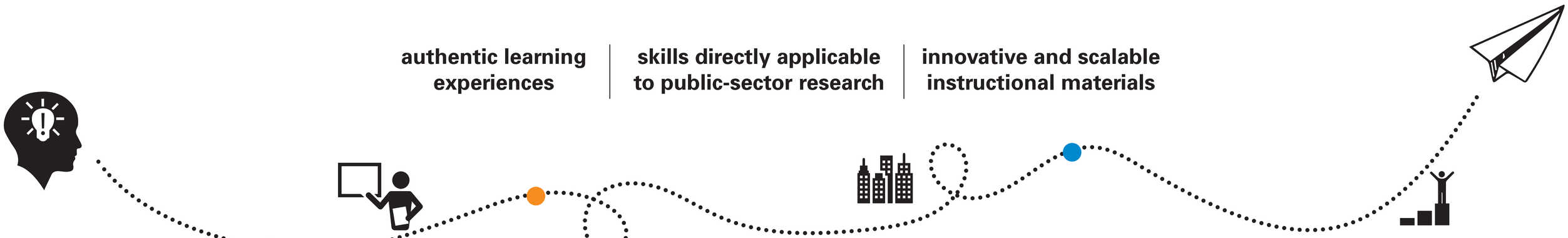
Preparing the Public-Sector Research Workforce to Impact Communities through Data Science

Libby Hemphill, Christopher Brooks, Lynette Hoelter, and Clifford A. Lampe; University of Michigan

*Training **undergraduate students**, **graduate students**, and **community stakeholders** in collecting, extracting, cleaning, annotating, and analyzing data generated and used by government organizations to further enable **data-based decision making**. Working with the Midwest Big Data Hub (MBDH), we identify interested **community partners** who will guide the development of innovative and scalable instructional materials and suggest relevant data sources for in-person and online training offered by faculty at the University of Michigan. Direct community involvement ensures **authentic learning** experiences centered on skills directly applicable to public-sector research.*

authentic learning experiences | **skills directly applicable to public-sector research** | **innovative and scalable instructional materials**



Citizen Interaction Design (CID) Course

within the University of Michigan School of Information (UMSI)

- Students work with community partners from the cities of Ferndale, Lansing, and Royal Oak, Michigan, to solve problems identified by the community partners
- Students use real data to solve real problems, devising an implementable cyberinfrastructure solution (not just recommendations) so they engage with the challenges and barriers faced in the public sector
- All course materials will be available under Creative Commons licenses

Network of Collaborators

- Frequent, engaged interactions with community partners in Michigan as well as throughout the Midwest Big Data Hub (MBDH)
- Collaborative work between project staff and community partners to develop content for CID course and MOOCs so learning experiences are authentic and solutions are valuable
- Establishment of a human infrastructure for supporting data science literacy in the public sector

Massive Open Online Courses (MOOCs)

with content devised from CID and relationships with community partners through the MBDH, freely available to anyone interested

MOOC 1: Building a Data Culture

- Understanding "data" and the different types of data needed to solve various problems
- Manipulating data, including preparing for and conducting basic statistical analyses
- Describing data patterns and presenting results visually and textually
- Learning CI tools available that take data analysis and presentation to the next level

MOOC 2: Using Advanced CI for Community-driven Data Science

- Setting up a local Docker image with tools (such as Git, Jupyter, Python 3+, and common data science modules) necessary for research
- Deploying the Docker image to cloud services, making use of the large-scale, low-cost computing infrastructure needed when handling large datasets
- Understanding reproducibility and transparency
- Following commonly accepted data science practices and workflows to identify, articulate, and answer potential research questions

NSF contract #: 1829724

SCHOOL OF INFORMATION UNIVERSITY OF MICHIGAN

ICPSR