.

3) Design of a cloud-based platform for connecting people with chemical and materials knowledge and tools that satisfy a wide range of unmet needs in an open community that sustain ongoing collaborative development.

DIFFERENTIATORS

Chemical network research generally focuses on chemical properties or transformations. There are a few commercial products based on such premises, the most prominent being MilliporeSigma Synthia (formally Chematica), Elsevier Reaxys, and CAS SciFinder[n].

This project develops a framework centered on the needs of stakeholders from industry, government, and the public at-large. Industrial collaborators (Dow Chemical, Procter and Gamble, and IBM) on this project are composed of leaders in chemicals and materials manufacturing and information technology spanning vertical depth and horizontal breadth in the business of chemistry and leading innovation in cloud-infrastructure. The strength of our chemicals and materials domain working groups will drive cross-industry consensus on the creation and evolution data models and methods, as well as standards on tool and data interoperability.

ROAD MAP

During the first year of Phase II Team A7017 will deliver a cloud-based Knowledge Management System to demonstrate functionality across an open data store and a SysML modeling platform. A key aspect of this framework is making it appropriately general so that it is extensible beyond the initial case studies (see Additional Information section).

In collaboration with academic teams, chemical and materials manufacturing partners will develop and seed an open pattern library with exemplars from manufactured consumer products. The consumer product and chemical domain experts will deliver large high-quality datasets relating to chemical and material properties for populating the design component aspects of the open library.

The second year of Phase II will expand the digital infrastructure necessary for continued growth of OAK. This cloud-based infrastructure includes a quantitative modeling, data transfer and security protocols, machine learning workflows, and visualizations for understanding the information within the network. The team will also deliver results of initial user surveys measuring the stakeholder perception of utility and impact.

PARTNERSHIPS

The Dow Chemical Company (Dow) and the Procter and Gamble (P&G) Company offer expertise in the design of our platform, as well as serving as stakeholder archetypes. Through Phase II, both companies will continue to contribute expertise, data, and tools to enhance the design the knowledge management system to accelerate chemical innovation. Specifically, Dow contributes expertise in chemical manufacturing, high-throughput experimentation, and chemical modeling. P&G adds expertise in chemical product formulation, market analysis, and chemical systems modeling.

IBM will also lend expertise in model-based systems engineering to help the team development the necessary information technology architecture for future success. As a Phase II partner, IBM will provide their systems modeling software platform, Rhapsody®, for creation of the ASK framework and support development of an open, collaborative environment through the IBM Garage.

INTELLECTUAL PROPERTY

OAK will contain both public and proprietary knowledge. The platform will allow for data sharing to have multiple levels of privacy from totally public to completely confidential. Intellectual property will be shared across co-developers and the commercialization team.

The OAK team is supported by legal experts in intellectual property and pre-competitive collaboration to ensure that data sharing is done in a responsible manner.

ADDITIONAL INFORMATION

During Phase II, the OAK effort will be focused around manufacturing case studies designed to target different aspects of the chemical economy. In one case study, we will map the pharmaceutical supply pattern of a tableted generic drug product. By analyzing the chemical supply chain, we will identify "capability gaps" in pharmaceutical manufacturing - and ultimately the specialty and fine chemical supply chain in the United States. In the second case study, we focus on product supply patterns in personal care products. Specifically, we focus on the role of water-soluble polymers, functional fibers, and performance fabrics and the roles they play in personal care.

The OAK team will promote track success by collaborating with the C-Accel cohort regarding network ontologies, automated data federation, intellectual property management, and nearby regions of the global Open Knowledge Network.

TRACK A

HORIZONTAL

TEAMS

NSF | Convergence Accelerator

TRACK A | Horizontal

NSF seeks to create an Open Knowledge Network (OKN), which has the potential to drive innovation across all areas of science and engineering and unleash the power of data and artificial intelligence to achieve scientific discovery and economic growth. The overall goal of the Open Knowledge Network track is to **enable the creation of a nonproprietary shared knowledge infrastructure that allows stored data (both structured and unstructured) to be located and its attributes and relationship to other data and to real-world objects and concepts to be understood at a semantic level.** This track will have a particular focus on exploiting publicly available U.S. Government data and similar public datasets. Teams may address "horizontal" challenges that apply to all domains, such as developing the underlying representation of facts, querying services that perform reasoning tasks with the data, or developing secured access capabilities.

SPATIAL DECISION SUPPORT FOR THE OPEN KNOWLEDGE NET

Sean Gordon
sng3@pdx.edu

John Gallo   **Jeffrey Hamerlinck**
Kathleen Stewart   **Xinyue Ye**

OVERVIEW

The Spatial Decision Support for Open Knowledge Networking project (SDS-OKN) is creating a cloud cyberinfrastructure of tools and templates to help multi-stakeholder groups create and share knowledge for improved collaboration and decision making. This SDS-OKN Cloud combines new knowledge management techniques (semantic web, knowledge graphs) with the capabilities of spatial mapping and analytic models (spatial decision support). The team's initial focus is on human-environment problems, including those pertaining to water quality, agriculture, biodiversity and wildfire, but tools and techniques will be broadly applicable to collaborative decision making and problem solving in any domain using geographic information.

DESCRIPTION

Semantic web and knowledge graph (SW/KG) technologies offer new capabilities to find and link information across the internet and enhance our abilities to browse, search and even reason with this information. Their use to date has been concentrated in the high tech and commercial sectors, with little application to public goods and services. While SW/KG can make existing information more accessible, they have limited ability to generate new information, especially for analyzing public resources across space and over time. This project is overcoming these limitations by creating tools for non-technical users to gather information, create their own knowledge graphs, and link these to spatial decision support capabilities for addressing space/time problems. Also included is a decision workbench for experts, and a robust programming interface (API) for cross-linking with partner tools.

Human-environment problems are the initial focus because they involve public resources necessitating management across space, time, disciplines, organizations and jurisdictions. For example, water quality has long been a top environmental concern of the American public. Each state has tens to hundreds of water management districts and is responsible for hundreds to thousands of multi-stakeholder water quality remediation plans under the EPA Total Maximum Daily Load program. The SDS-OKN will help users synthesize this type of information and translate it into actionable knowledge for public agencies and citizens.

Any OKN ultimately serves people who provide the inputs and use the outputs. But all new technologies face barriers to adoption, particularly if they are attempting to cross disciplines, organizations and jurisdictions. The SDS-OKN team brings expertise in technology adoption and participative applications to address these challenges. We are also leveraging our experience and collaborations with a rich array of regional partnerships to design techniques and governance systems for integrating these technologies into multi-stakeholder decision making.

The team will work with investors and partners to finalize and initiate a business plan, which could include contracts with collaboratives in other regions seeking assistance with the tools, and contracts with partners leveraging the API and desiring additional functionality.

DIFFERENTIATORS

The project has the following strengths:

• The ability to link semantic technologies to spatial technologies to generate new information and share it easily.

• Commitments of collaboration from multi-stakeholder partnerships at regional, state, and local levels.

• Inclusion of governance techniques that integrate technologies into and across multi-stakeholder, multi-sector, multi-discipline, decision making.

ROAD MAP

Year 1

• Release a SDS-OKN Cloud v1.0 toolset which enables collaborative knowledge graph creation and population, as well as the chaining of spatial and knowledge analysis operations.

• Work with existing collaborators to deploy v1.0 to multi-stakeholder partnerships.

• Develop cross-cutting knowledge schema for land use/cover, wildfire, environmental justice, zoning regulations and education.

Year 2

• Release a SDS-OKN Cloud v2.0 toolset, which will enable: 1) an option to turn-on automated, semi-supervised implementation of v1.0 capabilities, 2) creation of public-facing websites customized for knowledge sharing, and 3) creation of a customized workspace for each region for analytic collaborations.

• Use cross-cutting knowledge schema for sharing information among regional stakeholder communities.

• Launch a web portal where any interested group can access our tools and educational materials.

• Initiate the business plan, securing clients and funds for Phase III (Year 3 and 4)

PARTNERSHIPS

Multi-stakeholder Partnerships

These partners provided user needs and prototype testing in Phase I and will support deployment and further product testing in Phase II:

• Puget Sound Partnership (several hundred organizations)

• The Tulalip Tribes of Washington

• Oregon/Washington Cascade to Coast Landscape Collaborative (1000+ members)

• California Central Valley Landscape Conservation Project

• Maryland State Parks Equity Initiative

• Santa Barbara County Conservation Blueprint

• Deschutes National Forest & County, Oregon

Professional/Industry/Government

These partners are providing access to professional networks, data and the development of data standards and spatial data analysis software.

• Spatial Decision Support Consortium

• University Consortium for GIScience

• Esri, Inc.

• Open Geospatial Consortium

• CitSci.org

• US Geological Survey

• US Fish & Wildlife Service

• US Environmental Protection Agency

• USDA Forest Service

• Washington Department of Ecology

INTELLECTUAL PROPERTY

Tools developed will be released under open source licenses, allowing maximum dissemination of the created materials and non-sensitive data.

ADDITIONAL INFORMATION

More information, including a 5-minute explainer video, can be found on the project website:

http://www.sdsconsortium.org/sds-okn

KISMET - KNOWLEDGE OF INTERNET STRUCTURE MEASUREMENT, EPISTEMOLOGY, AND TECHNOLOGY

kc claffy
kc@caida.org

David Clark

OVERVIEW

We will create a knowledge network that will improve the security and functioning of three key but inherently vulnerable systems that underpin all activity on the Internet: the addressing, routing, and naming systems. Lack of attention to security in early design decisions, persistent disagreement about the best path to improvement, and lowest-cost operational practices surrounding these three layers have allowed malicious actors to execute and scale harmful misbehavior including interception and disruption of traffic, denial of service and phishing attacks, and distribution and execution of malware.

DESCRIPTION

Abuse of the Domain Name System is rampant: the Internet Corporation for Assigned Names and Numbers (ICANN) reported in March that there were about 740K active domain names associated with abusive practices. The Internet routing system shares similar vulnerabilities but little reporting; only high-profile episodes make headlines, such as two uses of BGP in 2018 to steal cryptocurrency and create $29M of fraudulent advertising revenue.

The premise of our work is that we cannot improve security by responding to individual events. Rather we must shift the landscape by understanding the operation of these Internet systems and behavior patterns of malicious actors, with the goal of finding operational practices that will thwart attackers.

Our Open Knowledge Network is structured to yield two types of knowledge: technical knowledge (Stage 1), with significant scientific and engineering components (tool developments, data analysis), and actionable knowledge (Stage 2) of direct use to society. To enable Stage 1, we will gather data sources and expertise related to these systems, in order to support scientifically reproducible

data analytics that can identify ongoing security threats, evaluate proposed mitigations, and track improvements. In Stage 2, we will cultivate an international operational community, drawing from disciplines including law and economics, to develop and socialize operational practices that can prevent or mitigate vulnerabilities.

Our success metrics will include reduced "time to insight" regarding structural characteristics of these systems, methods to develop, evaluate, and instantiate enhanced practices in the Internet naming systems, and demonstrating the utility of the OKN in verifying compliance with enhanced practices. We will analyze models of sustainability for the OKN, learning from other critical industries, e.g., financial and energy sectors.

Our long-range goal is to use our OKN to identify, incentivize, and validate operational practices that improve Internet security.

DIFFERENTIATORS

Approaches to solving fundamental Internet vulnerabilities have been: (1) largely technical, failing to overcome non-technical barriers; (2) complex and expensive; (3) global in scope. These properties are not compatible with the pressure of a competitive ecosystem, or in some cases governments with counter-incentives to address infrastructure security issues. We have spent years studying the vulnerabilities and tracking the impact of proposed solutions. Our conclusion, re-affirmed by our Phase I work, is that the path to better security does not lie in proposals to make global changes to the Internet protocols, but in finding operational practices that regions of the Internet can implement to improve the security profile of those regions. We identify important changes in Internet traffic patterns that reduce the necessity for global

agreement in order to improve security in regions of the Internet, increasing our likelihood of success.

Many organizations around the world generate data that can contribute to the OKN, some of whom are collaborators and partners in this effort. Navigating the diffuse nature of information about network infrastructure is a significant challenge. We will pursue a community-driven approach to sharing data using consistent formats and APIs, and identify metadata that enables discoveries that cut across data sets and domains. Our critical assets include partners with commercial network engineering expertise, data science techniques, and technology to manage data integrity, availability, and privacy.

ROAD MAP

Y1: Establish OKN, including annotated maps of routing and naming systems

- Demonstrate use of OKN to identify properties of network and DNS service providers associated with persistent misbehavior or large attack surfaces.

- Develop tools to understand key challenges in secure network configuration.

- Use OKN to verify compliance with operational practices required by the Mutually Agreed Norms for Routing Security initiative (MANRS).

- Workshops with industry experts.

- Develop extension to MANRS to prevent hijacks; validate benefit with OKN.

Y2: Generate and validate actionable knowledge.

- Create and demonstrate utility of DNS+BGP knowledge graph to improve security properties

- AI-driven assistant for security configuration.

- Develop catalog of best practices for DNS ecosystem. Demonstrate use of OKN to verify utility. Socialize/refine with industry actors.

PARTNERSHIPS

Securing the Internet requires national and international partnerships that combine previously disconnected data sources, user communities, operators, regulators, and other stakeholders to overcome conflicting incentives that hinder development and deployment of data-driven solutions. The leaders we have selected span government (ESnet, NIST, FCC), industry (Verisign, PIR, Comcast,

Farsight, Interisle), non-profit (ISOC, DNS-OARC, SIDN Labs, M3AAWG), and academic (MIT, U Oregon, BYU, Virginia Tech, U Waikato) sectors. Our partners bring diverse data sources, but more importantly they bring deep understanding of network engineering, operations, cartography, curriculum development and training, research infrastructure, economics, policy development and other disciplines fundamental to the creation of this prototype OKN. This interdisciplinary collaboration is required to not only create and sustain the OKN, but also to use it to evaluate and apply technical knowledge in economic, legal, and social contexts.

INTELLECTUAL PROPERTY

All software source code brought into project will be BSD, GPL, or UCSD licensed. Data sharing will rely on CAIDA Acceptable Use Agreements as a basis, with modifications as needed by data owners.

THE FEDERALISM DATA AND ADVANCED STATISTICS HUB (F-DASH)

Jason Windett
jwindett@uncc.edu

Samira Shaikh   **Isaac Cho**
Stephanie Moller   **James Hendler**
Jeffrey Harden   **Nathaniel Birkhead**

Federalism Data and Advanced
Statistics Hub (F-DASH)
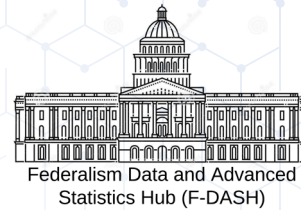
OVERVIEW

The fragmented nature of state and local governments leads to inefficiencies in collecting and analyzing data on policies and the entities involved in the policy process. The *Federalism Data and Advanced Statistics Hub* significantly advances extant efforts to link open data with *dynamic* data collection of government institutions on political actors, proposals, rules, laws, and outcomes at the federal and subnational level.

DESCRIPTION

Public policies have effects on social welfare, work-family, education, and the labor market that directly influence the everyday realities of citizens. Policies also have implications at the aggregate level, including economic growth, industrial development, the natural environment, the built environment and economic inequality. The effects of public policy have been vividly illustrated by the state and local government response to the COVID-19 contagion.

Scholars have established that insufficient availability of data undermines the ability to assess the intended and unintended effects of policies upon enactment. There has been prior work to fill this void, but the efforts remain narrowly focused by time, geography, or subject. Our goal is to go substantially further in breadth and depth than most current efforts to make the data and its analysis easily accessible to the American public. F-DASH has the potential to influence nearly all future research on U.S. public policy — from health to education to criminal justice — to become a multi-user resource for citizens, journalists, educators, students and an essential and innovative source for policy-makers and scholars of federal, state, and local governance.

In Phase 2, we propose to deliver the next version of F-DASH, a comprehensive knowledge graph related to federal, state, and local policy-making in the United States. Our deliverables are: 1) an integrated, centralized knowledge graph consisting of existing fragmented datasets compiled by scholars and government sources; 2) tools to dynamically capture, clean, visualize, and curate data, including digital source documents related to public policy outputs; and 3) a user-friendly open data hub linking federal, state, and local open data sources around common themes, policy areas, or social characteristics.

Once built, the F-DASH will require managing. To sustain the dynamic collection process, maintain webspace, etc. we will pursue a combination of internal support from our universities and from external grant making organizations such as the Knight Foundation, the Rita Allen Foundation and the Democracy Fund.

DIFFERENTIATORS

The F-DASH team has three key differentiators: *First,* the team has balance. The team is equally divided between social scientists and computer scientists, meaning that the development of the F-DASH uses cutting edge tools in artificial intelligence and visual analytics, while also reflecting the state of art in federalism and public policy. *Second,* the F-DASH team is diverse. The team has equal gender balance, with broad subject matter expertise - from data linking to ethics - and while it has many complementarities, it has no redundancies. *Third,* the F-DASH project meets a clear need, has a well-articulated goal, which has attracted the support of numerous community partners. Our effort combines existing data resources with an unprecedented effort of real time data collection across all 50 states' three major political institutions (legislature, governor, and courts of last resort). By linking data on actors, alternatives, and decisions with their outcomes, true evidenced based decision making is possible with our tool.

ROAD MAP

The team has a goal of entering beta testing by Fall of 2021. To do so, by the *data curation* team is to have collected all materials for governors, legislators, courts, statewide actors by the end of Summer 2021. The *open data team* is to have linked all entities across data types, and merged with extant data portals by Spring of 2021. The *data analytics teams* will have fully implemented all appropriate R packages, and have developed and tested analytic tools beyond what R packages facilitate by Summer of 2021. The *data access team* will work with the data curation and data warehouse teams to build the knowledge graph, migrate data into the knowledge graph by late Summer 2021, and develop the multi user interface by Winter 2021. *Track Success* led by PI Windett will work with track partners to integrate data and link OKNs and merge federal court data with State Court data. This will be done right before the beginning of Beta Testing. After Beta Testing, and responding to user feedback – the subsequent push will be to extend the OKN to local levels of government including cities, counties, and school districts.

PARTNERSHIPS

Our partnerships, set up to maximize complementary benefits to our project, and also to ensure Track Success are the following:

Open States is an organization that aggregates legislative information from all 50 states, Washington, D.C., and Puerto Rico. Open States is working with the F-DASH team to adapt existing tools to state courts and governor websites. This partnership allows us to conduct dynamic data collection as well as build upon their established data collection efforts to significantly reduce our time spent in scraping data.

Roper Center for Public Opinion Research at Cornell University specializes in public opinion surveys. Our partnership will develop a database of approximately 40,000 survey questions and responses from the 50 U.S. states from 1970 to the present, to assess the impact of public policies (e.g. *What is the most important problem facing schools today?*). Each of these surveys has been selected for preservation by meeting strict methodological and disclosure requirements, including use of probability-based sampling techniques, minimum sample sizes, and availability of complete question wording with all response options.

National Opinion Research Council is an objective, non-partisan research institution specializes in public opinion polling. NORC will develop numerous instruments to gauge public responses to issues facing their state, as well as elite opinion surveys of attorneys and judges. This survey is collaborative with the Northwestern Open Access to Courts Records team based on an original dataset of attorneys who have argued cases in front of federal courts of appeals, district courts, and state courts of last resort.

The Center for American Women and Politics (CAWP) is nationally recognized as the leading source of scholarly research and current data about women's political participation in the United States. Its mission is to promote greater knowledge and understanding about the role of women in American politics, enhance women's influence in public life, and expand the diversity of women in politics and government. CAWP has collected validated data on the gender of all candidates for state legislative offices and higher. This data will be linked with current election returns, institutional behavior, and decisions made by individual officials.

INTELLECTUAL PROPERTY

Initially, only documents and data that are freely available to the public will be used in the project. Additional information, including proprietary information of some parties, may be accessed via the Data Warehouse Level under appropriate agreements to be negotiated in the future.

CREDIBLE OPEN KNOWLEDGE NETWORK (COKN)

Chengkai Li
cli@uta.edu

Dolores Albarracin   **Yinghui Wu**
Xiaojing Liao   **Jun Yang**

OVERVIEW

In a world where decision making is heavily driven by data, lack of information credibility leads to a populace misled and divided by widespread misinformation and algorithmic decisions that are irrelevant or even catastrophic. Our team is building framework and tools to assist software developers and domain experts to ensure the credibility of their decision-making software powered by open knowledge networks (OKN). Applying our technologies and tools to practical domains, we construct knowledge graphs and software to (a) assist healthcare professionals and app developers in mitigating health misinformation and (b) help cybersecurity professionals in accurately assessing and mitigating software vulnerability.

DESCRIPTION

Credibility as perceived by information recipients stems from both "objective" and "subjective" factors. Objective credibility involves the factual accuracy of information, and subjective credibility involves perceptions that make information convincing to recipients. To improve objective credibility, we will develop data veracity tools for repairing incorrect and missing data, and we will develop data provenance tools that explain query and analytic results. To improve subjective credibility, we will develop strategies to contextualize results from queries and analytics according to information recipients' needs and profiles, and we will develop communication strategies to present such results in a way that is likely to be judged credible by target recipients.

In addition to providing a novel framework of credibility and a suite of tools that can be instantiated in a variety of domains, we directly apply our tools to produce credibility-aware knowledge graphs and software in two highly important domains: cyber threat intelligence and

mitigation of health misinformation related to vaccination hesitancy and to the COVID-19 pandemic. Furthermore, we will collaborate with several other teams (A6677 GIS knowledge, A7017 chemicals and materials systems, and A7115 civil infrastructure) to apply our framework and tools in their domains.

Without being addressed, credibility challenges will become a road block to the ambition of OKN. Opening up the knowledge network also opens the door to low-credibility decisions, since an OKN without credibility assurance may lead to inaccurate or even disastrous decisions. Outcomes from our ongoing project revealed critical misinformation in high-profile security vulnerability repositories such as the National Vulnerability Database, which is used for assessing the security of applications deployed on Amazon Web Services.

DIFFERENTIATORS

Decades of research and development in computing technologies such as data extraction, integration and cleaning, semantic web, database query, and security has resulted in vast experience in tackling important challenges in creating an OKN. However, little has been done to systematically ensure the credibility of software powered by knowledge graphs. This project will complement the aforementioned advancements and together will help realize the vision of OKN.

Ensuring credibility is imperative for all the convergence accelerator projects. While other teams are developing technologies for intelligent decision making based on OKNs in various application domains, our cross-cutting technology is an important enablement of the vision of OKN.

There have been many efforts in studying information credibility and mitigating harmful misinformation, especially in the context of news and social media,

including highly influential works by members of this team. However, except for methods of repairing erroneous data and augmenting incomplete data, there is no prior state-of-the-art for ensuring the credibility of knowledge graphs and software that use such data.

ROAD MAP

The project so far has produced initial versions of credible knowledge graphs and decision-making tools for health misinformation mitigation and cybersecurity. We have also created a public dashboard and a browser plugin for mitigating COVID-19 related misinformation.

The Phase II tasks will be carried out in parallel from onset. The anticipated completion dates of key milestones and deliverables are as follows:

Month 6: Framework of credibility and data model.

Month 9: Knowledge object collection on cyber threat intelligence and health misinformation.

Month 13: Tools for objective credibility.

Month 15: Tools for subjective credibility.

Month 15: Software for health misinformation mitigation and software vulnerability prioritization.

Month 15: Adapted tools to assist other teams in producing credible decision-making systems.

Month 18: Evaluation of deliverables with partners.

Month 19: Refined datasets, tools and software based on the evaluation results.

Month 21: Final evaluation with partners.

Month 24: Releasing datasets, software and website.

PARTNERSHIPS

Our interdisciplinary team has significant critical mass: 1) leading computer scientists in data-related areas; 2) renowned social scientists in credibility, persuasion, psychology, and communication; 3) domain experts and partners in healthcare, vaccine, health misinformation, and cybersecurity; 4) collaborators from the Army Research Lab, Qatar Computing Research Institute, and collaborators with major contributions to the largest knowledge graph products at Amazon, Google and IBM, and to Google's fact-checking products.

The project is informed by interviews, datasets and use cases from partners such as Fortiphyd Logic, Inc., Siemens,

Metafact, FactMata, and Duke University Health System, as well as the aforementioned Track A teams. The deliverables will be evaluated and deployed by these partners. A preliminary version of our vulnerability severity prioritization and assessment tool has been deployed by Fortiphyd Logic to mitigate security risks in industrial control systems.

INTELLECTUAL PROPERTY

This project is not using any proprietary technology. We are mostly using publicly available data. For development and evaluation in industrial environments at our partners, steps involving their proprietary data will be directly performed by them. With regard to intellectual property produced in the project, we aim to open source all tools and software and make all datasets produced in the project available to the public.

ADDITIONAL INFORMATION

https://idir.uta.edu/cokn/

LINKING THE OKN TO THE WEB WITH END-USER PROGRAMMING

Sarah Chasins
schasins@cs.berkeley.edu

Rastislav Bodik
Kyle Crowder

OVERVIEW

Our AI engine learns from users as they browse data they want to collect from webpages—e.g., from Google, Craigslist, Facebook. It writes complex data extraction algorithms automatically, giving users the data they need to make decisions and do their jobs, even if they don't know how to write computer programs. We're using this technology to advance important societal goals, including helping the Seattle and San Antonio Housing Authorities set their housing voucher policies to help low-income families move to high-opportunity neighborhoods.

DESCRIPTION

Imagine you work for the housing authority in a city with soaring rents. You've been tasked with deciding the reasonable rent for each neighborhood. We need to know the fair rent so we can offer housing vouchers in a way that lets low-income families move to high-opportunity neighborhoods.

What do you do? You check government datasets and realize the rent data is out of date, incomplete, or summarized at the city level instead of the neighborhood level. You realize Craigslist has the data you need, but how will you get the data out of its thousands of webpages and into a spreadsheet you can analyze? Today, data scientists who don't know computer programming can't get this data— unless they're willing to spend 10 hours a day in front of the browser, copying and pasting.

This is where our AI engine comes in. We've developed a tool that observes how the user collects web data manually, then takes over to collect the next thousands or millions of rows. After our tool runs, you have the Craigslist listings for the whole city, and you can do your analysis to set housing voucher policy. Our team is already working with three housing authorities on exactly this problem.

If you want to write a data collection program in the time it takes to drink your coffee, set this document aside and download our browser extension. You should expect to finish writing your first web data collection program about *ten minutes* from now.

By the end of this Phase II project, your second cup of coffee should get you all the way to (1) linking your new dataset with the existing datasets in a vast knowledge network and (2) creating a dashboard for visualizing your chosen datasets.

The live nature of web data lets us answer questions we can't answer with traditional, static data sources— surveys, censuses. Our tools unlock these datasets without the engineering effort that only for-profit companies can afford. Our *Extensible OKN* democratizes access to data and lets non-coders turn the web into a ready-to-analyze dataset that yields new knowledge and improves decision-making.

DIFFERENTIATORS

First, we connect the OKN to the web, contributing the dynamic live datasets that users demand. Second, our team brings together world-leading authorities in *automatic programming* to make OKN tasks accessible to non-coders. Finally, our existing relationships with policymakers let us achieve broad impact immediately.

Live Datasets. It is clear that users want and need live, dynamic datasets for real data-driven decisions. We are confident that integrating web data will make our OKN popular with users and useful for a much broader array of tasks.

World's Top Experts. Our team brings together world leaders in *automatic programming*, a set of techniques for generating programs based on examples, demonstrations,

and other inputs that non-coders can provide. In particular, our team includes the foremost experts in technology that combines automatic programming and Human-Computer Interaction (HCI), the subfield of computer science that focuses on making computing tools learnable and usable. Our team's prior work on automatic programming+HCI produced the state-of-the-art Helena web data collection tool. We are building on Helena to implement our Extensible OKN. With only a handful of automatic programming experts in the world, this confluence of expertise is hard to replicate.

Immediate Impact. Our team is a tightly integrated combination of a technical team and an application-focused team that spans multiple social sciences. Our application-focused team uses our technologies for key policymaking purposes such as improving public health and social services. Because of our track record of using cutting-edge technologies to shape policy decisions and our existing close collaborations with policymakers, we have already had the chance to start test-driving our tools for real policy challenges. This ensures our tools meet real user needs and can tackle high-impact challenges. Our Phase I public sector partnerships have already proven that our tools help policymakers improve program outcomes.

ROAD MAP

Our work will produce these key deliverables:

- **Data Collection:** Easy-to-use tool for extending the OKN with dynamic datasets collected from the web. *(Oct 2020)*

- **Data Dashboards:** Easy-to-use tool for building dashboards to visualize incoming data and communicate data analyses. *(Dec 2020)*

- **Data Integration:** Easy-to-use tool for linking new datasets (static or dynamic) to other OKN datasets. *(May 2022)*

- **Data Quality Assurance:** Easy-to-use tool for building trust in an OKN dataset, monitoring data quality over time. *(Aug 2022)*

- **Policy-Making OKN:** OKN datasets for supporting data-driven public policy decisions. A full-fledged knowledge network created with our own easy-to-use tools. *(Aug 2022)*

PARTNERSHIPS

Housing Authorities: Our work with the Seattle, King County, and San Antonio housing authorities focuses on

(i) particular policymaking decisions, e.g., the housing voucher design question, and (ii) how we can design our tools to support data-driven decision making in policy and government settings.

Apple: Our ongoing work with Apple focuses on how we can design our technologies to support dataset creation in industrial settings.

Google, Amazon, and Apple Knowledge Graph Teams: Our partners from the top knowledge graph teams in the world serve as our expert advisors, pointing us to any problems we neglect and steering us away from solutions they've tried and discarded.

INTELLECTUAL PROPERTY

Our existing tools are open-source software, built on top of open-source software. We will open-source the tools we develop throughout Phase II (BSD license). The animating goal of our IP strategy is to encourage the growth of our developer community and user community, so that our tools will remain sustainable long after the end of this award.

KnowWhereGraph ENRICHING & LINKING CROSS-DOMAIN KNOWLEDGE GRAPHS VIA SPACIALLY-EXPLICIT AI

Krzysztof Janowicz
janowicz@ucsb.edu

Pascal Hitzler   **Dean Rehberger**
Mark Schildhauer   **Wenwen Li**

OVERVIEW

We are developing a cross-domain knowledge graph and AI-based GeoEnrichment services as forecasting and investigatory tools to provide analysts with data on-demand. While we have broad potential applications, our initial targets are decision-makers and data scientists working in the food systems and supply chain sector to enable them to forecast environmental impacts on their products from the health of their soil to the sustainability of the products on the consumer's plate.

DESCRIPTION

Geographic Information Systems (GIS) are a prime example of an integration technology that demonstrates how knowing *when* and *where* things happen is crucial to understanding *why* they happened or will happen. GeoEnrichment services are a powerful component of enterprise GIS that provide location-driven, on-demand enrichment of an analyst's study area. Today, a majority of time invested during a GIS project is still spent on data retrieval, entry, cleaning, and apportionment instead of the actual analysis. GeoEnrichment addresses this problem by giving access to curated and GIS-ready information such as demographic data. While highly successful, these services do not provide open cross-domain data beyond pre-defined categories, cannot handle interlinked data, and offer limited support for data integration. Our project will address these challenges and re-envision GeoEnrichment by combining it with an open knowledge graph that makes AI-ready data at the human-environment interface available to diverse stakeholders to utilize and to grow.

This will be made possible by the team's spatially-explicit machine learning models for graph summarization and for predicting missing graph links. Our initial focus will be on food systems, including perspectives from soil health, labor, climate, and supply chains. We will enable the food industry and grocers to encourage farmers to invest in soil sustainability measures such as cover crops and reduced tillage to ensure the sustained yield of their fields while preserving environmental quality. As a technology-driven project, our goal is to demonstrate how the extraction of meaningful features, i.e., independent variables, from our graph will inform downstream models in applications by industry, nonprofits, and government agencies.

DIFFERENTIATORS

Spatial is special. Our team provides unique expertise in representing and integrating geo-data using knowledge graph technologies and GeoAI-based services: We have contributed to international semantic standards; our partner Ersi offers the market's most advanced spatial analytics technology, while Oliver Wyman, PCA, and In10T provide expertise in applying remotely sensed imagery and machine learning models to global food markets, supply chains, and farms. The academic team has a track record in technology transfer, the development of vocabularies, and lifting geo-data to the graph, while our partners NCEAS, USGS, and USDA are among the largest providers and integrators of geo-data. This combination enables us to rapidly develop our enrichment services, connect them to vast amounts of data, test them from within a leading GIS, and apply them to food systems and supply chains to simulate the effects of weather, soil, and crop production on food sustainability and efficiency in consultation with the Food Industry Association and humanitarian aid partner Direct Relief.

ROAD MAP

Year-1 milestones: *(M1.1)* A knowledge graph with data at human-environment nexus, including remotely sensed imagery and outputs of scientific models. *(M1.2)* Published semantically-enabled schema for data alignment and

deduplication *(M1.3)*. A GeoEnrichment prototype (MVP) integrated into Esri's ArcGIS. *(M1.4)* Prototype and graph tested with food systems and supply chain pilots by mining predictive features. Existing GeoEnrichment tools will be used as a baseline to evaluate success. **Year-2 milestones:** *(M2.1)* Refined graph and schema with contributions from external partners. *(M2.2)* Enhanced services based on feedback, including end-user interfaces. *(M2.3)* Broadened Geo-Enrichment beyond GIS by developing APIs for dashboards and question- answering. *(M2.4)* Deployment to broader applications and follow-on use cases. *(M2.5)* Final services and graph; preparation for public-private partnership.

PARTNERSHIPS

Academia Partners

UCSB's Center for Spatial Studies, National Center for Ecological Analysis and Synthesis, and Ecoinformatics (NCEAS), Climate Hazards Center *(new for Phase II),* **KSU's** Center for AI and Data Science, and ASU offer expertise in knowledge engineering, GeoAI, data synthesis, and in modeling environmental effects on crop production. **MSU** contributes expertise in precision and digital agriculture and complements our data with a historic perspective on soils and slavery (Matrix). **USC** *(Phase II)* will contribute expertise in environmental economics with a focus on air pollution on labor.

Industry Sector Partners

Esri will provide expertise for developing and testing the knowledge graph-based Geo-Enrichment services. **Oliver Wyman** will test project capabilities with its customers in supply chain optimization and commodity markets. **Princeton Climate Analytics** and **In10T** (Phase II) will show the usage of the graph and services within their environmental predictive intelligence and deploy to farms.

Nonprofit Partners

Food Industry Association: *(Phase II)* will provide access to the food industry with a focus on sustain- able agriculture. **Direct Relief** *(Phase II)* will apply project work to humanitarian aid supply chains.

Government Partners

USGS will provide expertise in lifting its National Map portal data to the graph, while **USDA ARS & NRCS** will provide data and expertise about sustainable agriculture and soil data.

INTELLECTUAL PROPERTY

We are committed to openness and will release products under a permissive BSD-3 license for the software, and CC0 or CC-BY for data. We will adhere to semantic technologies standards by W3C and OGC. While encouraging the least restrictive licensing, data or software licensed under more restrictive conditions will be accommodated to allow for broad participation from the industry.

KNOWLEDGE NETWORK INFRASTRUCTURE WITH APPLICATION TO COVID-19 SCIENCE AND ECONOMICS

Michael Cafarella
michjc@umich.edu

Matthew Shapiro
Oren Etzioni

OVERVIEW

Knowledge Networks are a novel and potentially transformative form of data, but building applications on top of them is too difficult, time-consuming, and expensive. We will build a Knowledge Network Programming System that makes it far easier to build novel knowledge-powered applications, while also improving the knowledge resources themselves.

DESCRIPTION

Knowledge Networks like Wikidata are a compelling new type of data — akin to a "structured world wide web" — that have enabled new applications, such as structured web search and voice assistants. Unfortunately, only the most technically sophisticated organizations have had the resources to build these difficult-to-engineer applications. As a result, most of these next-generation applications never actually get built, and users cannot benefit from them. Our Knowledge Network Programming System will dramatically reduce the cost of building knowledge-powered applications by combining programming tools with recent advances in data management and machine learning. This will happen in three ways. First, it will make Knowledge Network data items easy to use by integrating it directly into the Python language, much like integers and strings. Second, it will make knowledge applications easy to debug by automatically recording all data operations for future examination. Third, it will make knowledge application development more collaborative by making data sharing a basic language primitive, thereby including in the development process not just programmers, but also crowd workers, data field workers, analysts, and even news article readers: anyone who creates, manipulates, or consumes knowledge can make an application better.

We will test the system using several Knowledge Networks developed as part of this project. The first is the CORD-19 network, which describes 55K scientific papers on COVID-19 and related historical coronavirus research. *This is already public* and has been covered in the Wall Street Journal, the New York Times, and elsewhere.

The second is a network that describes macroeconomic statistics in the United States, including recent budgetary and economic responses to COVID-19. We believe these networks will be useful for our project at the same time they support national priorities.

Concretely, this project will yield new software, in the form of the programming system and toolset. It will also yield novel data resources, in the form of the above Knowledge Networks. Finally, it should yield novel applications that both illustrate the programming system and are useful on their own.

DIFFERENTIATORS

All Knowledge Network applications that we are aware of rely on traditional software engineering tools. We are unaware of any system that addresses application development per se, even though other data types (say, relational databases) have extensive dedicated tooling. The programming system is unusual in its application of data management and machine learning methods to goals traditionally associated with programming languages.

The Knowledge Networks help with validating our system and with crucial social, but are especially notable for the agility with which we can create them. *The CORD-19 Knowledge Network was released in March, 2020.* This was possible thanks to infrastructure work that was funded as part of the initial phase of this program.

The research team is unusual in its level of experience with shipping data development systems, knowledge network production, and relevant domain expertise. PI Michael Cafarella is an associate professor of Computer Science at the University of Michigan. He has published on databases and is one of the co-creators of the Hadoop system. PI Oren Etzioni is the CEO of the Allen Institute for Artificial Intelligence. The Allen Institute is a nonprofit research organization that arguably employs the largest set of Knowledge Network engineers outside a major tech firm. PI Matthew Shapiro is the Lawrence R Klein Collegiate Professor of Economics at Michigan, and an expert in macroeconomics and public finance. He serves as chair of the Federal Statistics Advisory Committee.

ROAD MAP

There are several core deliverables for this work: (1) the programming system, (2) the scientific and economics Knowledge Networks, (3) infrastructure used to produce those networks, and (4) application code built to demonstrate the programming system.

The programming system is already under development at the University of Michigan and a first component has been released. The Allen Institute has released an initial version of the CORD-19 Knowledge Network and continues to update it. The economics Knowledge Network is under development at Michigan. The infrastructure software is under development; we have front-loaded our budget to accelerate this work so it can be useful to other teams during the NSF program. All new code will be committed continually to an open-source repository, and turned into a minor release roughly every three months. Upon each programming system release, the production pipelines for Knowledge Network will be modified to fit the latest system features.

In the long term, we believe the system potentially has many users in the scientific, policymaking, and corporate worlds. We are collaborating with large institutions and a venture capital firm to ensure the project has users and funding beyond this program.

PARTNERSHIPS

The University of Michigan will focus on building the programming system itself, as well as producing the Economics Knowledge Network. In Phase 1, Michigan produced a beta version of the programming system software stack, publicly available via a repository at GitHub.

Additional Senior Personnel from the University of California, Berkeley at the University of Washington, Seattle will work on front-end software for adding data to a Knowledge Network.

We are working with researchers in the natural sciences and social sciences --- plus several large technology firms --- to evaluate the system.

INTELLECTUAL PROPERTY

Code and data created for this project will be released into the public domain. The only exceptions will be when some purchased source datasets (say, containing certain economic statistics) have restrictions that prevent us from doing so.

TRACK B

TEAMS

TRACK B TEAMS | Table of Contents

TRACK B1
TEAMS

TRACK B1

The NSF Future of Work at the Human-Technology Frontier (FW-HTF) Big Idea seeks to respond to the challenges and opportunities for the future of jobs and work. FW-HTF supports convergence research to understand and influence the impact of artificial intelligence on workers and work, understand and develop the human-technology partnership, design new technologies to augment human performance, illuminate the emerging socio-technological landscape, understand the risks and benefits of new technologies, and foster lifelong and pervasive learning.

The AI and Future of Jobs track will support the development of mechanisms that connect workers with jobs of the future, such as predictive artificial intelligence tools, economic and labor market analyses of needed skills for future workplaces, and educational technologies needed for adult learning. Ensuring fair and ethical treatment of workers will be a key principle for the tools, resources and activities associated with this effort. Projects may be focused on particular industries or regions, specific populations such as veterans, or particular workplace types such as small businesses, manufacturing, or K-12 schools.

TOWARD FAIR, ETHICAL, EFFICIENT, AND TRUSTWORTHY CROWDSOURCING TO SUPPORT JOBS OF THE FUTURE

Chuan Yue
chuanyue@mines.edu

Benjamin Gilbert   **Monchu Chen**
Qin Zhu   **Lei Ding**

OVERVIEW

People who do piece work online often have lost traditional jobs, have disabilities, or cannot work outside of the home. They are invaluable, though, to researchers in different disciplines, government agencies, aid organizations, and companies, especially with the surge of AI applications. We are creating a fairer and more ethical, efficient, and trustworthy online labor market to support vulnerable workers and connect workers to more sustainable and rewarding future jobs.

DESCRIPTION

On crowdsourcing platforms, job requesters post jobs, and online workers complete the jobs and get paid. Job requesters are often researchers conducting important studies, government or humanitarian organizations responding to crises, and companies constructing high-quality training datasets for AI applications. Crowdworkers are often people who could not be employed in traditional workplaces, or who have lost their jobs, have significant family responsibilities, or have health problems. They are vital for the prosperity of the U.S. economy, humanity, and society.

Crowdsourcing has created a vast and rapidly growing online labor market. However, today's crowdsourcing platforms cannot well support crowdworkers, job requesters, and the healthy growth of this important online labor market due to a number of problems that intertwine from the *fairness, ethics, efficiency,* and *trustworthiness* perspectives. For example, workers are often significantly underpaid or mistreated with their task submissions rejected; job requesters and workers can easily fail to act with honesty and respect towards each other, or be accountable for their actions; workers cannot efficiently identify and complete tasks or

improve their skills; and mutual trust is lacking between workers and requesters.

Our project will change this. We benefit from a convergence of the research and development from multiple intellectually-distinct disciplines including Computer Science, Economics & Business, and Humanities & Social Sciences. By performing fundamental research with rapid development advances through partnerships with crowdsourcing platform providers and customers, our project will deliver client-side and server-side techniques that can be used to create *fair, ethical, efficient,* and *trustworthy* (FEET) crowdsourcing platforms. These platforms will support millions of American workers for jobs of the future, and for connecting them in both traditional and gig economies. Our deliverables will also enable job requesters to receive high-quality and trustworthy task submissions for them to confidently conduct important studies and make crucial decisions.

To achieve this, we will: (1) build incentive structures based on economic theory to increase fairness in crowdsourcing, (2) build interactive user guides and training modules to incorporate ethics into crowdsourcing, (3) build task description enhancement and task recommendation AI models to improve crowdworkers' efficiency, (4) build security and privacy risk detection AI models to protect both crowdworkers and job requesters, (5) build algorithms to optimize human-AI collaboration in crowdsourcing and support crowdworkers for their skill growth needs, (6) build interfaces to connect crowdworkers to traditional jobs and traditional workers to crowdsourcing jobs, and (7) integrate our techniques on the client-side into a web browser extension, and on the server-side into crowdsourcing systems.

DIFFERENTIATORS

In this project, we uniquely identify and investigate four categories of challenging problems of crowdsourcing that intertwine in a complicated manner, along with many other factors such as the diverse language, cultural, and educational backgrounds of crowdworkers and job requesters.

We are the first to take a *convergence approach* to address these significant socio-technical problems of national importance. These problems can be addressed only by a convergence of the research and development from the multiple distinct disciplines represented in our project team.

From the track success perspective, we believe that *future jobs must be human centered* — workers should be able to flexibly, sustainably, and enjoyably work either at home or in the field. We know that connecting gig and traditional economies will be critical for workers and jobs of the future. Our project is in a unique position to enable this connection as we are the only Track B team focusing on the gig economy and online workers.

ROAD MAP

We will: (1) start to build all our proposed techniques or models listed above and complete half of them in Year 1, (2) start to intensively collaborate with other Track B teams in Year 1 for track integration and success, (3) complete the building and evaluation of all our proposed techniques or models in Year 2, and (4) work intensively with other Track B teams on the wide deployment of our and their developed techniques or models in Year 2.

PARTNERSHIPS

Appen Limited, a leading crowdsourcing and AI service vendor, which acquired the popular Figure Eight crowdsourcing platform in 2019, is a major partner in both phases. In Phase I, we worked with Dr. Monchu Chen, Director of HCI and Machine Learning at Figure Eight (now Principal Data Scientist at Appen), on trustworthiness of crowdsourcing and prototyped two security and privacy detection mechanisms. Accenture Labs (Arlington, VA), a leading R&D lab on discovering innovative solutions for both Accenture and its clients, is another major partner in both phases. In Phase I, we worked with its R&D Principal Dr. Lei Ding on their crowdsourcing service needs. Drs. Chen and Ding have joined our Phase II PI team.

We will work with stakeholders from industry, institutions of higher education, non-profits, and government entities, as well as crowdworkers in Phase II for evaluation or providing feedback.

INTELLECTUAL PROPERTY

All the deliverables including techniques, mechanisms, models, and the corresponding code developed in this project will be freely shareable under open source licenses to promote their wide deployment and use.

PERSONALIZED AI GUIDANCE FOR SEAMLESS WORKFORCE TRANSITIONING

Xiaoli Zhang
xlzhang@mines.edu

Melissa Walden   **Sid Saleh**
Sebnem Duzgun   **Sridhar Seetharaman**

OVERVIEW

The Personalized AI Guidance Ecosystem (PAIGE) is a seamless workforce transition system for the future of work that:

- guides job seekers through fast and fair job transition with an interactive, explainable, adaptive AI approach

- helps employers select best-fit candidates in an accelerated, cost-effective manner

- identifies required training to fill initial skill gaps and to upskill over time for long-term job fulfillment and growth

- provides scalable programs with higher impact for governments and education providers

DESCRIPTION

Our vision is that through human-centered, scalable AI, the PAIGE workforce transition tool will assess job seekers not only on the basis of their skills, education, and previous experience but also on the basis of their personalities, interests, and future goals. With this comprehensive assessment, individually tailored training recommendations will be generated for job seekers for fast and successful transition into best-fit jobs (Fig. 1).

PAIGE will also enable employers to recruit the most qualified employees with the highest potential for long-term success. Additionally, training providers will be able to customize their programs to specific individual/job combinations, which will improve training outcomes.

Our Phase II testbeds are two workforce groups that are significant to the Colorado economy: transitioning military personnel and displaced workers from the fossil fuel industry.

The first testbed, veterans, make up nearly 10% of Colorado's population, and both unemployment and
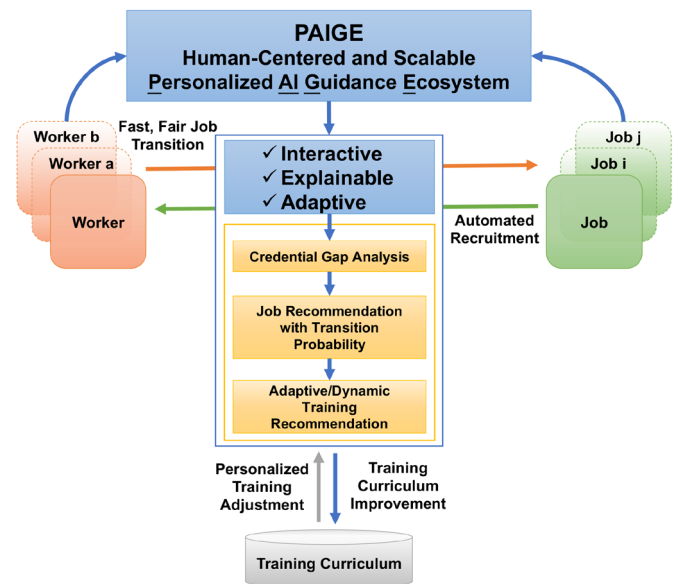
Figure 1. Program vision

underemployment are existing problems. The second testbed includes workers from Colorado's mines and coal-fired power plants, which are threatened by the low price of natural gas, the expanding market penetration of renewable energy, and increasing automation. The challenges faced by these groups are mirrored in workforce populations and industries across the U.S. and support the broader impact potential of our tool.

With our convergence-focused partners—public, private, and nonprofit entities focused on veterans and/or the energy sector (Fig. 3)—we will first build and train the PAIGE tool with our testbed groups and then scale the tool to workforce and employer groups throughout the nation. The substantial impacts of the COVID-19 and oil crises underscore the need for a fast, fair, cost-effective, and scalable workforce transition tool such as PAIGE.

DIFFERENTIATORS

PAIGE will improve the current state of the art of workforce transition systems by combining comprehensive individual assessment, explainable job recommendation, and adaptive training recommendation features in one easily accessible, user-focused tool. Additionally, it will build workforce transition knowledge for the future of work to support scalable, cost-effective, and personalized job and training recommendations that, given the quantity of training needed, are not currently feasible for traditional tutoring/training programs. The PAIGE tool will fill the gap between workers and current job or resumé listing organizations by providing a new mechanism of workforce guidance and assessment methods. It will also advance the understanding of job transition and skill retraining by revealing the subtle and complex links between worker attributes and job requirements that are beyond human interpretation.

ROAD MAP
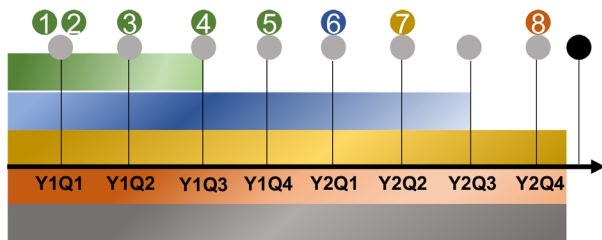
Task 1: Human-Centered AI for Personalized Workforce Transition
1. Featurized definition of job reqmts and worker skills
2. Online interactive user interface
3. Algorithms/software for competency gap analysis
4. Algorithms/software for explainable job rec.
5. Algorithms/software for adaptive training rec.

Task 2: Scalability, Sustainability and Understandability of Personalized Workforce Transition
6. Graph, modular workforce knowledge

Task 3: Workforce Transition Ecosystem Model
7. Ecosystem framework w/ business models

Task 4: Validation/Evaluation
8. Human subject experiment results and analysis
● Continuous updates and expansion to new domains

Task 5: Strategic, Convergent Activities
◉ Workshops, advisory board meetings, and other quarterly events

Figure 2. Tasks and deliverables

PARTNERSHIPS

Our partners include Texas A&M University, the National Renewable Energy Laboratory, EPRI (a nonprofit research organization in the energy sector), Colorado Community College System, Colorado Workforce Development Council, Jefferson County Business and Workforce Center, Colorado Department of VA, consultants from LinkedIn, and eight Track B1 teams. Our partners will contribute

to building, training, and deploying PAIGE by supplying resources such as data sets, testing evaluators, and training platforms. They will also form an advisory board to help chart a course for commercialization and future expansion.

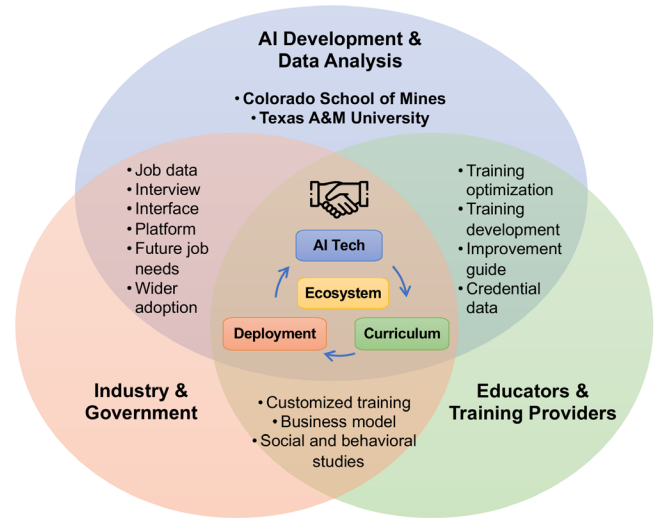
Figure 3. Team strength

INTELLECTUAL PROPERTY

Phase II will culminate in a patentable deliverable: the PAIGE workforce transition data hub. IP includes the AI algorithm as well as the tool interface. The Colorado School of Mines Office of Technology Transfer excels in negotiating joint IP with research partners and in working with industry to bring research-inspired discoveries to market.

MIDWEST COMPACT FOR CONNECTED ECOSYSTEM

Jeff Grann, Ph.D.
jgrann@credentialengine.org

Ken Sauer, Ph.D.

OVERVIEW

Our project strengthens America's learn-and-work ecosystem by assembling a critical mass of linked open data using Credential Engine's (CE) mature, open technologies including the Credential Transparency Description Language (CTDL) schema and Registry. We use a state agency strategy, partnering with the Midwestern Higher Education Compact (MHEC) and several national organizations, to harness their collective power to develop and sustain AI-powered applications supporting learners of all ages, employers, educators, and other stakeholders.

DESCRIPTION

The U.S. learn-and-work ecosystem is complex with 730,000+ credentials (e.g., degrees, licenses, certificates, industry certifications, badges) and inadequate systems to enable learners and employers to effectively differentiate amongst these credentials to make informed choices. Key data for identifying which credentials lead to high-wage, high-growth careers too often resides in inaccessible silos. We intend to integrate important and helpful credential data into a useable form for making decisions, such as students' career exploration, displaced workers' enrollment for upskilling training, and employers' assessment of applicants' suitability for a job. We do this by convening state agencies and national partners through MHEC to establish a critical mass of public linked open data about credentials and sustain these new information systems over time.

Our objective is within reach. The CE Registry holds data as linked open data on the web serialized as JSON-LD using the rich CTDL schema. CE has refined these technologies over 5+ years, partnering with hundreds of organizations to publish thousands of credentials and hundreds of thousands of unique data points to the web as linked open data that is human and machine readable. CE will expand its publishing technologies to provide a new staging service for combining data from multiple sources and further develop the CTDL to address prioritized use cases (such as, transfer and quality assurance services). CE's current primary users are state agency leaders responsible for education, employment, and economic outcomes. Indiana prototyped this work first and models what is possible for CE's 15 current state partners across the U.S.

Our project uses proven methods for interstate convenings to develop midwest state agency data systems and empower more than 35 million workers to evaluate credentials and advance their careers. We partner with MHEC, agencies in all of its 12 midwest states, several national organizations, and Lumina Foundation. The six states already engaged will accelerate their plans and help others get started. Activities include analyzing states' data systems, defining requirements, and supporting state and regional priorities. Progress will be assessed via quantitative reports and an annual survey of governors' offices about progress toward credential transparency and a connected learn-and-work ecosystem. Our approach positions us to scale nationally through established agreements with the three other interstate compacts and through master business contracts for technical services such as AI-powered applications.

Learn more https://bit.ly/ConnectedEcosystem

DIFFERENTIATORS

This approach is the best choice among competing options because it exemplifies the highest level of convergence research. CE is uniquely positioned as a non-profit with a mission to improve credential transparency and literacy; we bring the required openly-licensed schema, experience, technical expertise, and public/private infrastructure for success. In the past, commercial and federal efforts have

tried to implement piecemeal solutions but failed because they couldn't integrate their solutions across the full credentialing marketplace. CE's robust partner network is activated to develop, grow, and sustain a critically needed "GPS" system for credential transparency via role responsibilities, workflows, strategic plans, and policy.

The MHEC partnership brings a distinctive ability to scale by convening 12 midwest states as a community of practice, along with multiple national organizations. This approach has been used successfully to establish credit articulations for military service and the highly-successful State Authorization Reciprocity Agreements (SARA), a voluntary regional approach to state oversight of postsecondary distance education created in 2012 backed by Lumina Foundation and now supported by 49 states and ~2,000 colleges/universities. We are similarly uniquely positioned to scale and sustain our deliverables via MHEC-negotiated master contracts that meet state procurement requirements, an approach that saved the region at least $29M in 2019. States in the other three regional compacts can also use MHEC master contracts.

ROAD MAP

Q3 2020. Build Midwest Compact, Partner contracts, State agency MOUs. Baseline quantitative metrics. CE releases beta staging service

Q4 2020. Partner discovery, goals, teams, and workplans to publish, connect, and consume data. **Prototype:** Kickoff Compact w/state agency leaders.

Q1 2021. Partner outreach. Test publishing workflows. Feedback on CE staging service. **Milestone:** Data mapping and workflows.

Q2 2021. Milestone: All partners begin publishing to Registry. **Evaluation:** Publish quantitative and survey results in annual report

Q3 2021. Launch CE staging service. **Milestone:** Application requirements defined.

Q4 2021. Document system requirements and RFPs

Q1 2022. Milestone: State and national partners publish scoped data to Registry

Q2 2022. Dissemination for Phase 3. **Milestone:** Application integrations with Registry. Master contracts published. **Evaluation:** Annual report

PARTNERSHIPS

Phase 1 & 2 core project team: Credential Engine, Indiana Commission for Higher Education, Midwestern Higher Education Compact, Lumina Foundation. **User group:** Multiple Indiana state agencies; current CE state partners in Illinois, Kansas, Michigan, Minnesota, Ohio; new state partners in IA, MO, NE, ND, SD, WI. **National publishing partners:** ACE, CareerOneStop, Coursera, Distance Education Accrediting Commission, SkillsCommons, Unizin. **Evaluation consultant:** Guidera Strategy. **Application partners:** AACRAO, Coleridge, Educause, IMS Global, 8 track partners. **Policy partner:** Education Commission of the States. **Other interstate compacts:** NEBHE, SREB, WICHE.

INTELLECTUAL PROPERTY

CE's technologies and data are open, modeled on World Wide Web Consortium specifications for the semantic web. The Compact model for advancing a regional vertical open knowledge network will be openly licensed.

ADDITIONAL INFORMATION

https://luminafoundation.org/campaign/learn-and-work-ecosystem/

COMPETENCY CATALYST

Robby Robson
robby.robson@eduworks.com

Myk Garn   **Jeanne Kitchens**
Ashok Goel   **Elliot Robson**
Matt Lisle

OVERVIEW

Due to digital transformation in the workplace, companies need to reskill their workers frequently and rapidly, but when companies turn to colleges for help, they find it hard to correlate college course topics and programs with the workplace skills they need. This lack of clarity results in inadequate training, time lags, and workers not being reskilled.

Competency Catalyst (C2) will offer a solution to this problem. The solution, called *Skillsync,* is a virtual assistant that enables employers to define their reskilling needs, connects them to colleges, and helps colleges rapidly create targeted reskilling programs that are informed by and aligned to these needs. Skillsync is made possible by a C2 *platform* that converts job and course data from national sources into underlying knowledge, skills, and abilities (KSAs) and provides *alignment services* that allow college users to interpret and meet employer needs. In Phase II Skillsync will be used by human resources (HR) and talent directors, and by continuing education departments nearby colleges to meet the reskilling needs of about 250 existing employees from diverse industries whose jobs are at risk due to modernization. After Phase II we aim to use Skillsync to dramatically increase America's reskilling capacity, while using the C2 platform to catalyze an ecosystem of reskilling applications.

DESCRIPTION

By 2022, an estimated 75 million jobs may be displaced, and 133 million new ones created by digital transformation of multiple industries. Given the high cost of turnover, reskilling current employees is a cost-effective strategy, with many benefits. Reskilled employees are more likely to stay with their company and become high performers. Increasingly, companies are looking to their local colleges,

not just for graduates to enter the workforce, but to also meet their immediate reskilling needs. Yet, when businesses turn to their local colleges for reskilling programs, they feel out of sync. As Daniel Jackson, President of Carroll County, GA Chamber of Commerce told us, "90% of faculty think they are doing okay but only 20% of businesses think colleges are keeping up."

Our research revealed that the root cause of this disconnect occurs when companies describe their training needs to colleges. Companies talk in terms of jobs and skills, but colleges listen in terms of courses and knowledge. This results in misunderstandings, time lost describing and re-describing needs, lengthy course and program revisions, and in training that, when finally delivered, does not cover all the skills that employees need.

The C2 platform uses recently developed natural language processing and machine learning methods to extract KSAs from job descriptions, course syllabi, and related documents and to organize these into "competency frameworks" that employers can effectively use to describe training needs. Similar methods are used to align these KSAs with learning opportunities. Although the proposed Phase II work focuses on reskilling existing employees, the output from the C2 platform will support employee skills assessment, career planning, and other applications being developed by C/Accel projects. C2 will also make the competency frameworks it generates freely available in open formats,

meeting a critical talent pipeline need identified by the U.S. Chamber of Commerce, Open Skills Stack, and others.

skillsync

Skillsync will be designed with extensive input from users and stakeholders in partnership with business, colleges, UPCEA, CAEL (national associations of training providers with hundreds of college members), and the Business Higher Education Forum (BHEF). It will be piloted with companies, who will use it to define reskilling needs, and with colleges, who will respond by developing and enrolling workers in reskilling programs. It will be developed using participatory research and human-centered design methodologies.

To ensure that workers from underserved populations will benefit from the industry-academia partnerships it enables, Skillsync will be designed with input from diverse users and the underlying AI algorithms in the C2 platform will be examined to avoid bias. Both diversity and fairness will continue to be points of emphasis as Skillsync moves towards productization.

DIFFERENTIATORS

Skillsync is the first application designed to align, expedite, and improve company-college reskilling partnerships. As a marquee feature, Skillsync will include an *intelligent virtual assistant* that allows company users to ask questions about college offerings as if they were conversing with a counselor or advisor. This technology, called "Jill Watson," was developed by Dr. Ashok Goel's research team at Georgia Tech and has made national headlines as a virtual teaching assistant. Its capabilities will be expanded via supervised machine learning to answer the questions about college courses and programs that HR and talent directors typically ask.

The C2 *platform* will leverage the open source competency and skills system (CaSS) developed by Eduworks and the Credential Engine technology that plays a central role in efforts to improve America's talent pipeline with open data tools. This will enable it to "plug and play" with these efforts, extending the reach and impact of the essential and novel KSA analysis and competency alignment services it offers.

ROAD MAP

In year one the C2 team will complete construction, testing,

and validation of minimum viable prototypes of the C2 platform and Skillsync. In year two, beta versions will be piloted by eight company/college partnerships and used to create reskilling programs for their workers. By the end of the Phase II grant, Skillsync will be validated, field tested, ready to deploy, and known to key communities of colleges and companies. It will be ready to be productized as a commercial solution for the $170 billion corporate reskilling marketplace. The C2 platform will be integrated with the Credential Engine and be offered as a freely available open data service.

PARTNERSHIPS

The C2 project builds on three types of partnerships: (1) *Development and Data Partners* who will contribute job or course data assist with development of core functionality, including Georgia Tech, the Credential Engine, Open Syllabus, the DXtera Institute, Udacity, and LibreText; (2) *Deployment Partners* who will participate in testing, validation, and dissemination, such as the University of West Georgia, the U.S. Chamber of Commerce Foundation's T3 Innovation Network, BHEF, CAEL and UPCEA; and (3) *Affiliate Partners* who will use the C2 platform and services to support their applications, including other C/Accel projects such as SAFER (B-6656) and DIRECT (B-6915).

INTELLECTUAL PROPERTY

The intellectual property used in the C2 platform and Skillsync is either open source or owned (and in some cases patented) by Eduworks, who will also have full or joint rights that can be exercised without restriction in any new technologies developed as a result of the Competency Catalyst project. The one exception is "Jill Watson," which is owned by Georgia Tech and will be deployed in Skillsync as a service, with commercial use to be negotiated.

PREPARING THE BUILDING INDUSTRY FOR THE FUTURE OF WORK

Shahin Vassigh
svasigh@fiu.edu

Shu-Ching Chen   **Miguel Alonso**
Masoud Sadjadi   **Atorod Azizinamin**

OVERVIEW

Our buildings and infrastructure must operate more efficiently, be built faster and at lower cost, with dramatically reduced environmental impact. Although technology has advanced significantly in the past decades, the US construction industry and workforce have been slow to change in order to meet these demands.

The **Robotics Academy** is a suite of internet accessible tools driving the future of construction — manufacturing buildings and infrastructure in factories with intelligent automated robotic machinery, delivering completed structures to their final destination.

DESCRIPTION

The Academy platform addresses the needs of workers, businesses, and investors through three distinct components. *The Immersive Learning* suite delivers robotics training and educational curricula in Virtual Reality (VR) and Augmented Reality (AR) using immersive headsets and mobile devices. Extensive training is offered in Robotics Hardware programming and operation, Communication Software, Safety Protocols, Safety Standards, Risk Mitigation, and Safety Sensors. The Curriculum is use-case driven, using real-world tasks from construction, manufacturing and material handling environments.

Second, the *Innovation Network* supports technology exchange and innovation. It includes the Robotics Knowledgebase, Research and Development Exchange, the Roboticists Forum, and the Use-Case Showroom.

Third, the *Automation Marketplace* includes the Venture Café, a Crunchbase type exchange for businesses, entrepreneurs, investors, and an employment matchmaking service.

For the over 4 Million construction workers threatened with potential job displacement, the Academy provides training and upskilling to increase income, extend careers, and reduce workplace injuries.

The Academy provides access to financing sources and highly skilled workers for the 3.2 Million traditional construction firms and the 1,365 business currently in the modular building space looking to transform and grow, while experiencing a 1 Million worker labor shortage.

The market for modular construction has the potential to reach $260 Billion. This platform dramatically improves the ability of over 2 Million inventors, 4,000 venture capital and investment banking firms, and entrepreneurs to raise capital and invest.

The robotics Academy will eventually benefit the entire US Population by fostering innovation and increasing the market penetration of modular construction leading to 1) better building design, construction and performance, 2) lower cost homes and buildings; and 3) reducing carbon emissions.

DIFFERENTIATORS

The Academy is the only platform of its kind, for any industry — an integrated one-stop shop for training, innovation, and investment that will immediately be the nation's leading construction industry resource.

The entire platform is powered by AI and unique industry analytics to deliver a robust, and powerful expert system curated personalized content, training, employment and investment services, all accessed through a seamless web-based dashboard.

It delivers high-quality training using real, hands-on experiential learning anywhere, anytime, on any hardware platform, supporting workers, businesses, and entrepreneurs.

The Academy supports innovation and new investment in the construction industry. Its technology and embedded educational and learning theory can be adapted for other industries, and will deliver transformational advances in educational technology, immersive software design, automation technology, and understanding the economic impacts of AI and robotics.

The platform is positioned to become the leading resource for what will become a rapidly growing market, rapidly scale, and use a number of different monetization strategies to quickly become financially sustainable.

ROAD MAP
The Robotics Academy deliverable are as follows:

Immersive Learning: Year one: use continuous improvement and agile software development techniques to build the platform with a minimum viable product ready before the end of year 1. Completed elements include: 1) training curriculum, 2) personalized learning AI algorithm, 3) AR and VR learning platforms, 4) remote robot lab functionality to provide software hardware trouble shooting, 5) platform to connect the affiliate labs for in-person training, 6) testing the learning software incorporating user feedback. *Year two:* 1) finalize, deliver and disseminate the learning software through the project website.

Innovation Network: Year one: complete: 1) the AI knowledge algorithm, 2) the roboticist forum, 3) R&D Exchange, 4) Robotics Wikipedia, 5) the robotics Use Case Showroom, and 5) Newsroom. *Year Two:* incorporate changes and finalize.

Automation Marketplace: Year one: complete: 1) the AI knowledge curation and personalization algorithm, 2) employment matchmaker, and 3) the Venture Café, 3). *Year two:* 1) complete testing and evaluation, and 2) finalize the Automation Marketplace product suite.

PARTNERSHIPS
We will work with the University of Miami and University of Southern California to fully develop the Robotics Academy. Our *Technology Partners* include: 1) *Microsoft,* providing technical support for the integration of Azure AI for cloud computing and voice recognition, and the Hololens 2 for AR functionality, 2) *Xennial Digital,* integrating their educational platform with a user data collection API, 3) ABB Robotics, sharing their industrial robotics training curriculum, 4) SISU, collaborating on developing immersive

courses, 5) RoboDK, providing support for integrating their robotics API into the training software.

Our *Building Industry Partners* include international companies SKANSKA, MOSS Construction, Coastal Construction, Gensler, Perkins and Will, McKenzie, and DPR. They will provide guidance on the training needs of the industry and serve as a source for users during testing phases. Our *Non-Profit Partners* the Beacon Council and CareerSource South Florida will provide expert advice on the development of the training curricula, Employment Matchmaker, and Venture Café products. We will also work with three K-12 major school systems across Florida to test and perfect the training curriculum.

INTELLECTUAL PROPERTY
The IP involved in the current project consists of a machine-learning-based framework for estimating student confidence in their decisions while interacting in an immersive learning environment. It is comprised of three components: the immersive learning environment, the model training environment, and the model inference environment. The IP has been disclosed to the FIU Office of Technology Management and is under review.

SITUATIONAL AWARENESS FOR EMPLOYMENT RESILIENCE (SAFER)

Katy Börner
katy@indiana.edu

Martin Storksdieck   **Kosali Simon**
Xiaozhong Liu   **André Skupin**

OVERVIEW

This project seeks to transform data on millions of jobs and associated skills into powerful tools for data-driven decision support aimed at workers and career coaches. We will develop and deploy personalized apps and comprehensive data visualizations of near real-time labor market data in order to enhance situational awareness of emerging job risks emanating from advances in automation or fast-moving threats, like the COVID-19 pandemic. Workers and coaches will be provided with knowledge and tools that foster increased awareness and a more proactive stance in recognizing impending job risks, transferring existing skills to future-proof jobs, and accessing reliable information on training opportunities to fill skills gaps.

DESCRIPTION

We believe it is of paramount importance for American workers to be better informed about how to chart paths to future-proof, resilient employment. Workers, career coaches, and employment agencies alike need to be empowered by both overall understanding and granular, actionable insights on the evolving labor market and resulting employment risks, whether due to encroachment by automation and AI, or other slow- and fast-moving threats to current jobs and career trajectories.

Focusing on the needs of career coaches and workers, we will: (1) develop and deploy a SAFER app (called *Movahed*, shown on right) and employment maps that let users explore job risks and possible career paths in alignment with self-reported skills, interests, and preferences, (2) scale up training of coaches to prepare them to use the app as part of career counseling services, and (3) to make the data collected via the app available to coaches, so they can guide mentees effectively and serve workers with maximum effect, even and especially in times of increased demand, whether in response to regional and national labor market trends, plant closures, or a pandemic.

The value of observing labor market trends as patterns on geographic maps has long been recognized and the ability to seamlessly zoom from state to regional and local employment patterns is taken for granted in current offerings. Why is it that no such interface — highly visual and interactive — exists for the employment space of skills and jobs? Our project will produce the first-ever, comprehensive map of the employment knowledge space so anyone can navigate the landscape of professions and skills with the familiar ease reminiscent of Google maps. The employment maps will help communicate labor market trends in the context of specific skills and professions, holistic and dynamic. With deployments ranging from web apps to large-format displays at public venues, from libraries to science museums, viewers can be reached, informed, and engaged in diverse ways, and guided to the SAFER app for personal career exploration and job coach referral.
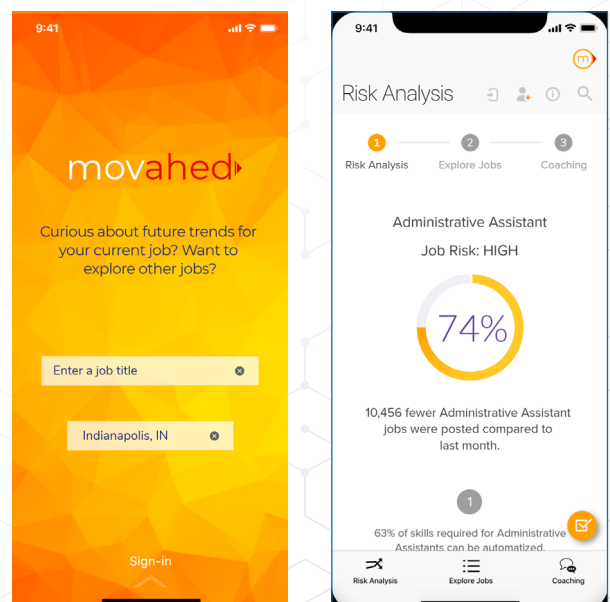
DIFFERENTIATORS

Job sites like Indeed, LinkedIn, and others serve the needs of employers, without ensuring jobs are future-proof. However, job seekers and employees need to carefully consider what job to take next to stay employed in the long run and what re- or upskilling choices to make to secure those jobs.

Most commercially available labor market information data mining and visualization services exclusively serve employers or institutions (who have the ability to pay for these services). We are unique in supporting data-driven career decision making *by individuals* (alone or facilitated though job counseling or coaching)—in a personalized way supported by self-assessments used to understand existing skills, risk aversion and preferences.

SAFER will not compete with existing job boards, but fill glaring gaps, from the *individual level,* empowering workers through personal risk scores and increased awareness of trends affecting their livelihood, to the broader *coaching and management level* – tracking and contextualizing employment risk relative to the skills landscape.

We believe that all individuals can benefit from both personal and technologically mediated career coaching, as very few will work the same job their entire life. Our approach is designed to address behavioral inertia around improving career prospects by individuals in a variety of interconnected ways: by raising awareness of global/local/personalized risks, providing engaging experiences for exploring alternatives, and by lowering barriers for behavior change. Our team is prepared to study—and to develop a theory of change—which population groups benefit from our system in arriving at desirable future career paths.

ROAD MAP

Mid-fidelity mockups of the *Movahed* app have been user tested and will be used in coach training by Skillful Indiana. Feedback will be used to implement the SAFER app and associated training materials by Q3 2020. During Q1 and Q2 of the project, we will produce a comprehensive map of the employment domain, deploy it through an *Employment Map* (see below) website in Q3,
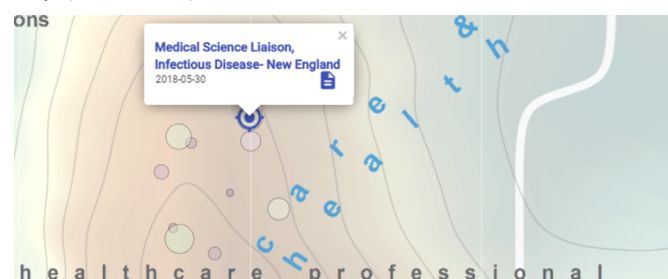
produce interactive educational materials in Q4, and make it available to career coaches and the public.

In partnership with Skillful Indiana and InStride, we will create and field test the app and employment maps as part of a comprehensive socio-technical career guidance support system that involves outreach and awareness raising campaigns, ongoing analysis of data that feeds the app and maps, a support system for coaches, and the beginning of a system for recommending re- and upskilling resources. We will field-test the system with hundreds of coaches and thousands of job seekers. At the end of Year 2, we will focus on scaling apps and maps to other states and partners for broader impacts. During Year 2, we will add monitoring and analytical capabilities to the app and the mapping system, with the *Employment Maps* ingesting risk scores and labor market data for detailed visualization, akin to the familiar tracking of hurricanes or pandemic spread.

PARTNERSHIPS

We will work closely with Skillful Indiana, a non-profit initiative of the Markle Foundation that is focused on accelerating the creation of a skills-based labor market in Indiana, and with InStride to do front-end evaluation, usability testing, and field testing of SAFER. We will collaborate with the American Library Association to disseminate information about SAFER to their membership of 57,000 libraries. In close collaboration with the Association of Science and Technology Centers, we will co-design outreach strategies and embed embed SAFER into the initiative on *AI and the Future of Work*. We embrace "track success" and hope to leverage synergies with the NASWA/NLx, Credential Engine, Competency Catalyst, Ripl.org, NCSU, and BHEF teams on creating a shared data infrastructure and channels to broadly disseminate our app to reach workers across the country.

INTELLECTUAL PROPERTY

BigKnowledge will apply its proprietary, patent-pending workflow and software to provide project partners with API access to visualizations and inference functionality and enable public access to the employment base map at least for the duration of the project. Any existing IP rights will be unaffected by the incorporation of respective components into project processes and products. Data IP rights are under discussion.

COLLABORATIVE AI PLATFORM FOR SMART AND AUTONOMOUS JOB SEARCH, RECRUITMENT, TRAINING, AND WORKFORCE DEVELOPMENT

Nihar Mahapatra
nrm@msu.edu

Christopher Nye
Jason Huang
Amanda Chuan

OVERVIEW

AI&Work360 is an extensible, collaborative AI platform and marketplace that continually learns from data to enable personalized, context-adaptive, and frictionless job search, transition, and pathways, internal and external recruitment, and training and workforce development. It is aimed at individuals, organizations, training providers, policymakers, and software developers. It emphasizes AI fairness and support for students, veterans, persons with disabilities, and minorities, and has built-in pandemic awareness.

DESCRIPTION

By 2030, up to 14% of the workforce is expected to change occupations as the world of work is disrupted by technological advances. In OECD countries, about 14% of jobs are at high risk of automation and 32% face the prospect of radical transformation. Certain population subgroups like students, veterans and other military-affiliated individuals transitioning to the civilian workforce, persons with disabilities, and minorities face special challenges. Consequently, they end up with poorer employment prospects, with lower likelihood of getting hired or getting a lower-paying job that does not fully utilize their skills.

To address the above challenges, we will build AI&Work360, an AI platform that streamlines the workflow for job search, transition, and pathways, internal and external recruitment, and training and workforce development. Although the platform is meant for all types of workers, it has specific AI services for students, veterans, persons with disabilities, and minorities. It will make the job search process much more effective, easier, faster, transparent, and in alignment with individual preferences and goals. Similarly, it will help organizations meet current and future challenges by making recruitment and workforce development more responsive to their needs, efficient, and fair.

Our platform has the potential to make a positive impact on the lives of tens of millions of Americans who may be either unemployed, underemployed, or preparing to change jobs or occupations. This will be enabled by our partners who will use our deliverables and assist with scale-up to have national-level impact. These include: 50,000 students enrolled at Michigan State University (MSU) and the nearly 14.5 million college students in the United States; over 18 million veterans, over 18 million people with a disability that are employed; the over 40 million Americans who lost their jobs due to COVID-19; millions who will join the workforce in the coming decade; millions of displaced workers; and the millions of workers at our partner organizations.

DIFFERENTIATORS

Our platform differs from other career assistance tools and recruitment services in a number of transformative ways: (1) For effective person-job matching, our platform has an occupational knowledge graph that maps the world of work, tracks how it changes geographically and over time, and interlinks worker characteristics to job and occupation characteristics across all occupations. (2) It streamlines the workflow for job search, transition, and pathway planning, internal and external recruitment, and training and workforce development, and exploits synergy between the different stages of each workflow and across workflows so that the combined effect is greater than the sum of their separate effects. (3) It allows for personalization to the characteristics of the individual user or organization and adaptation to the surrounding context (geography, industry, occupation, job, labor market conditions, etc.) to optimize decision making. (4) It provides collaborative, configurable autonomy so that

AI services within each stage or across stages of the workflows supported can be performed autonomously depending upon user preference or in collaboration with other users in a person's social network. (5) For workers, it will make the job search, transition, and career pathway planning process much more effective, easier, faster, transparent, and in alignment with individual preferences and goals. (6) It will help organizations meet current and future challenges by making recruitment and workforce development more responsive to their needs, efficient, and fair. (7) It provides a variety of predictive data analytics to minimize uncertainty in job search and recruitment. (8) It not only helps individuals find the right jobs, but also organizations find the right people, and thereby creates a win-win. (9) It is extensible because other AI services can be integrated into the platform through the marketplace to extend its functionality over time. (10) Finally, it has pandemic awareness built into its AI services for both individuals and organizations.

ROAD MAP

Y1-Q1: Complete occupational knowledge graph and back-end infrastructure.

Y1-Q3: Finish and obtain partner feedback on app front-end. Complete cross-cutting services and research tasks.

Y1-Q4: Initial prototype complete and enter testing and iterative updates. Initial partner rollout.

Y2-Q1: Enter alpha stage of application, product branding and marketing campaign started. Other AI&Work360 workflow services and research tasks completed, enter testing. Start social media marketing.

Y2-Q2: Enter beta stage of application. Complete mobile application and enter mobile tests.

Y2-Q3: Release candidate complete. Partner feedback and focus groups for mobile and workflow services completed.

Y2-Q4: Release final web-based prototype and mobile applications with revisions from partners. User interviews and marketing complete, with initial user base established. Preparation of SBIR Phase II proposal submission and commercialization with MSU Technologies Office.

PARTNERSHIPS

We have a broad and diverse range of prominent worker- and organization-centric partners in academia, government, and industry, both civilian and military, for general users, as well as those with specific needs. Our partners are providing us with insights, data, personnel time, pilot testing commitment, and help with national-level scale-up. Organizational partners include the Fortune 50 companies *Ford, Johnson & Johnson, Boeing,* and *Proctor & Gamble.* For students and future workers, we have *Career Services Network* and other units within MSU and University Innovation Alliance representing most areas of work. For veteran workers, we have partners locally at MSU, *Michigan Veterans Affairs Agency, Department of Veterans Affairs,* and *the Department of Defense,* which has the world's largest workforce. For persons with disabilities, we have a tie-up with MSU *Project SEARCH* and through them with its national-level organization and *Michigan Rehabilitation Services.* For both worker and organization perspectives, we have partnerships with *Center for O*NET Development* and the largest jobs site *Indeed.* Finally, we have labor market data providers *Emsi, Chmura,* and *Burning Glass Technologies.* We are also collaborating with many B1- and B2-track teams with complementary or supplementary functionality or data.

INTELLECTUAL PROPERTY

Preliminary versions of: occupational knowledge graph, transferability of competencies, and career pathway planning procedure were incorporated into a demonstrative software tool in Phase I. Copyright to the software is solely owned by Michigan State University and the IP is administered according to MSU Copyright policy and the terms and conditions of the Phase I contract. This Phase I software will serve as a foundational base for further research and to continue the development in Phase II.

THE NATIONAL LABOR EXCHANGE RESEARCH HUB

Emma Northcott
enorthcott@naswa.org

Charlie Terrell
Marty Romitti

OVERVIEW

The National Labor Exchange Research Hub (NLxResearchHub.org) establishes the first national, open ecosystem of real-time and historical labor market information to improve connections between jobseekers and employers. The NLx Research Hub will offer cost-efficient, timely, and transparent data to researchers and practitioners developing artificial intelligence (AI) tools to better predict the future of work. It will also provide self-service tools to help jobseekers answer questions such as: "What jobs are available? What skills do I need to get those jobs? Will the skills be transferable across industries? What support is available to ease my career transition?"

DESCRIPTION

Many jobseekers make career decisions based on out-of-date and incomplete information. Employers struggle to identify critical skills needed by their work force to stay competitive in a global economy. Automation and catastrophic events like the COVID-19 pandemic add to the uncertainty.

The National Labor Exchange (NLx), a partnership between the National Association of State Workforce Agencies (NASWA) and DirectEmployers Association, has provided a public-private job board since 2007. For 13 years, the NLx has compiled millions of job postings that represent a treasure trove of information for researchers predicting the "future of work," and for those helping jobseekers and employers navigate that future. NLx data can support dynamic information tools to help people make decisions in rapidly changing labor markets. This data source, which grows year over year, will fuel the NLx Research Hub and provide the best information currently available on the types of jobs in demand, and the skills needed for positions of the future.

The NLx Research Hub will give researchers unprecedented access to job posting data through an online data warehouse. Provided at no cost to state agencies, educators, economic developers, and local communities, the NLx Research Hub can provide data to help plan careers, develop training courses, expand businesses, and grow local economies.

DIFFERENTIATORS

The key differences between the NLx Research Hub and other sources of jobs data will be quality, cost-efficiency, transparency, timeliness, and trust.

The NLx is a trusted partner to over 300,000 employers and the workforce development agencies of all 50 states, DC, Guam, and Puerto Rico. This unique network will help the NLx Research Hub's data and tools reach a national audience and create impact long after the conclusion of Phase II.

For-profit job boards and data products often keep their sourcing and aggregation methods private. This makes it difficult to validate the data and connect it to other information sources. Moreover, it is inefficient for every locality or state to buy a separate data license. The NLx continues to offer jobseekers a no-cost source of unduplicated job content, posted by verified employers and updated every 24 hours. The NLx partners will now expand its impact by making job posting data more readily available to inform the future of work.

ROAD MAP

Key milestones in Phase II include building the technical architecture for the NLx Research Hub, creating a governance framework for researchers, developing research with end-users, and prototyping worker-centered tools. Major milestones are listed below (deliverables are

indicated with *), based on an anticipated start date of September 1, 2020.

11-20: Gather technical infrastructure requirements for the NLx Research Hub.

02-21: Achieve full Data Trust, governance, and data access policies and procedures.*

03-21: Complete back-end data infrastructure.

03-21: Launch state research and development pilot (KY, MI, MN, MT, OR, PR, VA).

06-21: Define state research use cases.

07-21: Finalize scripts to tabulate data for research.

08-21: Soft-launch the NLx Research Hub, making raw data and background resources available to state workforce agencies and approved researchers.*

09-21: Hold a public workshop on research applications of NLx data.

01-22: Complete prototyping for state pilot tools.

02-22: Start 5-month beta testing of prototype NLx Research Hub dashboards with jobseekers and frontline staff in American Job Centers.

05-22: Launch state pilot tools through CareerOneStop, state websites, and American Job Centers.*

06-22: Deliver a national training webinar on applied research techniques.

08-22: Complete research/tool development using AI to generate labor market insights.*

08-22: Publicly launch the NLx Research Hub.*

PARTNERSHIPS

NASWA represents all 50 state workforce agencies, DC, and U.S. territories. DirectEmployers is a leading HR association serving over 900 member companies. These two entities provide the content and technology behind the nation's largest nonprofit job board, usnlx.com, the job vacancy data source behind the NLx Research Hub.

The Labor Market Information Institute, a research center housed within the Center for Regional Economic Competitiveness (CREC), will help develop the NLx Research Hub user community and evaluate the impact of the Hub on that community.

The W.E. Upjohn Institute for Employment Research will

conduct labor market analysis and test the prototype tools with jobseekers.

SRI International will develop AI applications for self-service queries and future tool development.

BrightHive will support data governance activities.

Analysts from state workforce agencies in Kentucky, Michigan, Minnesota, Montana, Oregon, Puerto Rico, and Virginia will participate in a state pilot to develop research and deploy data tools through their state websites, CareerOneStop, and American Job Centers.

INTELLECTUAL PROPERTY

The team developed and signed a Data Trust Agreement to govern labor market research as an approved use of NLx data. As the trustee of the Data Trust, NASWA designated a Governance Board which will meet quarterly to monitor the sharing and publication of NLx data. The Board will consult legal counsel to address IP issues not already covered by the Data Trust Agreement.

WE-RISE: INTERNET OF EMPLOYMENT

Huiling Ding
hding@ncsu.edu

Xipeng Shen

Min Chi

Tim Menzies

Xiaolei Fang

WE-RISE
"Internet" for Employment

OVERVIEW

WE-RISE (Workforce Empowerment and Retraining through Internet of Solutions of Employment) is the first AI infrastructure that removes the interoperation barriers among workers, employers, trainers, and supporters in workforce development. Available as a plugin to individual users, it brings free, one-stop-shop solutions: credential gap analysis and recommendations of job, training, and wrap-around services to millions of job seekers as they search for jobs using their favorite recruiting websites.

DESCRIPTION

Open and extensible, WE-RISE tackles challenges in workforce development by connecting non-credentialed workers with middle-skill career technical education (CTE), particularly short-term, non-credit, and purpose-driven training programs, to prepare workers for in-demand jobs. It enables interoperability and connectivity among stakeholders in the workforce ecosystem by implementing standardized data formats, automatic data conversion tools, and dynamically linked and unified ontologies of jobs, skills, and training. Bridging job requirements and training, WE-RISE creates alignment between industry needs, individual skills, and CTE programs to enable better career guidance and a better match of worker skills to employer needs. After workers upload their resumes and/or skills, WE-RISE provides them with instant skill gap diagnosis and recommendations to help them transition to bright outlook jobs in manufacturing, a sector predicted to witness 2.4 million unfilled jobs in 2028. Moreover, our training modules on automated recruiting and soft skills help prepare workers for job interviews and workplace success in today's challenging labor market. Cultivating talent pipelines to fill in the skill gaps, WE-RISE will help alleviate unemployment caused by automation and the ongoing COVID-19 outbreak.

WE-RISE will be widely tested by critical partners in the ecosystem: manufacturers, job centers, job seekers, and community colleges in the four states of Florida, Massachusetts, North Carolina, and Texas. To ensure post-Phase II scalability and sustainability, we will explore two options with our partners: forming a non-profit consortium or startup company targeting workforce divisions, job centers, and training programs as possible paid clients or seeking grant from funding agencies such as Ford or the Mellon Foundation to scale up and deliver impactful solutions to American workers.

DIFFERENTIATORS

Disconnected from WB boards and college websites, pay-for-performance recruitment websites "harvest" resumes to hire talent for businesses. LinkedIn Premium charges a minimal monthly fee of $30 to offer white-collar subscribers job skill analysis. Aiming to transform workforce development, WE-RISE provides targeted service to non-credentialed workers who have been largely neglected and "Linked Out". *Unlike existing solutions, WE-RISE finds the job seekers in the places they are job seeking, provides detailed assessment on why they are not qualified for positions they apply to, and direct them to career-oriented training to bridge those gaps – free of charge.* Functioning as a personalized and portable "career studio," WE-RISE will assist job seekers with skill gap analysis before referring them to training programs for upskilling and to wrap-around service providers for needed assistance. WE-RISE can help job centers, training programs, and service providers drastically increase their reach to potential customers by matching these services with millions of job seekers who conveniently access the full-package solutions offered by WE-RISE from all major job search websites.

ROAD MAP

We will devote Year 1 to developing core modules of our minimal viable product (MVP): interoperable software infrastructure, credential gap analysis, and personalized recommendations of jobs, training, and services. In Summer 2021, we will provide partner and public APIs covering data format conversions, ontology access, and core capabilities. We will also offer a technical workshop and continued technical support via discussion forums to interested partners and recruit 100 pilot users through an educational conference focusing on soft skills and career pathways. Year 2 will be devoted to user testing and cross-validation of our workforce solutions through collaboration with multisector partners in workforce divisions, manufacturing, unemployment centers, recruiting services, and community colleges. We will also test the interoperability and scalability of our infrastructure in hospitality, a sector hard hit by COVID-19, to assist with workforce upskilling and possible transition to sustainable careers in manufacturing.

PARTNERSHIPS

Solidly integrated into the workforce ecosystems, WE-RISE has multi-state connections in public and private sectors. *Career centers (NC Works Commission, Florida, Massachusetts, and Texas)* will share datasets about worker and employer needs. *Community colleges (NC Community College System, i.e., Wake Tech, BlueRidge; Lake-Sumter, FL; Kilgore, TX)* will share datasets about credentials and training programs. *Manufacturing Extension Partnerships and NC Chamber* will connect us with small- and medium-sized manufacturers for domain expertise. Nonprofit partners and professional organizations will support us with research and teaching of soft skills and career pathways in high schools, colleges, and workplace settings. Finally, *recruiting websites such as Indeed, Jobs.com, and LayOffers.com* will share employment data and insights into automated hiring processes. All partners will test our core modules when assisting job seekers.

INTELLECTUAL PROPERTY

The project will be based on one patented technique by the PIs on reuse-centric optimizations of machine learning. We will generate at least three new patents, i.e., WE-RISE cyberinfrastructure design, ESA plus-based credential gap analysis, and personalized training recommendations. The partners will receive the priority to test drive WE-RISE and to license the IPs after the release of the MVP.

SKILL-XR: AN AFFORDABLE AND SCALABLE X-REALITY (XR) PLATFORM FOR SKILLS TRAINING AND ANALYTICS IN MANUFACTURING WORKFORCE EDUCATION

Karthik Ramani
ramani@purdue.edu

Alex Quinn   **Thomas Redick**
Kylie Peppler   **Niklas Elmqvist**

OVERVIEW

Rapid technological change has increased the need to re-skill the manufacturing workforce. Although virtual and augmented reality (VR-AR or XR) are promising for training, technology barriers prevent easy content authoring and sharing. We are developing a new approach, using Artificial Intelligence (AI) + XR, for the manufacturing industry and educators to create low-cost and scalable applications for (i) hands-on workforce training, and (ii) rapid skills-assessment.

DESCRIPTION

We propose a flexible way for employers and educators to easily author personalized AI+XR applications that simulate the 1:1 apprenticeship training model (tailored to each employee). Our platform enables the user to create do-it-yourself applications using 10% the time cost relative to rudimentary video or written instructions. This allows for training and analytics for workforce preparation at the scale of an entire enterprise.

Through in-depth discussions with over 90 potential users in manufacturing, education and hardware installation, we identified pain points (difficulty of use, lack of feedback and analytics, amongst others) that current approaches to XR don't overcome. Armed with these insights, we iteratively developed our initial focus to (a) a smart easy-to-use XR authoring platform for training and learning applications, and (b) a wearable camera with AI to monitor and provide real time personalized task analytics of users activities for skills assessment.

In Phase I, potential users informed our technology strategy, engaged with our initial prototypes, and provided feedback through iterations. In Phase II, we will test and deploy more robust "real" use-case prototypes through our industry and educational partners to obtain trainee analytics data. We will also conduct testing in our hands-on distance classes, play-labs in museums, and with small businesses. Our technology creates an entirely new worker training and task analytics paradigm coupled in a positive feedback loop. Hence, its potential market value is at minimum the current training-learning market ($370 billion). Because Skill-XR is a general purpose technology, it is expected to expand the training and analytics market to entirely new use cases that would otherwise have been too costly to consider. We plan to launch a start-up at the end of Phase II with our manufacturing and education partners as first customers.

DIFFERENTIATORS

Early prototypes and research revealed major barriers for users even to consider XR including lack of self-authoring capability, prohibitive cost, and lack of useful data analytics. Our foundational approach solves these problems by using simple building blocks that come together to make a modular platform so anyone can rapidly produce personalized XR applications. The features we support even in our prototype are not seen in prior art or products. As a result, our platform is far superior to existing alternatives like video or training manuals, because of spatial immersion, real-time feedback, monitoring of performance, and content-based network effects. Instead of costly repeated demonstrations or training sessions for every procedure, we can scale training content across time and space. Furthermore, our proprietary AI-based sensing technology provides task analytics which allow managers to track and optimize productivity across time and space.

Our multi-disciplinary team brings together decades of experience in our respective fields: spatial interactive interfaces, AI+XR, visual data analytics, and cognitive and learning sciences. We have a rich history in building teams

and startups that work across previously-siloed verticals in a convergent manner. And perhaps most critically, we've spent over 30 years collaborating with and understanding the manufacturing industries of the Midwest.

ROAD MAP

In year 1, our initial user-inspired technology pilots will focus on high demand use cases from our partners, moving from low-medium fidelity to high fidelity prototypes and improving at each iteration. In year 2, we will perform a full evaluation study of the productivity gains from our platform using task-specific analytics on motion and time taken for several tasks and parameters including handling of objects, cognitive load, knowledge of tools, operations, procedures, and workflow and planning effectiveness. By leveraging SkillXR's shared knowledge of objects, environments, interactions, and self-improving AR, our product-service model will be poised to touch hundreds in year 2, tens of thousands in year 3-4, and millions of workers and children by year 4-6. *Imagine creating and sharing skills on the cloud with AR-Tube while transforming real environments to smart ones.*

PARTNERSHIPS

We've engaged partners across sectors to deeply understand their skills requirements, workflows, and difficulties in training. The lead partners, with clear use cases ready for Phase II, include manufacturing (Nvidia, PTC, Gaylor Electric, Wabash National, Kirby Risk, Delta Welding, and other undisclosed partners), education (Britannica Global, Ivy Tech Community College, Purdue Gifted Education Institute, Strong Museum, ZeroUI, E-Blox), government/non-profit (Indiana Economic Development Corp, and rural Carroll County in Indiana) and VC-based product-market fit advisors.

Our partners span industries and have differences in machines, type of knowledge, and skill levels. Still, they have similar patterns in tasks and unit operations. By understanding their workflows, we engineered our platform to facilitate cognitive, perceptual, and motor skill features to be universal across contexts. We've already shared early prototypes with partners, and received feedback that one primary piece of value is that we've made XR, previously unapproachable, into something they can consider using at scale to solve skilling pain points. In Phase II, we plan to formulate a startup and licensing strategies in order to commercially engage with our partners.

INTELLECTUAL PROPERTY

To date, our team has been awarded 5 technology patents, with 10 applications pending, and 10-20 more planned for Phase II. Industry partners have signed a mutual NDA. Finally, we have formed a patent strategy development team with the Shor Chan group and Maginot group alongside the Purdue Research Foundation.

DATA-DRIVEN RESKILLING MARKETS PREPARE AMERICANS FOR THE FUTURE OF WORK

Justine Hastings
justine_hastings@brown.edu

Susan Athey   **Jonah Rockoff**
Eric Chyn   **Seth Zimmerman**

OVERVIEW

We use machine learning and Artificial Intelligence to transform government administrative data into scientifically-valid measures of Return-on-Investment (ROI) to labor and reskilling programs. We deliver these metrics straight to workers and job counselors through a user-centric web- and mobile- app to empower informed decisions. Informed decisions prompt training programs to innovate and add value, creating a virtuous cycle of data-driven improvement and positive impact. The result is improved employment and economic opportunity for all workers per dollar spent on training.

DESCRIPTION

Even before the COVID-19 pandemic, 39 million American workers were predicted to lose their jobs to automation. In the last 4 weeks alone, 30 million Americans have filed for unemployment. Now, more than ever, we need an effective market for jobs training and reskilling.

However, as everyday Americans struggle to reskill and prepare themselves for new jobs, they must rely on word-of-mouth information to help them choose reskilling programs. The data needed to guide them towards effective programs are locked up in decades-old government systems. Meanwhile, state leaders face the current economic crisis with minimal data insights to guide their policy investments. Billions of federal and state dollars and scarce time and money from struggling families are invested in reskilling programs with no evidence of effectiveness. It's no wonder we face a persistent skills gap with this inefficient and poorly-incentivized system.

Our team partners with state and local government to use data, science, and technology to create the tools and metrics needed to foster an effective, data-driven reskilling market to benefit all Americans.

Our project is called Data for Opportunity in Occupation Reskilling Solution (DOORS). We are working with states across the country to unlock the power of these data, measure which reskilling programs work, and put that information in the hands of the people.

To create ROI metrics with administrative data, we designed a secure, cloud-based *Research Data Lake* (RDL) environment which houses administrative data and optimizes them for rapid insights. Metrics can be quickly developed, put into practice by policy makers, and delivered directly to the public through web and mobile apps. Our approach is sustainable, increasing government's capacity to implement data-driven policy, partner with researchers to measure what works, and continually iterate and improve.

Workers and job counselors will use our metrics to make informed decisions. Policy leaders will rely on our metrics to fund effective program expansion. Training programs will use them to innovate and measure success. DOORS will catalyze the reskilling market, opening new doors to employment and economic opportunity for all Americans.

DIFFERENTIATORS

"I want to see if [a program] is worth it. I cannot spend valuable time on something that will not benefit me in the end." — *Training program seeker.*

"This will make us more credible. People will like us better if we have data [and] can answer their questions with confidence." — *State job training counselor.*

DOORS delivers an end-to-end solution to actually deliver scientifically-valid ROI measures into the hands of American workers who need them and will use them going forward. By forging and incorporating key partnerships throughout our process, DOORS produces science that will actually

be used by the American people to make their lives better. DOORS partners include tech and finance industry leaders and four state partners who serve a combined population of 23 million Americans who are committed to using DOORS to deliver a reskilling ecosystem that effectively and efficiently up-skills their workforce for the future of work.

Scott Jensen, the Rhode Island Director of Labor and Training, said: *"This partnership is helping us fully leverage our data to make better use of Federal funds, help workers re-enter the workforce, and measure the impact of workforce development programs to prepare all workers for in-demand jobs."* Rhode Island has already been able to produce ROI measures for its leading jobs program, and in partnership with our team leveraged DOORS to spin up an emergency COVID-19 benefits system in just 10 days using the RDL to deliver Pandemic Unemployment Assistance to 30,000+ Rhode Islanders. It was one of the first in the nation to deliver these benefits to people in crisis.

DOORS is open-source and scalable to any U.S. state or local government. Ten states have expressed interest in implementing DOORS to date. DOORS is efficient and self-sustaining. Government learns the value of technology and science as we invest in their capacity to own and manage DOORS going forward, increasing evidence-based policy capacity and impact long-term.

ROAD MAP

With Phase II funding, we will:

• Launch a fully-functional RDL for partners in four states and ROI measures by Q1, 2021.

• Measure the impact on demand for and effectiveness of training programs by Q1, 2022.

• Launch dashboards and evaluation tools for government partners to measure continual improvement by Q1, 2022.

• Launch the full, productionized open-source system for any state or local government to adopt by July 1, 2022.

PARTNERSHIPS

Public partners:
Colorado Dept. of Higher Education; Rhode Island Office of the Governor and Dept. of Labor and Training;

Virginia Offices of the Governor and Secretary of Administration; Washington Employment Security Dept.

Non-profit partners:
• *Research Improving People's Lives (lead institution):* In Phase I, developed ROI measures; built technology and data resources. In Phase II, scaling across states.

• *The Centers for Civic Impact at Johns Hopkins University:* In Phase II, training government to use the ROI measures.

• *Ideas42:* In Phase II, designing training for the web- and mobile-tool.

• *The National Bureau of Economic Research:* In Phase II, bringing together academics, policy makers, and industry to share learnings.

Private and technology partners:
• *Amazon Web Services (AWS):* Donating technical expertise and cloud environment.

• *Facebook Research:* Donating technical expertise to integrate ROI measures into social media platforms.

• *Maycomb Capital:* Donating performance to quickly expand effective reskilling programs.

INTELLECTUAL PROPERTY

We are an open-source platform that may incorporate some proprietary software for cloud computing, which is a traditional business relationship bound by terms and conditions. There are multiple competitors in the market for cloud computing. Our system can run on any of them. The RDL is open-source and can be run for free. The code to produce the ROI measures and app will be freely available, and the measures will be publicly available through an API.

DIRECT: A FRAMEWORK FOR DIAGNOSIS, RECOMMENDATION, AND TRAINING IN CONTINUOUS WORKFORCE DEVELOPMENT

Beverly Woolf
bev@umass.edu

Andrew Lan   **Ina Ganguli**
Shlomo Zilberstein **Andrew Cohen**

OVERVIEW

Manufacturing constitutes a significant segment of the US economy, given 244,000 small plants and 52,000 large factories distributed across the US. However, with the increasing adoption of new technologies, such as robots, a significant disparity exists between the skills of current US workers and the requirements of future jobs, resulting in an estimated 2.4 million unfilled jobs in the next decade. DIRECT will help both unemployed and incumbent workers find jobs, plan future career paths, select training activities for upskilling and reskilling, and obtain on-the-job training. It will help the nation's global competitiveness, have a huge impact on increasing the scale of manufacturing and contribute to a diverse US workforce. DIRECT is especially relevant in the aftermath of the COVID-19 pandemic, as the US will experience an unprecedented wave of job seeking.

DESCRIPTION

Using artificial intelligence-enabled components, the *Diagnosis, RECommendation, and Training in continuous workforce development* (DIRECT) framework establishes a vertically comprehensive upskilling and reskilling manufacturing pipeline. It (i) connects workers to future jobs and helps them plan future career paths using AI algorithms trained on large-scale worker, job, and training resources, (ii) provides efficient on-the-job training (OJT) through an intelligent AI-enabled platform. It prioritizes fair, equitable, transparent, and trustworthy processes. DIRECT is also horizontally generalizable to a wide range of worker populations and industries.

DIRECT has three major components:

- The **job matching** component matches workers to available jobs they are qualified for by using AI algorithms to compare mastered skills provided on worker resumes to required skills listed on job postings.

- The **career path planning** component generates a diverse set of long-term future career paths for workers given their backgrounds, preferences, and real-life constraints and recommends training resources to fill skill gaps by using worker career trajectory data to quantify the effectiveness of training resources.

- The **continuous training** component provides personalized just-in-time OJT to manufacturing workers through an intelligent electronic platform, resulting in efficient upskilling, increased productivity, and improved worker experience.

DIFFERENTIATORS

Our project team's diverse combination of expertise (computer science, economics, psychology, and sociology) provides an exceptional opportunity to manage various aspects of job search and training and to develop an effective and comprehensive solution to upskilling and reskilling.

Our collection of industry partners, including both large, international manufacturing corporations with factories across the US and small, local manufacturing plants in Massachusetts, Connecticut, Michigan, Illinois, and North Carolina, enables us to develop and test solutions that will scale to a wide range of US workers and manufacturing businesses.

Our collection of public partners, including several state workforce agencies, enables us to directly benefit workers in real need, especially low-income and low-skilled job seekers. This focus is in stark contrast to existing tools that are primarily developed using data on predominantly high-skilled workers with detailed online resumes.

DIRECT will promote equity across workers with different gender, race, and socioeconomic background. We will conduct interviews and surveys to elicit the perspectives and real-life constraints of underserved groups. To deliver fair and transparent algorithmic recommendations, these findings will be incorporated into the development process of the AI algorithms and worker interface.

DIRECT will especially be relevant in the current COVID-19 pandemic by helping unemployed workers find jobs quickly and reducing the need for face-to-face interactions in career counseling and OJT settings.

ROAD MAP

Deliverables	2020-2021	2021-2022
Build **datasets** on workers, job seekers, jobs, and training resources with our partners	✓	
Develop **DIRECT's AI-enabled components** for job matching, career path planning, continuous training, and worker interfaces	✓	✓
Evaluate DIRECT's usability, effectiveness, and worker experience through **quantitative and qualitative studies**		✓
Test DIRECT's generalizability and create an **open-access version** for the public		✓

PARTNERSHIPS

T = Training resources and technology

D = Data for training DIRECT algorithms

P = Physical plant for testing DIRECT

Partner	Contribution
Stanley Black & Decker (T, D, P) Manufacturing company well-known for making industrial tools with 60,000 employees and 56 plants in 25 states	Data on workers and jobs, training materials, testing site for continuous training with real workers
MassHire (T, D) Workforce agency with 29 Career Centers in Massachusetts	Data on job seekers, jobs, and training resources, worker interviews and surveys, testing site for job matching and career path planning

Partner	Contribution
Hope Street Group (T, D, P) Workforce agency in Michigan with 2,500 manufacturing plants	Data on workers, jobs, and training resources, worker interviews and surveys, testing site for all components
PBC Linear (T, P) Manufacturing company that makes bearings in Illinois with 200 employees	Training resources, worker interviews and surveys, testing site for continuous training
Arizona State University (T) AI and Robotics	Intelligent training with collaborative robots
National Association of State Workforce Agencies (D) National job posting repository with 50 state workforce agencies	Data on jobs
Competency Catalyst (D) Platform for worker credentials and training resources	Data on workers, jobs, and training resources
North Carolina State University (T, D) Manufacturing plants and community colleges in North Carolina	Data on workers, jobs, and training resources
CMU Robotics Academy (T) Training materials from youth to adult robotics training programs	Robotics training materials, robotics bootcamp

INTELLECTUAL PROPERTY

The developers will support free use of the software produced and will reveal code in the expectation that other developers will reciprocate. We will share code with others both within and outside the workplace and engineering communities. We intend for these open source tools to increase worker learning, engagement, and productivity and improve the fairness and transparency of the platforms. The DIRECT components will be expanded to a variety of manufacturers (e.g., tools, ball bearing) and for different segments of the workforce (e.g., youth, unemployed, disadvantaged).

SHAPING FUTURE CAREERS FOR AMERICA'S MOST VULNERABLE WORKERS: A TECHNOLOGY-ENABLED MATCHING AND TRAINING PLATFORM

Arthur Y. Huang
arthur.huang@ucf.edu

Erik Brynjolfsson   **Vince Albanese**
Zhishan Guo   **Nan Hua**

OVERVIEW

The economic restructuring necessitated by the COVID-19 pandemic is a harbinger of a more fundamental transformation of our economy and workforce. No sector has been harder hit than the HOT industries (Hospitality, Operations, and Transportation) and no sector has more to gain from reinvention. This project focuses on the key leverage points for improving the outcomes for workers in an increasingly dynamic and digital labor market: connecting workers' potential to rising career development opportunities, recommending viable career paths, and developing an efficient and secure system for job matching and skill training.

DESCRIPTION

This project integrates the newest advancements in deep learning, natural language processing, linguistics, cybersecurity, and the distributed database system to build a technology-enabled career facilitation platform which can enhance workers' and businesses' adaptive capabilities to prepare for the future of work. This project includes the following key components:

• Development of a new CAPABILITY (Credentials, Aptitude, Personality, Adaptability, Background, Ideals, Leadership, Initiative, Temperament, and Years-ofexperience) attributes taxonomy, which expands the O*NET and EMSI skills databases.

• Development of dynamic knowledge graphs visualizing the evolution of CAPABILITY attributes for different jobs and work reorganization trends.

• Development of a career transition models which can customize career paths based on workers' CAPABILITY attributes, interests, constraints, desires, perceptions of risk, expected salaries, and rising career opportunities.

• Development of soft skill training modules.

• Development of a trustworthy career facilitation platform for matching workers to jobs in an efficient, secure, and trustworthy fashion, as well as providing customized career path recommendations.

DIFFERENTIATORS

We are the only team in this track that focuses on the labor-intensive service industries. This project provides innovative solutions to support vulnerable workers in the hospitality, operations, and transportation (HOT) industries when times become exponentially more distressing. The HOT industries contribute about 24% of the total US GDP and employ about 32% of the US labor force. The myriad entry-level jobs in these industries are often the first stepping stones to success in many other parts of the economy; however, the economic significance has not been sufficiently reflected in the support for their core workforce. The average hourly wages, benefits, job retention, and career development support are among the lowest of all industries. The research team's prior work revealed that jobs in the HOT industries have the highest Suitability for Machine Learning (SML), an indicator for the potential reorganization of a job as duties change as a result of technological penetration. The HOT industries are due for major restructuring as machine learning becomes widely adopted.

The research team will harness deep learning models and numeric models to derive unique insights from massive job postings and resume databases. Focus groups and interviews will provide unique angles about the trends of work reorganization and the critical CAPABILITY attributes for job success. The federated trust framework has proven its merit by providing secure and reliable data storage and exchange solutions. It is through the agency of such

sophisticated technologies that this project aims to prepare America's vulnerable workers to become more adaptable and resilient. This novel career facilitation system will include 6 key functions: (1) Revealing job trends and skill trends, (2) visualizing a job-level CAPABILITY attribute taxonomy, (3) recommending career paths, (4) providing a trustworthy mechanism for verifying workers' CAPABILITY attributes, (5) recommending skill training for rising opportunities, and (6) matching workers with potential jobs. This platform includes an innovative credential verification mechanism to enhance the efficiency and effectiveness of job matching. The Green Armanda Foundation has committed to sponsoring the operations of this platform following Phase II.

ROAD MAP

In Year 1 we will develop the CAPABILITY attributes taxonomy, identify job trends, and build the distributed database infrastructure. In Year 2 we will develop a job matching and career transition model and soft skill training modules, and perform platform testing and validation.

Year 1	• Development of the CAPABILITY attributes taxonomy • Visualization of the dynamic graph of the CAPABILITY attributes by job • Work reorganization trends • Setting up the platform infrastructure
Year 2	• Career transition model • Soft skill training modules • Credential verification mechanism • Job matching and training mechanism • User testing and validation

PARTNERSHIPS

The project includes a variety of organizations who have committed to full engagement during Phase II in forms such as sharing data and resources, participating in interviews and focus groups, recommending industry experts, and getting involved in platform testing and trial use. Our partners are listed as follows:

Career Center: CareerSource Central Florida

Industry Assoc.: Florida Trucking Assoc.; USIndependent Workers' Assoc.; Central FL Hotel and Lodging Assoc.

Gov: FL Dept of Economic Opportunities.; City of Orlando; Florida House of Representatives Workforce Development Committee

Worker Assoc.: Assoc. of Independent Workers

NPO: The Green Armada Foundation

Businesses: Amazon.com; iWorker, Inc.; Pack-IT, Inc.; NDM Marketing, Inc; Case Mason Filling, Inc.; Dorabot Robotics, Inc.; Coakley Williams Hotel Management Inc.; Red Lobster; RCI Vacation Network; iWorker; Kelly Services

Database companies: BurningGlass, Inc.; Linkedin.com; National Labor Exchange (NLx)

INTELLECTUAL PROPERTY

The research team will create several patents, including the new CAPABILITY attributes taxonomy, a new career transition model, and the job matching and training recommendation platform. Some of the materials developed by the team such as the soft skill categorization database, soft skill training content, and the career path recommendation system will first be used by our industry partners for internal training for some period of time, and then be made free to the general public.

INCLUSION AI FOR NEURODIVERSE EMPLOYMENT

Nilanjan Sarkar
nilanjan.sarkar@vanderbilt.edu

Susanne Bruyere   **Brian Scassellati**
James Rehg   **Zachary Warren**

OVERVIEW

Every year in the United States 50,000 autistic children become 40,000 unemployed autistic adults. Challenges in social communication have been one key factor preventing these individuals from finding and keeping jobs. Our AI-driven tools mitigate these challenges, empowering individuals with autism to gain meaningful employment, and enabling employers to access the capabilities of a more neurodiverse 21st century workforce.

DESCRIPTION

Our suite of tools includes: an AI assessment system that can uniquely profile the specific problem-solving talents of autistic individuals and match them to a job; a socially-intelligent, virtual-reality, job-interview simulator that reduces the social challenges of finding a job; a collaborative virtual-reality system that dynamically measures and teaches the social components of peer-interaction and team-based work; a social robotic system to assess and develop resilience to interruption on the job; and computer vision tools to assess and improve non-verbal communication.

Commercial deployment of these tools has the potential to transform what is currently an enormous societal cost into great value for the nation. The full system cost to support a single unemployed adult with autism is about $150,000 per year; the total support cost for unemployed autistic adults in the U.S. now reaches more than $175 billion annually. By impacting the meaningful employment of these individuals, we not only reduce these massive costs, *we actually position this large segment of the U.S. population to contribute substantially to the economy* through increased labor productivity, contribution of wage taxes, participation in the consumer and housing markets, and so on.

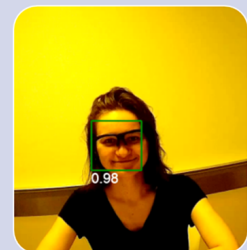| AI assessment system that can uniquely profile the specific problem-solving talents of autistic individuals and match them to a job	A socially-intelligent VR job-interview training simulator that reduces social challenges of finding a job	A collaborative VR system that dynamically measures and teaches the social components of adaptive peer-interaction and team-based work	A social robotic system to assess and train resilience to interruption	Computer vision tools to measure and assess non-verbal communication

DIFFERENTIATORS

Currently, neither individuals with autism, nor their potential employers, nor the vocational support systems dedicated toward matching people to jobs, have access to efficient, targeted, intelligent technologies that promote job matching/recruitment, ongoing assessment/support, or retention in an inclusive workplace. Not only are employers missing out on this untapped human resource; chronically unemployed people experience negative lifespan outcomes, poor mental health, and limited independence, with cascading societal repercussions and costs across decades.

In 2015, $268 billion was spent on autism services (for children and adults)—increasing to $450 billion by 2024—of which *not even 1% is currently spent on development of technologies* that can be rapidly brought to market at scale. A number of companies have boldly developed neurodiversity hiring initiatives, but there remains a gap in the development of technologies to support their hiring efforts and the individuals whom they hire.

ROAD MAP

Year 1 will focus on prototype refinement, evaluation of efficacy and market feasibility, and data capture to understand barriers to uptake. Year 2 will focus on deployment to real-world settings, commercialization with our private- and public-sector partners, and expansion toward even broader impact. *Our current work is focused on autism but has clear applicability to much larger underutilized segments of the workforce:* about 1 in 6 people has a "neurodivergent" condition (autism, ADHD, dyslexia), about 50 million people in the U.S. alone. We also envision our tools supporting the training, employment, and upskilling of workers recovering from a physical or cognitive trauma (such as a stroke or head injury) or those forced to train/work via virtual tools during a pandemic.

PARTNERSHIPS

We have a strong team of partners in this work:

1. Private-sector companies that hire for neurodiversity and seek cutting-edge tools for assessing and onboarding candidates, or supporting and upskilling their employees on the job: Auticon, Ernst & Young, SAP, The Precisionists, Inc.

2. Public-sector disability employment agencies that run regional training centers to support autistic adults with training and coaching to secure a job. Initial partners are representative of various regions across the U.S., including Tennessee, Iowa, Washington, and Southern California.

3. Companies committing tools and distribution platforms: Floreo, Tipping Point Media, and Microsoft have already deployed platforms for educational content for autism to share with us.

INTELLECTUAL PROPERTY

Our private-sector partners have committed to execute licensing agreements for use of our tools as they become ready for commercialization.

Our public-sector partners have committed to execute nonprofit agreements to deploy our technologies as scalable assessment and training opportunities for the individuals they support.

Our partners who have platforms for digital content have committed to work with us to embed our products within their existing platforms, potentially extending our reach to millions of current users.

TRACK B2

TEAMS

NSF | Convergence Accelerator

The ultimate goal of the National Talent Ecosystem track is to support research leading to innovative approaches for employers to support workers seeking the skills required for 21st century work related to AI, data science, predictive analytics, and other technologies of the future. Successful projects will prototype innovative approaches, such as learning environments, simulations and tools for analysis or assessment, and vehicles for recruitment and engagement, with the potential for wider implementation by industry, educational institutions, and other stakeholders engaging in the co-creation of a national talent ecosystem.

SAFE SKILL-ALIGNED ON-THE-JOB TRAINING WITH AUTONOMOUS SYSTEMS

Siddharth Srivastava
siddharths@asu.edu

Erin Chiou
Georgios Fainekos
Kurt Van Lehn

OVERVIEW

Manufacturers are moving towards programmable robots, for safety, efficiency, agility, and responsiveness to design changes—a shift viewed as essential for global competitiveness. *Our objective is to create and evaluate a proof-of-concept for a just-in-time Artificial Intelligence (AI) system that allows manufacturing employees without college degrees in CS/Math to easily program robots in an industrial setting.* This follows a new, human-centric vision for AI: *an AI system/robot must make it easy for its user to learn the entire range of its functionality.* Our system would automatically adjust the level of training to the skill-level of the worker and the task at hand.

DESCRIPTION

The manufacturing sector, which currently employs 13M workers, is moving towards AI powered programmable robots. However, day-to-day programming-related tasks such as debugging, or even understanding the behavior of today's AI/robot system---let alone adapting them to align with evolving needs---require expertise in computer science (CS)/AI. Less than 5% of the US workforce has college degrees in these fields. This raises the question: *who would program the programmable robots?*

Our vision is to create an AI system that can automatically estimate the skill level of a worker and adapt the training to make it more scalable, accessible, and demand driven. We plan to create an AI-based training system so lower skilled workers can program robots by giving high-level tasks such as "pickup gear A". Skill-aligned feedback from our system will make it easier for a worker to learn to debug robot behavior. This approach reduces the time burden placed on manufacturers as well as workers to upskill. This flexibility can reduce the costs of training as well as those of designing training modules.

We envision that this system will create a new job category, *robot tasker,* with applications in manufacturing and beyond. It spans the range between a robot operator (entry-level) and that of a robot engineer (specialized, high level). Our project proposes a solution that would assist millions of workers who have plateaued in their skills, not because they lack an ability or access to free online curricula, but because of constraints that make a time-investment in education a luxury.

In future spin-off projects we plan to use our insights in working with the manufacturing sector to make our AI solution more broadly applicable to other sectors such as healthcare, warehouse management and logistics.

DIFFERENTIATORS

Our innovative solution is made possible by our team's recent research results in AI, robotics, intelligent tutoring systems (ITS), and human-systems engineering. Prof. Srivastava's research makes taskable and explainable robotics feasible while Prof. Fainekos' research on safe control synthesis makes it possible to automatically create safety compliant robot behavior. Prof. VanLehn is a renowned expert in ITS and he has developed numerous widely deployed ITS for STEM learning. Prof. Chiou is at the forefront of research on new methods for long-term and short-term analyses of collaborative human-AI systems. Our team includes Prof. Subbarao Kambhampati, an expert on human-aware AI and the past president of AAAI, as well as Prof. Nancy Cooke, an expert on team effectiveness and the past president of the Human Factors and Ergonomics society. This unique research momentum enables our team to create a new paradigm of adaptive training systems that autonomously adapt to the worker and the robot task.

In contrast to approaches for creating curricula that require users to take time away from employment, our solution allows workers to learn on the job. In contrast to curricula that need to be created for every new combination of robot-task and worker-background, our training system would automatically adapt to the robot task and the worker, thereby reducing the cost of training even further. In contrast to approaches that require workers to learn how to program robots by providing precise robot-movement instructions, our approach will allow workers to safely task the robot using high-level commands. The proposed system will autonomously convert these high-level commands into safe robot programs when that is feasible, and it will give the worker skill-aligned explanations for infeasibility when it isn't.

ROAD MAP

Key milestones and deliverables for Phase II include (1) a proof-of-concept adaptive training system including (a) a system for tasking robots that allows low-skilled workers to easily adapt robot behavior to their needs; (b) a system for providing worker-skill aligned feedback based on self-explaining AI; and (c) a language for expressing safety constraints and regulations; (2) broad system wide evaluation of our proof-of-concept conducted with a user-group of manufacturing workers and stakeholders showing measurable progress using rigorous evaluation metrics; 3) a roadmap for future applications and deployment.

PARTNERSHIPS

We are collaborating with **Raytheon Technologies** *(RTX),* a major US manufacturing organization (resulting from a merger between United Technologies and Raytheon). RTX will co-develop and co-evaluate our solution from the point of view and practicality of deployment in the manufacturing industry. **University of Massachusetts** Amherst will address access to skilled talent pathways, aiding us in developing a system that matches workers to appropriate training.

The **Arizona Department of Education's Career and Technical Education division** *(AZCTE)* will bring expertise to help with the design and evaluation of training modules that serve disadvantaged members of the workforce.

Toyota Motor North America's Future Systems Research Department will support this initiative with a parallel project focusing on applications of our technology in other domains. **Modern Industries** and **Trinity Robotics**

Automation will help obtain feedback from a broader manufacturing perspective and arrange site visits to help the team develop broadly applicable solutions.

INTELLECTUAL PROPERTY

Code will be co-developed and shared by project partners. The project team plans to use non-exclusive rights IP agreements for work products. Team has a technology transfer office (ASU's *Skysong Innovations*) to assist with IP agreements.

CREATING A MARKETPLACE FOR UPSKILLING THE NATION'S WORKFORCE

Jennifer Thornton
Jennifer.Thornton@bhef.com

Brian Fitzgerald
Sara Hooshangi

Creating Solutions. Inspiring Action.®

OVERVIEW

McKinsey estimated in 2017 that the exponential growth in automation and the digital economy would require upskilling 375 million workers globally. This societal challenge has intensified with record-breaking job loss due to COVID-19: more than 33 million Americans have filed for unemployment as of May 2020. Yet even with increased need, effective digital technology (DT) upskilling for the workforce remains a challenge for employers to accomplish at scale. The Business-Higher Education Forum (BHEF) and its partners will support employers by creating an Upskilling Solutions Marketplace (Marketplace), connecting employers to skills-analysis tools, micro-credentials, and applications for worker engagement.

DESCRIPTION

BHEF will create an employer-focused Marketplace, where data providers share information on regional-skill demand and job trends, application (app) developers provide tools to capture employee skills and identify opportunities for upskilling, higher education institutions supply micro-credentials, and learners document skills, develop career plans, and upskill. The Marketplace will address employers' key barriers to upskilling; enable an increase in rapid upskilling connections between individuals, employers, and educational institutions; and provide additional solutions for the post-COVID-19 employment landscape.

The Marketplace will allow employers to overcome upskilling barriers by increasing continuous learning opportunities, improving access to skilled-talent pathways, and aligning the proficiency of essential DT skills for tens of thousands of American workers. Universities will develop upskilling opportunities with improved skills-demand data, and workers will better understand career pathways and the skills and training required of them.

BHEF and the Greater Washington Partnership (GWP) will work with four employers in the Washington D.C, Maryland, and Virginia region (DMV), Northrop Grumman, EY, General Dynamics IT, and Stanley Black and Decker, to perform usability tests on the interoperable, blockchain-based platform and provide feedback on the labor-market data, tools, and micro-credentials. BHEF will then scale the model to other regions. While BHEF and partners will lead the identification of an initial suite of tools and micro-credentials, the Marketplace will be open for additional higher education and app developers to offer products, tools, and training modules tools that meet employer needs for DT upskilling.

DIFFERENTIATORS

1. **Compelling workforce challenge:** The Marketplace will remedy frictions employers have when upskilling their workforce. It will serve as the first national platform for employer-facing labor market and skill-level workforce insights aligned with university micro-credentials.

2. **Synergistic responsiveness:** The Marketplace provides a framework for different, aligned initiatives to work together. Designed for agility, the Marketplace can support workers unemployed due to COVID-19 by translating labor market and skill-level needs to individuals and facilitating connections between companies and job seekers.

3. **Audience understanding:** This project has deep employer involvement through its Advisory Board, pilot partners, project team, and in-kind collaborators. Work is led by BHEF, a membership organization with a forty-year track record of bridging the gap between higher education and employers.

4. **Convergent team, technologies, and tools:** The diverse team and underlying technology platform for

the Marketplace will make use of existing and new technologies supporting credential management, data analytics, and competency frameworks.

5. **National Distribution:** Business, higher education, and intermediary organizations will deploy a comprehensive scaling strategy for national reach after pilot testing in the DMV.

ROAD MAP

Data and technology strategy:
- Strategy covering tools, technology, processes, rules, and requirements developed - *Jan 2021*

Marketplace platform ready for:
- Limited Phase-1 testing with fictional user (employee/learner) data - *May 2021*

- Phase-2 testing with fictional user data - *Dec 2021*

- Phase-2 testing with real user data - *May 2022*

Data and tools:
- Initial tool for jobseekers available (COVID-19 response) - *Nov 2020*

- Initial employer facing tool available - *Dec 2020*

- Competency mapping tool available - *Mar 2021*

- Additional upskilling tools available - *Dec 2021*

Pilot conducted in the DMV:
- Four higher education institutions develop and add DT micro-credentials to Marketplace - *Jul 2021*

- Four employers test Marketplace platform, data, tools, and micro-credentials with a subset of employees - *Nov 2021*

Scale:
- Additional DT micro-credentials from within and outside DMV on Marketplace - *July 2022*

- C-Accel teams' products integrated in Marketplace - *Aug 2022*

- Business stakeholders from three additional regions begin using Marketplace - *Sept 2022*

PARTNERSHIPS

During Phase I, BHEF and GWP led a consortium of universities and businesses in qualitative and quantitative research activities, which informed market intelligence on the DMV region's DT talent landscape. As part of this research, Virginia Tech served as an adviser on BHEF's higher education working group, and Emsi conducted labor market analysis.

In Phase II, BHEF and GWP will continue with a subset of Phase I business partners testing the platform and tools, four university partners (Virginia Tech, University System of Maryland, George Mason University, and Georgetown University) creating upskilling modules, and Emsi providing labor market data and employer-facing tools. IBM will provide Marketplace infrastructure, augmenting their blockchain credential platform and interoperable learner record system. Other businesses such as PwC, Accenture, Wiley, Ellucian, and Stanley Black & Decker have committed tools, in-kind support, and/or to serve on the Advisory Board. The convergent core team and advisers bring expertise in disciplines such as economics, workforce development, business, organizational development, data science, higher education, psychology, AI, and computer science.

INTELLECTUAL PROPERTY

The Marketplace will include user rules and licenses for the platform and shared content. Each partner, solution, and service provider retains ownership of any background and intellectual property created independent of this project and any solely created inventions and copyrightable materials under this project. Jointly created materials will be jointly owned. Each ownership party provides BHEF a nonexclusive, royalty free, revocable license to utilize the materials solely for work under this project.

FACTORY FORWARD

Ross Higashi
rhigashi@cmu.edu

Megan Andrew   **Petra Mitchell**
Sonya Francis   **Martin Scaglione**

OVERVIEW

Factory Forward is an integrated research and development project that tests novel methods for delivering training that unblock two key bottlenecks in the advanced manufacturing workforce pipeline.

BootUP combines career readiness training with a bootcamp-style intervention to fill critical entry-level worker shortages in the manufacturing sector.

SkillUP delivers one-hour, hands-on, games-based training sessions to help incumbent workers spot tech-addressable inefficiencies at work, driving sensible technology adoption.

DESCRIPTION

American manufacturers know that they must adopt advanced manufacturing technologies to compete with lower labor costs overseas, but have done so very slowly. An even larger disruption, driven by technologies such as machine learning and artificial intelligence is nearing. Yet broad-based workforce preparation in manufacturing is virtually non-existent today. Most factory floor workers receive no professional development, being trained narrowly on the operation of new machinery following its purchase.

SkillUP views this as a missed opportunity. SkillUP is a low-cost, lightweight professional development prototype for incumbent factory workers. A mobile workshop delivers a series of efficient, hands-on technology learning experiences based around different industry-ready technologies (e.g. machine vision inspection systems) that help workers to do their jobs better, less tediously, and more cost-competitively. Shared concepts encountered across lessons enable workers to spot opportunities to eliminate inefficiencies through technology. These grounded, sensible cases for technology adoption flip the status quo of "training after technology adoption" to "training that drives technology adoption", enabling the nation's 12.7M incumbent manufacturing workers to meet technological shifts and the changing nature of work on their own terms.

BootUP addresses a second manufacturing workforce threat: 469K unfilled entry-level positions, with no applicants in line to replace them. Largely, this stems from an unacceptable risk of entering training without a job guarantee — something no employer is willing to give. BootUP de-risks entry level hiring for all parties by connecting pre-workforce training and support, a collaborative two-week technical bootcamp, and a guaranteed job *interview*. Follow-up training and career coach access follow the trainee into early employment. A smartphone app maps this process and provides access to supports including coaches.

DIFFERENTIATORS

Most Track B projects are developing technology solutions to address the cost, effectiveness, and informational aspects of training. We complement these projects by developing programs to deliver them under real-world organizational conditions. Factory Forward interventions navigate complex personal and interorganizational constraints, allowing advanced technology training to reach companies and workers. These programs can easily be reused as vehicles for other content, media (e.g. VR), and in other sectors — including materials being developed other Track B teams.

ROAD MAP

Year 1: Pittsburgh pilot
Goal: Initial testing and intervention refinement

- **BootUP:** One cohort of 12 recruited, trained, placed in employment; data collected, program revised

- **SkillUP:** 3 content units served to 8 employers; data collected, program revised

- **Site selection:** Two additional regions (Sites B and C) chosen based on overlapping partner capability

Year 2: Pittsburgh expansion
Goal: Establish program scalability within a region

- **BootUP:** Two cohorts of 12

- **SkillUP:** 6 content units, 12 net employers

Year 2: Regional scaling test
Goal: Establish program scalability across regions

- **BootUP:** One cohort of 12 using new local partners and regionally adapted curriculum

- **SkillUP:** 3 content units to 8 employers with new local partners and regionally adapted curriculum

Year 3+: National sustainability
Goal: Leverage Phase II success to allow members of partner networks to stand up additional sites

- **Sustainability:** Sites apply for grants using Phase II as evidence of effectiveness

- **Activate management plan:** Engage national-level partners to coordinate and support program expansion

PARTNERSHIPS

In addition to researchers and developers at Carnegie Mellon University and the RAND Corporation, our team includes two prominent national networks with complementary roles in national workforce training. The National Institute for Standards and Technology's Manufacturing Extension Partnership (MEP) network is a nationwide network of providers who connect small and medium manufacturers to consulting and training services. These MEPs will co-design, adapt, and deliver training within their regions. Trainee recruitment, preparation, and support are conducted by Goodwill Industries International, a national leader in connecting vulnerable populations to stable careers through a whole-person approach. This effort is enhanced through a collaboration with Hope Street Group, whose SkillingAmerica curriculum is used to train career coaches (called "Navigators") to support trainees at a level beyond simple information or resource provision.

INTELLECTUAL PROPERTY

Factory Forward develops reusable training and collaboration models. SkillingAmerica belongs to Hope Street Group. CMU owns the SMART technician curriculum and microcertifications, as well as the learning management system (LMS) through which they are delivered. Licensing and spinoff opportunities for CMU-owned portions are available through CMU's Center for Technology Transfer and Enterprise Creation (CTTEC). Certain curricular technologies are released under open source license.

ADDITIONAL INFORMATION

We are open to additional partnerships that would expand the scope and generalizability of the pilot, or enhance its content. This includes sectoral agencies or foundations serving regional needs; partnerships in other sectors; and technology developers/vendors.

LEARNING ENVIRONMENTS WITH AUGMENTATION AND ROBOTICS FOR NEXT-GEN EMERGENCY RESPONDERS

Ranjana Mehta
rrmehta@tamu.edu

J. Gabbard   **A. Leonessa**
D. Srinivasan   **E. Du**
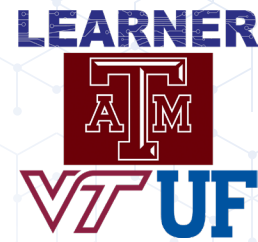
OVERVIEW

Emergency response (ER) workers perform extremely physically and cognitively demanding tasks that can result in serious injuries and loss of life. Human augmentation technologies (HATs) have the potential to augment workers physically and cognitively, thereby dramatically transforming the landscape of ER work, reducing injury risk, improving ER response, as well as helping attract and retain skilled ER workers. We propose *LEARNER, an agile and adaptive HAT-integrated ER training platform*, that accelerates HAT adoption for safer and efficient ER work, supports adaptive learning sensitive to ER workers' socio-technical opportunities and budgetary constraints, builds and retains skilled ER personnel, and ultimately accelerates next-gen workforce development across other industry domains.

DESCRIPTION

LEARNER is a novel mixed-reality learning platform that has physical, augmented, and virtual reality components, where ER personnel will learn to work effectively with two HAT classes: *powered exoskeletons (EXO) and head-worn AR interfaces (AR)*. These HATs will showcase the modularity of LEARNER across physical and cognitive augmentation that have distinct learning requirements. We will develop, integrate, and assess EXO and AR learning modules into the LEARNER system across different *access levels* (home to field house to training centers). Our industry partner SARCOS Robotics will design an upper-body EXO emulator interface integrated into the core LEARNER system. Concurrently, we will engage with our other industry partner Knowledge Based Systems, Inc. (KBSI), to develop and field a working LEARNER prototype that utilizes a unique paradigm for learning by adapting to a variable set of learners' characteristics and contexts, through the incorporation of physiological, neural, and behavioral markers of learning into real-time AR/VR scenario delivery. Finally, we will test and evaluate the HAT-integrated LEARNER prototype at a National ER Training Center (Texas A&M Engineering Extension Service; TEEX) in close guidance from our government partner (National Institute for Standards and Technology; NIST).

Completion of customer/market needs assessment and determination of LEARNER business model will move the prototype towards achieving scale in the ER community. We will work with existing and future partners to submit non-dilutive grants and to license copyrights to the LEARNER curriculum across ER and broader industry domains.

LEARNER will serve our responders by building a more capable and skilled ER workforce, safeguarding their health, improving their career longevity, and ensuring our nation's emergency preparedness.

DIFFERENTIATORS

Current VR based training platforms in ER are focused on extensively simulating the environment. LEARNER will *enable quick integration of emerging HATs (EXO and AR) into its platform* enabling an individual or a group of users to learn and collaborate in tomorrow's human-technology ER teams. Access to training resources (i.e., facilities, technology, budget, time away from duty) remains a critical barrier for effective and continued ER training. LEARNER *provides multiple training delivery methods, from the home to embodied immersive training,* that offer affordable and abundant opportunities for rapid repetition and skills refinement. ER workers are diverse in their makeup, in terms of their demographics, experiences, trust, and learning rates. In contrast, current ER training paradigms are static, with a one-size-fits-all approach. LEARNER utilizes *personalized learning algorithms to reduce skills gaps across ER teams* that enhance team operations.

LEARNER is *scalable across other work domains;* affording HAT learning in, for e.g., manufacturing, thereby creating opportunities for broader industry adoption.

ROAD MAP

Milestones: We will test novel ER-specific personalized training contents in immersive VR environments (m.5), test and release initial LEARNER software (m.7), establish AR/EXO-HAT learning module requirements (m.7), and demonstrate and release LEARNER prototype with integrated HAT modules (m.16). LEARNER pilot test will be completed with ER workers in a high-fidelity ER benchmarking study at TEEX (m.21). *Key inflection period* will be during the pilot test evaluation (m.21-23). Completion of LEARNER preliminary test with other Track B technologies (m.21) will provide strategies for broadening industry adoption for sustainability. *Key deliverables:* ER-based training scenarios, a personalized learning algorithm, AR/EXO-specific learning modules, and a functional LEARNER prototype.

PARTNERSHIPS

We have assembled a team of academic researchers across three universities, industrial (SARCOS Robotics; EXO integration), government (NIST; training standards/ testing), and non-profit ER workforce development (TEEX; ER training and evaluation) partners. Our advisory board is composed of leaders from ER stakeholder organizations (International Association of Fire Chiefs, National Volunteer Fire Council, National Fire Protection Association; to advise on ER training needs/constraints for LEARNER development), government institutions (FEMA, US Naval Research Lab, Texas Division of Emergency Management; to guide on achieving scale in ER and related markets), and relevant industry partners (Eksobionics, IHMC, Boeing, Ford, and American Society for Testing and Materials F48 committee on EXOs; to share experiences in technology development, rollouts, commercialization, and training and assessment in diverse industrial sectors). We will work with four other Track B teams to demonstrate scalability of LEARNER to manufacturing and construction.

INTELLECTUAL PROPERTY

The IP will include the LEARNER training platform, ER training scenarios, personalized learning algorithm, and EXO/ AR learning modules that would be protected in the form of copyrights and trademarks, registered through filings with USPTO to minimize infringement and illegal copying of materials. Licensees will be responsible for preventing infringement or illegal distribution of IP and IP owners would also make good faith efforts to prevent the same.

ADDITIONAL INFORMATION

LEARNER: http://www.me.vt.edu/learner

AI-BASED MULTI-LEVEL SKILL ANALYSIS AND ASSESSMENT FOR PERSONALIZED TRAINING

Aidong Lu
Aidong.Lu@uncc.edu

Chen Chen   **Weichao Wang**
Pu Wang   **Xintao Wu**

OVERVIEW

This project proposes to develop an immersive training platform that can be widely used in training and education. Our core technology is a set of AI-based skill analysis, assessment, visualization, and training recommendation tools. Our results will lead to scientific advances in our knowledge of augmented human abilities and practical solutions for integrating advanced AI in work and training. We will apply and test our platform in important sectors of nurses and firefighters, and create a community with our partners across the spectrum of training, education and workplaces to benefit both employees and employers. Our project has both societal benefit and commercial value for various training and skill assessment needs.

DESCRIPTION

This project addresses the critical needs for developing and sustaining a skilled technical workforce with an AI-assisted immersive training platform. We plan to develop both MR/VR versions that can be used flexibly for both formal on-site/off-site training and informal exercises/recruitment. Our platform will improve the learning experience, provide personalized instructions, predict job injuries, enable automatic skill assessment and job recommendation, and support visual analytics tasks for both workers and employers.

The core of our work is a comprehensive suite of AI-based analysis tools for understanding human actions and skills. We will encapsulate our tools with open-APIs to support various applications related to training and skill assessment.

Our platform primarily targets at two life-saving but overloaded workforces, firefighters and nurses. We will evaluate and create a comprehensive set of training modules for various online/onsite training needs through close collaborations among university fire engineering and nursing programs, fire departments, hospitals, and nationwide professional associations. We will summarize our evaluation results for the first-hand experiences and generalize them to other occupations.

With the developed platform that will be made publicly available, skilled workers will learn STEM skills and achieve easy transitions into new positions that require similar skills in later careers. The platform could be extended to a variety of skilled worker occupations such as health care and smart manufacturing.

This convergence research team specializes in complementary fields across disciplines, including computer science (AI, vision, behavior analysis, visualization, sensing, VR/AR), kinesiology, fire engineering and nursing.

DIFFERENTIATORS

Innovative Skill Assessment Technologies: different from traditional VR training platforms, our work focuses on skill assessment with support of advanced AI and big data technology, including multi-level motion analysis and modeling; training recommendation, skill certification, and future job recommendation; immersive motion visualization and visual analytics; smart motion capture via deep radar vision and environment sensing; injury analysis and prevention from kinesiology, that can be applied to other training programs.

Fundamental research on augmented capabilities of human through a systematic study of human performances on various skills, and the results will be summarized to guide and evaluate the future design of new augmented technologies.

Core software modules for skill analysis and personalized training with easy extension that can be provided to others through open platform APIs.

Benefiting both employees and employers with our data-driven platform and automatic skill assessment and job recommendation techniques.

ROAD MAP

Our long-term objective is to develop the next-generation of smart and augmented training platform for students and workers.

Our short-term milestones and deliverables include:

Year 1: (1) Core training modules for nurses and firefighters, (2) Complete curriculum for adoption at local participants and national associations, (3) Improve recruitment with serious games, (4) First-hand training experiences and evaluation results with hospitals and fire departments.

Year 2: (5) Offer our training at universities and online forums/platforms, (6) Establish new types of certificates through collaboration with NFPA, (7) Expand the platform with more skill sets and instructional features and reach out to the first responders and health care communities nationwide, (8) Promote our results to more participants and extend work to more occupations.

We plan to publish our training modules of nurses and firefighters free to the public and provide enhanced professional services to commercial users with advanced training modules and capabilities.

PARTNERSHIPS

We have three types of partners that span across the spectrum of education and training. Since our partners include both organizations offering training and workplaces recruiting workers, our collaboration will form a community for joint efforts on addressing the needs of the national talent ecosystem. We will continue to grow the community.

Higher Education - UNC Charlotte is our education partner who will develop new course modules and offer them to the undergraduate programs. The materials can be expanded to online courses and made available worldwide for broadening outreach.

Hospitals and Fire Departments - We have been collaborating on the development and evaluation of our training platform with participating hospitals and fire departments.

National/International Associations – We will work with NFPA (firefighter) and INACSL (nurse) to reach to more workers, publicize our results, and modernize their training/work worldwide.

INTELLECTUAL PROPERTY

Our project includes patentable and copyrightable technologies. We expect to submit provisional patent applications based on our accumulated results at the beginning of this project and follow up on the regular applications. We also plan to design and distribute copy-righted training programs for the application domains. All copyrightable inventions will be made available freely to all researchers working on the Project by means of royalty-free licenses for non-commercial use granted to all participating institutions for the duration of the Project, and can be negotiated with our university technology transfer office.

UNPACKING THE CAREER PATH

Denis Nekipelov
denis@virginia.edu

Tammy Wang

OVERVIEW

The proposed project will build a sustainable solution with an artificial intelligence (AI) tool to help candidates manage and tailor their career path through the complexities of the job market. The goal of this app is to educate and empower the American workers with a personalized analysis of the impact of their choices and their long term career outcome. Apart from the candidate's employment and educational history, other factors like their career interests, training options to skill themselves and market trends are also taken into account while providing personalized recommendations.

DESCRIPTION

Long term prediction for career trajectories of workers is a difficult task. Current unforeseen economic challenges and job displacements only add to the frustration.

We propose to alleviate this problem by developing a user-friendly app (PATH) which will educate candidates with potential career choices and provide a personalized recommendation to make their next career move. More specifically, the app will help the candidate explore his/her desired career goal and guide the candidate through the skills and experience levels required to attain the goal. The app will also provide the optimal trajectory by leveraging historical jobs data, market trends and AI-driven user-specific predictions.

Note that such an app is a critical missing resource in the job market today. Existing resources like job postings, professional networking apps, or generic career guides like blogs and books do not help a candidate with a personalized recommendation as proposed in PATH. Also, personal career mentors or counselors are far and few, and not available to most, especially minorities and underprivileged, owing to various cultural and socio-economic reasons. In fact, these groups are the target users for our app. Thus, we foresee a scalable app like PATH to have a significant impact across the demographic spectrum including women and minorities. It democratizes career mentorship by putting it in reach of everyone in the job market.

On the technical side, the heavy lifting prediction the of the app is based on rich industry data and done by an innovative econometric model which leverages recent advancements in machine learning, causal modeling, graph and natural language processing. Our preliminary model from Phase I already provides reasonable predictions and recommendations for candidates and will be extended at scale in Phase II.

A unique feature of the AI technology behind our tool is that it is designed to be modular built on a flexible framework and can be broadly applied to other job verticals. Co-PI Wang and PI Nekipelov will work with Phase II partners to seek and implement applications of this technology to other job verticals to provide advice on long-term career management to broader groups of American workers.

DIFFERENTIATORS

To our knowledge, a tool like PATH does not exist today. It has unique capacity that can offer career orientation to candidates even in complex job verticals such as high-tech industries. Moreover, the ability to scale across demographics and provide actionable insights for all candidates will be impactful to the society at large.

The team is deeply connected with a range of partners. These partners bring ample experience in launching sustainable long-term technology solutions which serve public interests. This will help our team not only build PATH but keep it running well beyond funding from the NSF.

PI and co-PI have already developed successful industry-scale platforms in their prior work in companies such as Microsoft and Zillow. The rich Silicon Valley recruiting industry data and experience from the Co-PI's recruiting firm (Riviera Partners) and a long track-record of developing real-world models for talent management systems are key advantages our team has for delivering this application.

ROAD MAP

The goal of the project is to launch the product in two distinct verticals with partners within two years. In the process, the sustainable approach to leverage the AI tool to continue to yield social impact will be identified, assessed and implemented.

In October 2020 the team will hire key personnel and research will be undertaken to identify the MVP feature set, develop features, and perform testing and collect feedback on these features. AI models will be researched and outreach will be performed to engage track partners.

Between October and December 2020, a COVID-19 career impact model will be created and related findings will be published to the partner network. In January 2021 the track integration plan will commence and internal feedback will be gathered for the product.

In March 2021 the first MVP will be released in the first vertical and user research will be conducted to plan for the first version launch. In the meantime requirements will be gathered for the next vertical. Advanced features including integration with Track Teams and soft skill training will be researched.

By July 2021, the team anticipates launching the Beta version for the first vertical. Around October 2021, MVP vertical 2 will be launched to collect user feedback.

The beta launch for vertical 2 is scheduled for December 2021. After this time the team will release integration feature updates and create a playbook around new partners. Shortly after March 2022, the team plans to launch with new partners and gather user feedback to perform further research. After July 2022, the product will be installed with new partners and the operation playbook will be updated to include the sustainable structure to operate the application to post grant period.

PARTNERSHIPS

The team will be working with partners from Phase I to conduct targeted user research through quantitative surveys, qualitative interviews, in-person focus groups to understand market needs and validate and fine-tune the applications to implement real-world use cases. Partners include Women in Product, WomenHack, Quality Assurance Commons for Higher and Postsecondary Education and Giving Tech Labs. Industry partners such as AISensa will be invited to help support the soft skill training as an additional function in the proposed application. Giving Tech Labs will help build messaging and positioning in Phase II. The partnership will expand to public sectors such as Virginia State Government, nonprofit organizations such as Center for Workforce Inclusion, and more via platforms such as X4Impact. These groups will be an important part of the product research and the go-to-market process. The project team also plans to partner with Track B teams to leverage each other's work such as job repository, data protection and data sharing.

INTELLECTUAL PROPERTY

The potential IP developed in the project using prior IPs will be guided under the policies of the University of Virginia and Riviera Partners. All partners will be required to sign a non-disclosure agreement and attest that all intellectual property created in this proposal is the sole property of Riviera Partners and the University of Virginia. Efforts are underway to patent and copyright the packages and any other proprietary technology or idea that may arise from project activities as a further method of protecting this intellectual property.

W. DOUGLAS MAUGHAN
Office Head
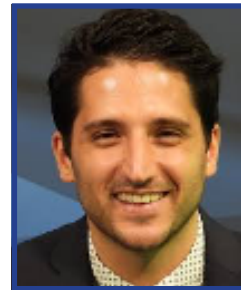dmaughan@nsf.gov

CHAITANYA BARU
Senior Advisor
cbaru@nsf.gov

LARA CAMPBELL
Program Director
lcampbel@nsf.gov

NAKHIAH GOULBOURNE
Program Director
ngoulbou@nsf.gov

IBRAHIM MOHEDAS
AAAS Fellow
imohedas@nsf.gov

LINDA MOLNAR
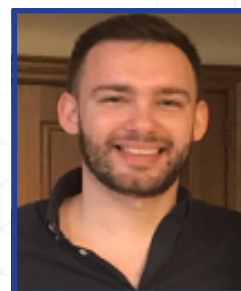*Program Director*
lmolnar@nsf.gov

MICHAEL POZMANTIER
Program Director
mpozmant@nsf.gov

TALITHA WASHINGTON
Program Director
twashing@nsf.gov

PRADEEP FULAY
Program Director
pfulay@nsf.gov

CODY BOWLES
Program Analyst
cobowles@nsf.gov

THANK YOU

ONLINE
nsf.gov/C-Accel

EMAIL
C-Accel@nsf.gov

FACEBOOK
facebook.com/US.NSF

TWITTER
twitter.com/NSF

YOUTUBE
youtube.com/user/VideosatNSF

INSTAGRAM
instagram.com/nsfgov