

De-siloing Produce Safety Sampling Data through a Standardized Database

C. Barnett-Neefs¹, M. Stasiewicz¹

¹ Food Science & Human Nutrition, UIUC

Goal: Establishing the foundation of a produce safety sampling database to de-silo data
Significance: Food safety models have a limited capability to scale up to due a lack of data
Approach: Database Programming to enable AI-driven Geospatial Risk Analysis

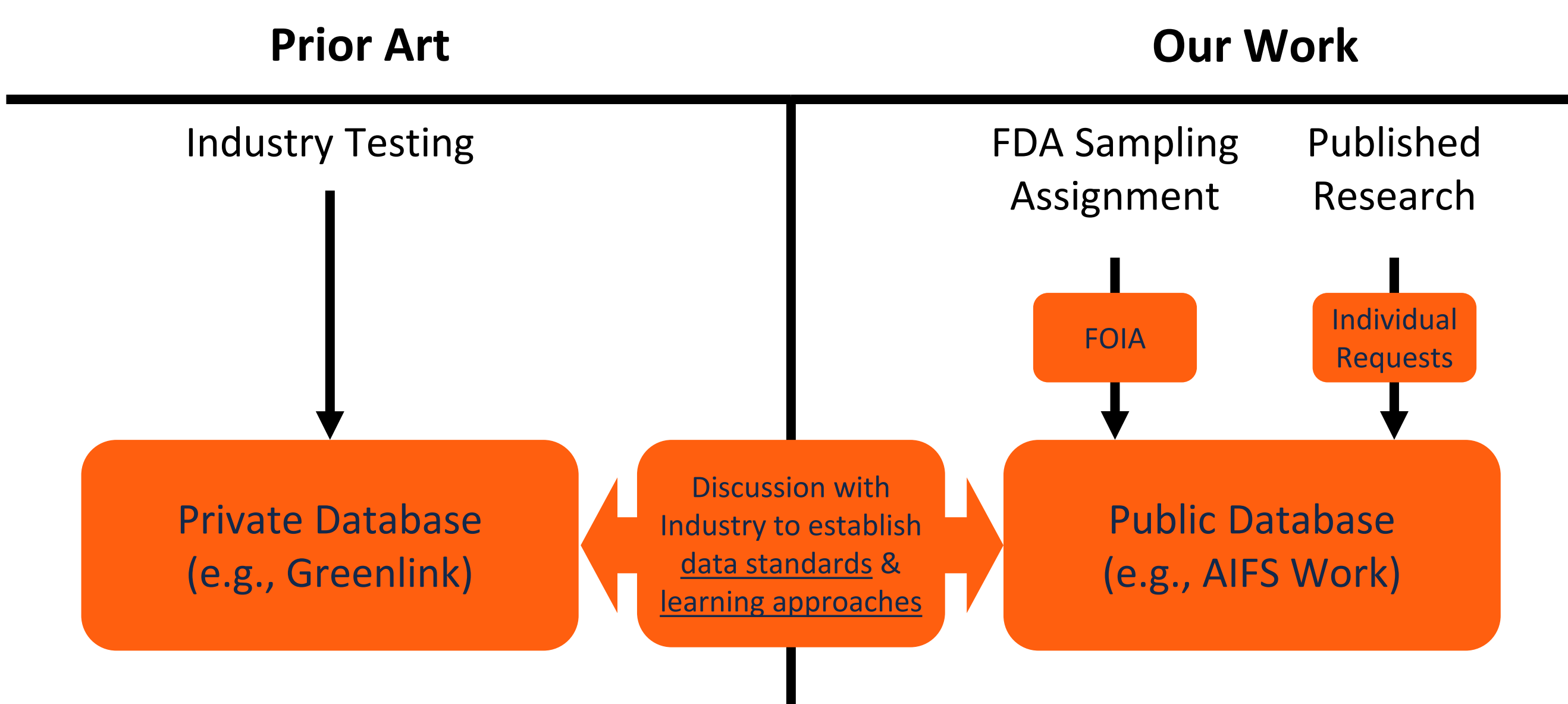
Vision

Establish a publicly-accessible unified produce sampling database to provide more informative data for the purposes of produce risk analysis and modeling

Database Creation

Sampling Data De-siloing and Standardization

- De-siloing of data can be accomplished through the creation of a public, standardized, produce safety sampling database
- Data sources include safety sampling assignments conducted by government organizations, and samples collected as part of published research
- Creating data standards is an ongoing process with industry to determine key factors involved



Database Seeding

- One potential source of starting data for the database may be projects funded by the Center for Produce Safety (CPS)
- Roughly 40 projects contain compatible pre-harvest data
- This database may also be integrated with organizations such as the CPS to provide an easily-accessed platform for researchers to publish produce sampling data
- Other potential future sources include governmental organizations (e.g., FDA Sampling Assignments)

Summary

- Established initial seed data sources and potential major data silos that could be integrated
- Determined preferable environmental factors to collect during novel data submission

Notable Outcome

Discussed with industry partners on data standards to use, database creation and hosting methodologies, major data sources to integrate, and how environmental risk factors may be included



Applications

Data Accessibility

- With sufficient momentum in the field, a de-siloed safety sampling database will likely encourage more usage, prompting a growth in data
- A unified database for produce sample testing represents a potential cornerstone for future produce safety projects, especially in the digital food safety space

Big Data Analysis Opportunities

- The additional dimensionality provided to the safety sampling data through environmental risk factor calculation may provide novel research opportunities
- With sufficient data, the safety sampling database may be an ideal candidate for AI-driven learning to better detect environmental risk factors for foodborne pathogens
- This data may be cross-referenced against other large-scale phenomena, such as animal migration patterns

Future Expansions

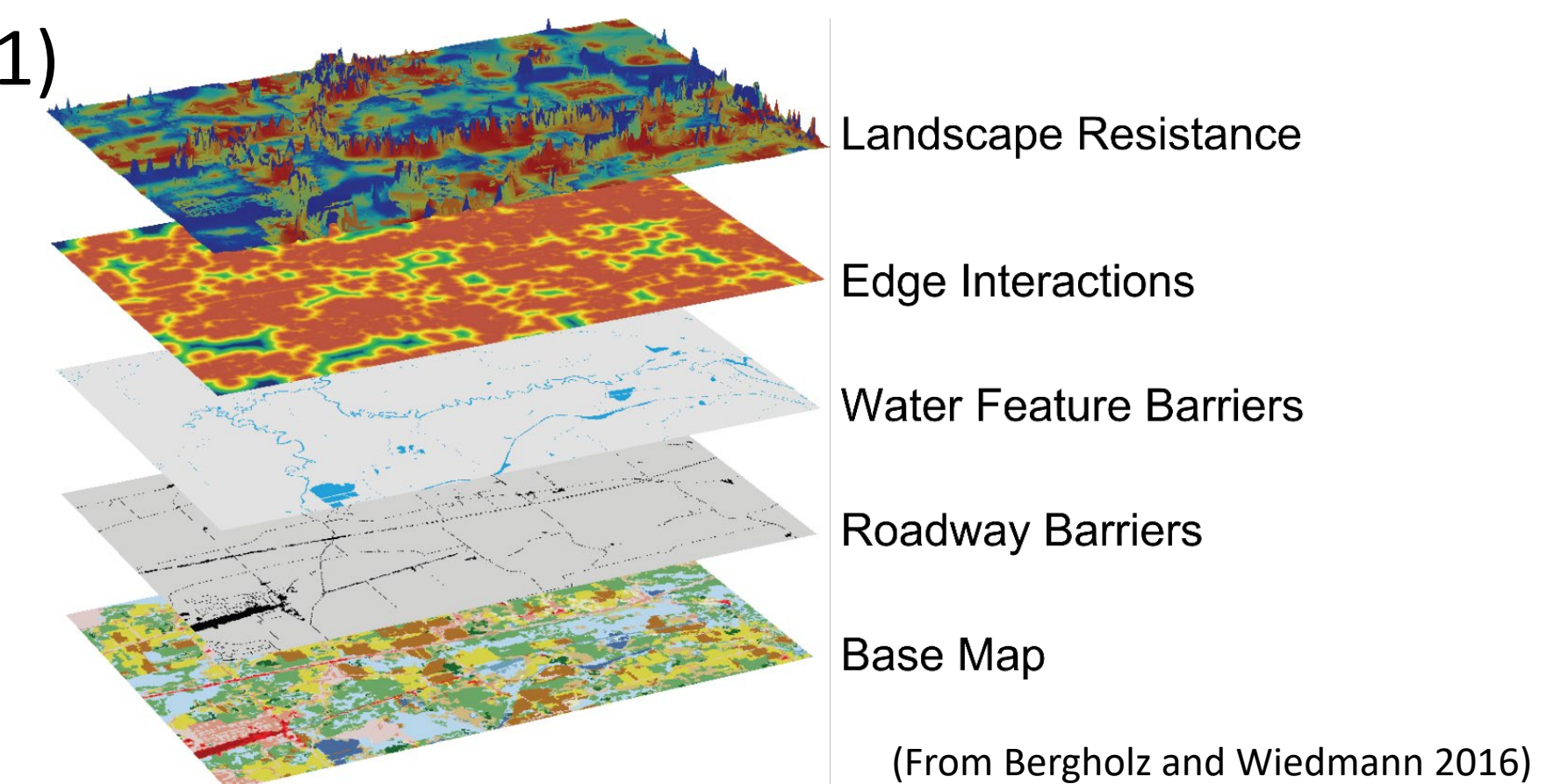
Pathogen Range

- The current scope of this project was limited to investigating factors and data associated with foodborne bacteria only
- Future development could enable the investigation of foodborne viral and parasitic threats as well

Geospatial Risk Analysis

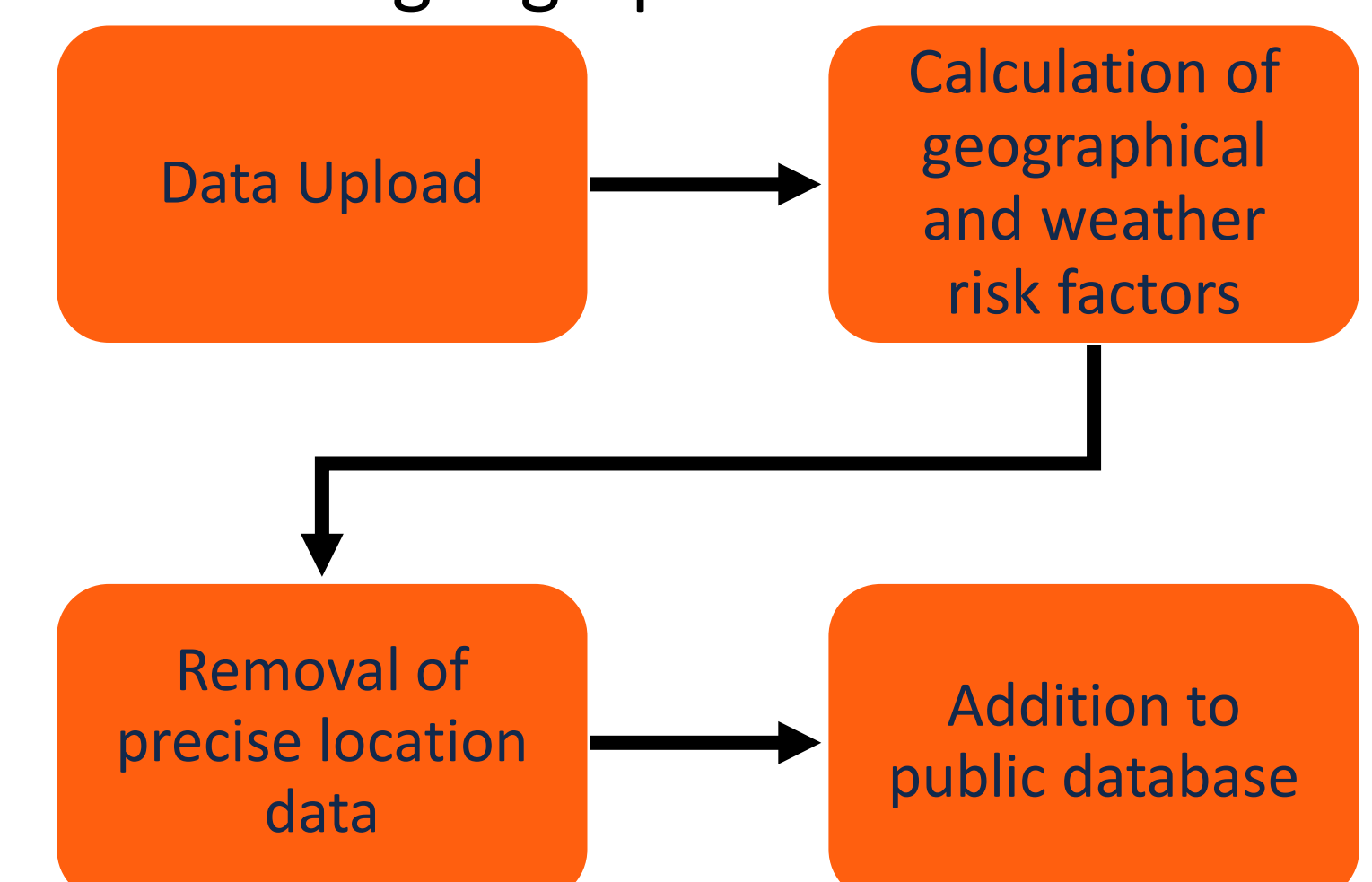
Environmental Risk Factors

- Several environmental risk factors have been identified as relevant to foodborne pathogens
- Factors can include climate, land-use, soil, and other spatial components (Liao, Guo et al. 2021)



Preservation of Anonymity

- As calculation of environmental risk factors requires precise location data, a processing step is required to remove it
- Upon upload to the database, submissions may undergo processing to convert precise location data into calculated risk factors based on local weather and geographical elements



Acknowledgement

This work is supported by AFRI Competitive Grant no. 2020-67021-32855/project accession no. 1024262 from the USDA National Institute of Food and Agriculture.



National Institute of Food and Agriculture
U.S. DEPARTMENT OF AGRICULTURE