

Projected sales of main products in 2013

Distribution of market share among the major industry players

# Community Inclusive Data Ecosystem Challenges and Opportunities



**Ayoung Yoon, Ph.D. & Andrea Copeland, Ph.D.**  
Indiana University Purdue University Indianapolis  
School of Informatics and Computing

# Data for Local Communities

<https://dataforcommunity.com/>



**Support for  
Communities  
Data  
Utilization**

● Phase 1.  
Community  
study (2017-  
2018)

Data intermediaries  
Community organizations  
General public

● Phase 2.  
Public library  
study (2018-  
2019)

Public libraries with  
data initiatives

● Phase 3.  
Prototype  
library service  
design (2019-  
2020)

3 partner libraries



# Potential for Data



DATA



KNOWLEDGE



ACTION

- Scholarly research
- Business
- Health
- Education
- Finance, etc.

- Reproducible research
- Federal, state level decision
- Policy change
- Budget, etc.

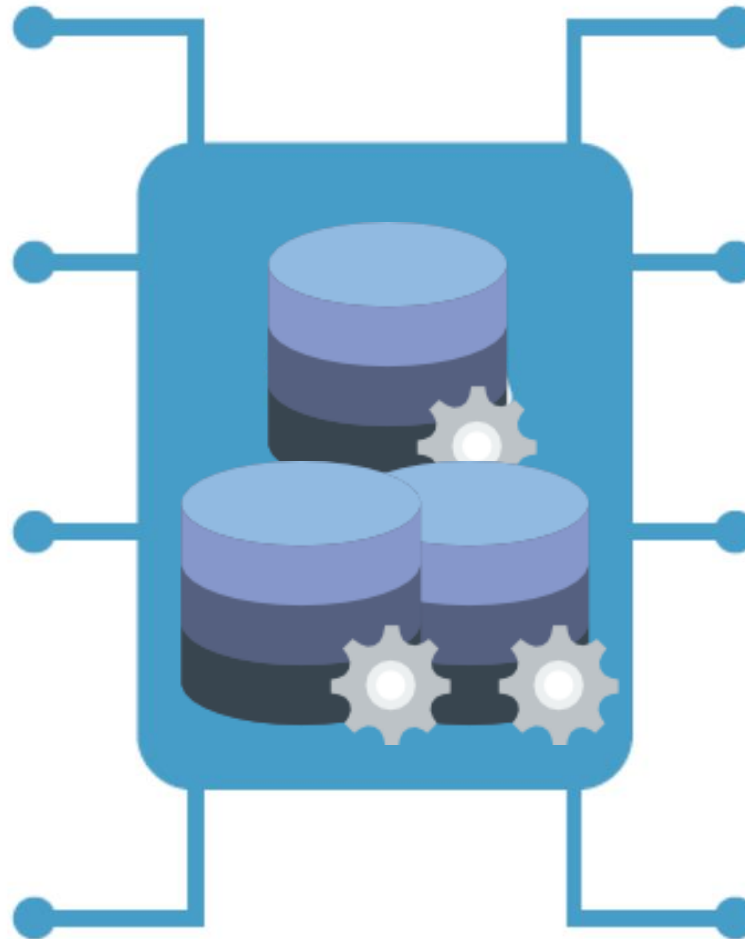
# Potential for Data in Smaller Communities

Everyday decision-making

Disaster planning

Design, monitor, and evaluate community policies

Increase institutional responsiveness



Community economic development

Meet community needs

Empower local/disadvantaged voices

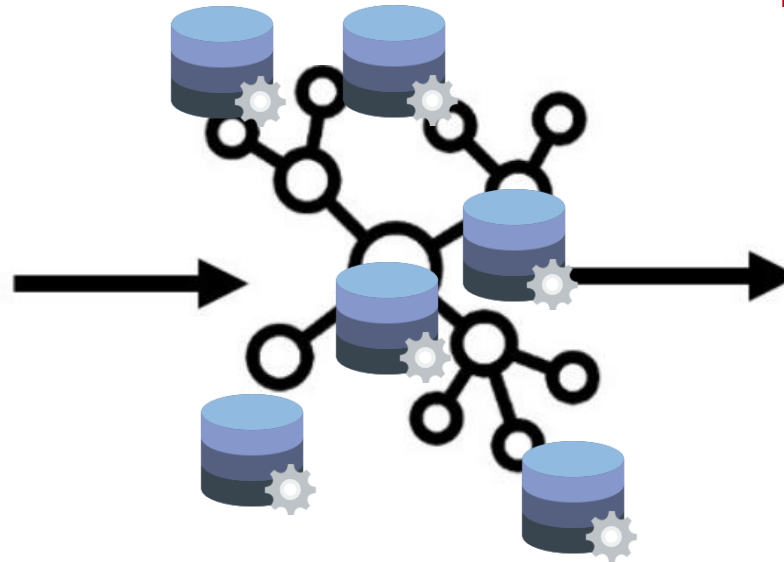
Enhance service delivery and effective service utilization

# Data Supply Chain

## Data Producers

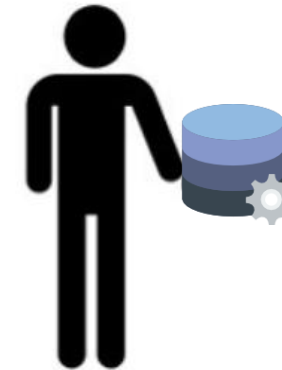


Government  
Research institution  
Business  
Non-profit



Repository  
Archive  
Network

## Data (re)users



Local government  
Community foundation  
Community non-profit  
Individual

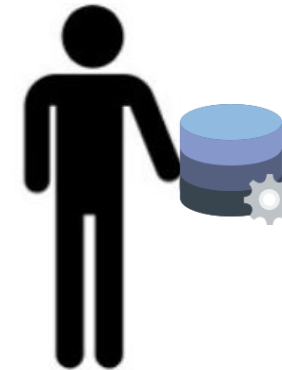
# Data Supply Chain

Data Producers

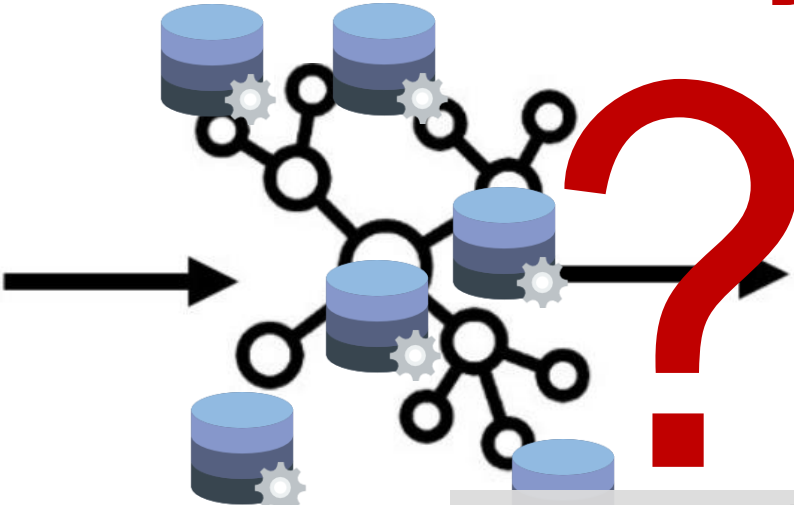


Government  
Research institution  
Business  
Non-profit

Data (re)users

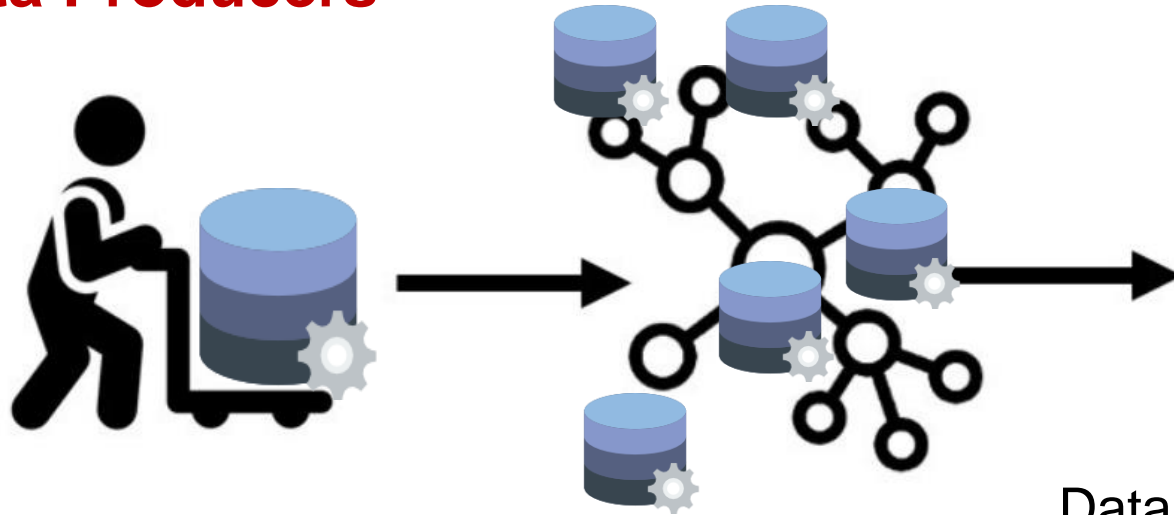


Repository  
Archive  
Network

- 
- Structural and technical infrastructure barriers for data access (Douglass et al., 2014)
  - Reluctance of data sharing (Yoon, 2017)

# Data Supply Chain

**Data Producers**



**Intermediary**



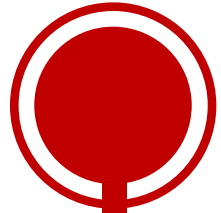
**Data (re)users**



Data context (Davies, 2014; Van Schalkwyk et al., 2014| Yoon et al., 2018 )

- Increasing data accessibility and utility
- Serving as 'keystone species' in data ecosystems
- Democratizing the impacts and use of open data

# Research Questions

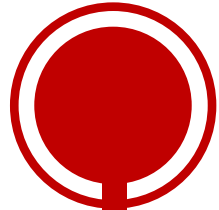


**What** are the **concerns** of communities regarding the utilization of data about their communities\*?

**WHAT challenges** communities faced during their data utilization process?



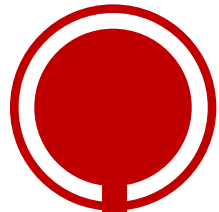
# Research Questions



## \*Communities

- Focusing on community-based organizations (CBOs)
- Modified definition from the National Community-Based Organization Network (n.d.)
- “Organizations characterized by locally-defined needs and services from locally-based and directed program designs, implementations, and evaluation.”

# Methods



## QUALITATIVE interviews

25 participants from community-based organizations across the states

Dallas, TX; Columbus, OH; Boston, MA; San Antonio, TX; Grand Rapids, MI; New Haven, CT; Indianapolis, IN; Milwaukee, WI; and Providence, RI.

Recruitment through data intermediary organizations in different U.S. cities

6 of 10 from National Neighborhood Indicators Partnership (NNIP)



# Organizational Profile



## Non-profit

Community foundations, impact initiatives, educational organizations, religious organizations, public libraries, and community service organizations

Missions: community development, including education, health, economic growth, safety, literacy development, domestic violence, culture, and the environment

# Organizational Profile

A great variety of data sources



- Census
- American Community Survey
- Economic data from the Bureau of Economic Analysis
- Data from the Department of Education
- KIDS COUNT data
- Bureau of Labor Statistics
- Data from the Department of Transportation
- Department of Health data
- Local level data such as the city school district data, regional voter records, regional Children's Services data, and community health data

# Concerns



“We’re hyper-focused on data.”

“Data are not the raw information but are, rather, contextual ”

The data needed to be accompanied with stories that were alive and supported the data.

# Concerns



“There [are] a lot of trust issues.”

- Access denial  
“I’ve asked for what data is going to come back to me. I’d like to learn about the air quality that you just tested in my backyard. (...) Basically they’re like, “Yeah, we don’t really give the data back.” ...As a resident of [name of town], that’s not acceptable.”

# Concerns



“There [are] a lot of trust issues.”

- Political use against communities

“if a journalist had access to the raw data, how might they use it? If a political opponent had, and so on and so forth, you know?”

# Concerns

“There [are] a lot of trust issues.”



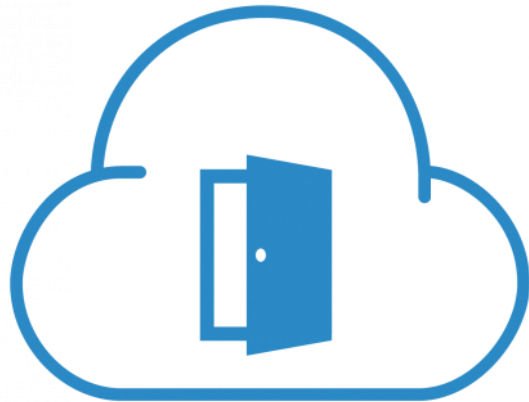
- Data quality

“We can’t collect the gold standard of data (...) data collection is based on what is feasible.”



# Challenges

## 1. Disparate approaches to data access

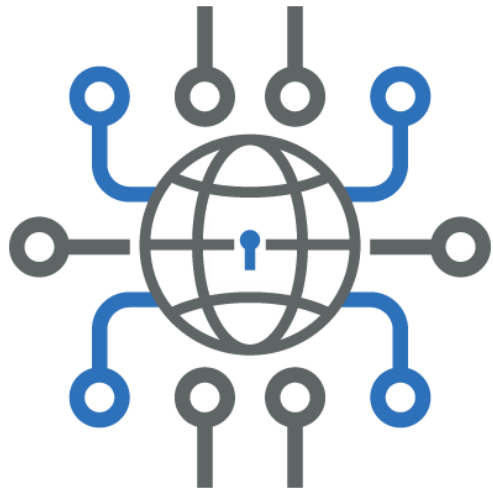


Not all open data are open.

- General public will not know “the data” exist.
- Staff capacity (data holders) is limited; time gaps when help is requested. (e.g., 1.5 year)
- Data confidentiality and privacy
  - Each stakeholders play different role; complexity of data sharing agreement; need more systematic approach.
- Issue of interoperability

# Challenges

## 2. Limitations of existing data in the community work context



The need for micro-level data

- Most micro level data are at zip-code-levels, which are “not particularly helpful in wanting to understand what’s really going on in a neighborhood”

Out-dated: “most data tend to be already old by the time we get it, because the government, census data is ten years.”

# Challenges

## 2. Limitations of existing data in the community work context



Need for (re)using private data

- Two major issues
  - “Data sets are not very trustworthy [as they] frequently have errors [because] they don’t have a regulatory context.”
  - Private data (with public nature) must be used through “exclusive sharing agreement[s because of the] security concerns.”
    - “We have a silent protest against (the notion of private data being private) ... we think that's a smoke screen for reducing oversight and public attention.”

# Challenges

## 3. Communities' widely varying capacity

### Data literacy



- Varied in staff skills and expertise related to data
- Data literacy as “a special skillset,” which “takes a lot of training and specialization.”
- More staff are “generalists, they weren’t trained as researchers, and there’s nobody on their staff who is either.”
- Strong preference working with data intermediaries or consultants.

# Challenges

## 4. Exclusive nature of existing infrastructure for communities' data work



### Community data infrastructure

- Many data owners generally “don’t invest in the creation of an infrastructure” that supports data sharing and (re)use for communities; “if they do, typically they want to own it.”
- Most existing data exchange systems as “still crippled on the backend...and there’s often some budget hungry proprietary software firm gatekeeping access to the data model of how the whole thing actually works.”

# Challenges

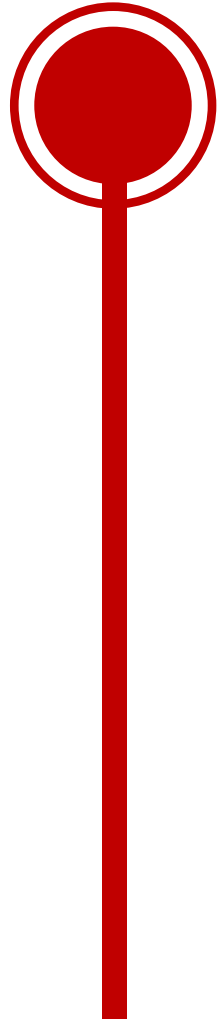
## 4. Exclusive nature of existing infrastructure for communities' data work



### Unfriendly experience

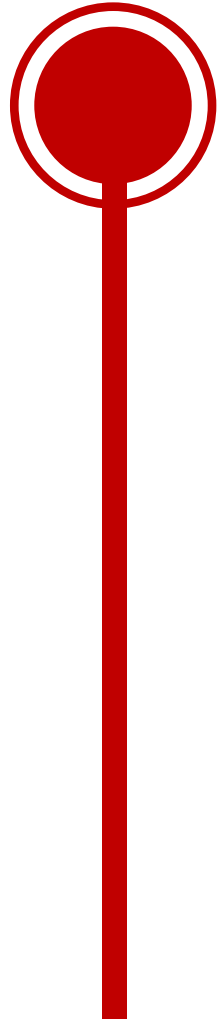
- No response when requesting data or negative response when working with state agencies.
- Difficulties working with academic partners due to different interests.
- Inequality in partnership and resource distribution when conducting community research

# Summary of Challenges



- Fear from communities regarding data potential and use.
- Existing data access systems are not always community friendly.
- The needed level of granularity and locality are not always reflected in existing datasets.
- The current data supply chain does not support the often time-sensitive nature of community work.
- Many CBOs experienced difficulties of training staff or hiring data experts within their organization.
  - Lack of local training opportunities.
  - Not all has local data intermediaries.

# At the same time

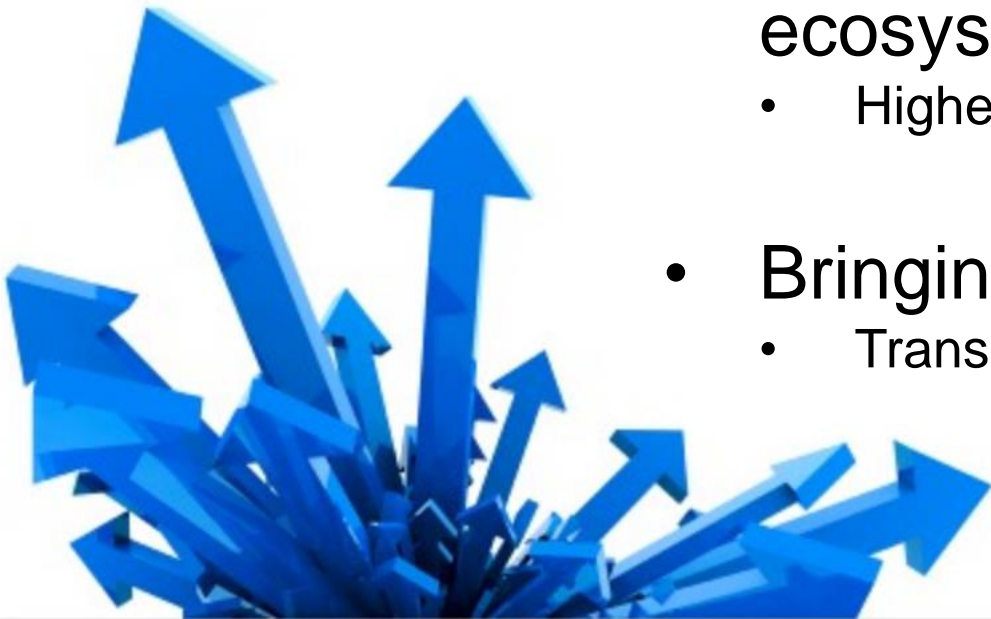


There are many many good stories about data, data (re)use, and data work from the communities.

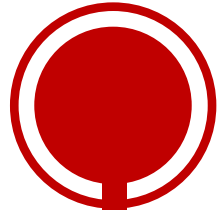


# Opportunities

- Creating educational programming
  - e.g. Microsoft and the Urban Institute's NNIP's data training; Data 101 by Western Pennsylvania Regional Data Center and the Carnegie Library of Pittsburgh
  - Extending or building partnerships within data ecosystems
    - Higher Ed, data intermediaries, public libraries, government, etc.
  - Bringing community voices into data ecosystems
    - Transparency, inclusiveness, active users (not end-users)



# Next questions for our future study



Has open data contributed to the development of data literacy skills?

How to assess their impact on data literacy skills acquisition?

Has open data increased participatory community-based problem solving?

# ACKNOWLEDGMENTS

This research is supported by the **Institution of Museum and Library Services (IMLS)** through the National Leadership Grant for Libraries program (#LG-96-17-0184-17), “Data Reuse for Local Community.”

The project team would also like to thank the nine data intermediary organizations and the staff who participated in our study.



# Data for Local Communities

<https://dataforcommunity.com/>

Yoon, A., & Copeland, A. (under review). Toward community-inclusive data ecosystems: Challenges and opportunities for community-based-organizations.

Yoon, A. & Copeland A. (in press). Understanding Social Impact of Data on Local Communities. *Aslib Journal of Information Management*.

Yoon, A., Copeland, A., & McNally, P. (2018). Empowering communities with data: Role of data intermediaries for communities' data utilization. *Proceedings of Association for Information Science and Technology (ASIS&T) Annual Meeting*, 55 (1), 583-592. <https://doi.org/10.1002/pa2.2018.14505501063>

Copeland, A., & Yoon, A. (2018). Data Reuse and Library Programming. *Indiana Library Federation (ILF) 2018 Annual Conference*. Indianapolis, IN.

Yoon, A. (Sep 2017). Data for community development and decision making. *Research Data Alliance (RDA) 10th Plenary Meeting*. Montreal, Canada.

# QUESTIONS?

## Project Website

<https://dataforcommunity.com/>

## Project Team

PI: Ayoung Yoon,  
[ayyoon@iupui.edu](mailto:ayyoon@iupui.edu)

Co-PI: Andrea Copeland,  
[ajapzon@iupui.edu](mailto:ajapzon@iupui.edu)

Project Assistant: Paula McNally,  
[pjmcnall@iu.edu](mailto:pjmcnall@iu.edu)