# Integrating Public Health Researchers into the Open Data Ecosystem

**Erika Martin, PhD MPH**
**Rockefeller Institute of Government & University at Albany**

---

**California Health and Human Services Open DataFest III**
**Sacramento, CA**
**March 14, 2016**

*This work was supported by the Robert Wood Johnson Foundation's Public Health Services and Systems Research Program (grant ID #71597)*

# Project overview

- Limited guidance on tailoring open data to different users
- Open data are only valuable when used
- How can we improve the quality and usability of data for public health research and practice?
  - ***Systematic review of health data offerings*** in three open data portals (HealthData.gov, Health Data NY, NYC OpenData)
  - ***Key informant interviews with practitioners*** publishing open health data to understand challenges and opportunities
  - ***Pilot open data linkage project*** to assess the feasibility of using open data for academic research

## CHARACTERISTICS OF DATA USE

- **Data characteristics**
  *(e.g. missing data, timeframe, data collection procedures, database design, data elements, population)*
- **Data user characteristics**
  *(e.g. intended use, expertise, skills, tasks performed)*
- **Platform promotion and user training**
  *(e.g. value propositions, financial resources, political support, information technology, regulations and data stewardship, legal interpretation of confidentiality protections)*

## DATA QUALITY AND USABILITY

- **Intrinsic data quality**
  *(e.g. accuracy, reputation, confidentiality, reliability, validity, objectivity)*
- **Contextual data quality**
  *(e.g. appropriate amount, completeness, concise representation, ease of manipulation, relevance)*
- **Platform usability**
  *(e.g. accessibility, functionality, learnability, representational consistency, visibility)*
- **Metadata quality**
  *(e.g. accuracy, completeness, consistency, interpretability, provenance)*

## DIRECT AND INDIRECT HEALTH IMPACTS

- **Short-term impacts**
  *(e.g. availability of health information, data-driven population health planning and monitoring, mHealth development, consumer empowerment, research grants and studies)*
- **Long-term impacts**
  *(e.g. improved population and patient health, enhanced decision-making, higher quality/value medical and public health services)*

# Systematic review of open data offerings

***(HealthData.gov, Health Data NY, NYC OpenData)***

- Most data offerings not designed for health research
  - Only one-quarter of open data offerings are structured datasets
  - Most offerings do not contain demographic variables
- Variation in quality and usability across platforms
  - Health Data NY scored highest on intrinsic data quality, contextual data quality, and adherence to Dublin Core metadata standards
- Gaps in meeting "open data" deployment criteria
  - All offerings met basic "web availability" open data standards
  - Fewer met higher standards of being hyperlinked to other data to provide context
- Platforms enable users to discover and access data in novel ways, with areas for improvement
  - Technical problems limit functionality, low web visibility, HealthData.gov is primarily a search engine

# Key informant interviews

## *(Policymakers and practitioners in New York and federal agencies)*

- Wide range of perceived benefits
  - *Internal benefits*: improved data/metadata quality, more efficient public health operations (e.g. data silos, FOIA requests)
  - *External benefits*: health literacy, data-driven improvements in healthcare delivery and built environment, community empowerment, improved data quality, timeliness, and usefulness
  - New users bring innovative ideas
- Numerous challenges to releasing data
  - *Critical challenges*: resources, cultural resistance, legal and regulatory issues, and data/metadata quality
  - *Other challenges*: technical issues with legacy systems and data platforms, knowledge gaps, addressing needs of diverse end-users
- General optimism that open data movement will continue
  - Yet success depends on sustained leadership, resources, cultural changes, promoting the use of data, and establishing governance

# Pilot data linkage project

## *(Mapping childhood obesity to the built environment in New York)*

- Many datasets readily available for public health research
  - Can use data creatively to evaluate multiple dimensions of the built environment (e.g. using restaurant inspections data for fast food availability)
  - Can synthesize data from different domains (health, agriculture, education)
- Challenges consistent with findings from other study phases
  - Lack of standard definitions for data elements severely constrains interoperability and ability to merge by geographic identifier
  - Incomplete metadata, e.g. missing codebooks
  - Data quality, e.g. incomplete addresses, inconsistent location descriptions
  - Data timeliness
  - High level of geographic aggregation limits value
  - Some data not easily discoverable (or available) in open data platforms
  - Data not yet 5-star, e.g. downloadable in multiple non-proprietary formats and with links to provide context
  - Limited usability, e.g. advanced statistical skills required to prepare data

# Questions?

- Contact:

  **emartin@albany.edu**

- For additional project information:

  **www.publichealthsystems.org/erika-martin-phd-mph-0**

- For materials from fall 2013 workshop on open health data in New York and links to open data resources:

  **www.rockinst.org/ohdoo**

# Bibliography


- Martin EG, Helbig N, Shah NR. Liberating data to transform healthcare: New York's open data experience. *JAMA* 2014; 311(24): 2481-2481.
- Martin EG, Helbig N, Birkhead GS. Opening health data: what do researchers want? Early experiences with New York's open health data portal. *J Public Health Manag Pract* 2015; 21(5): e1-7.
- Martin EG, Law J, Ran W, Helbig N, Birkhead GS. Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on three open data platforms. *J Public Health Manag Pract.* [In press; online ahead of print]
- Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. [Under review]
- Dwicaksono A, Brisette I, Birkhead GS, Bozlak CT, Martin EG. Evaluating the contribution of the built environment to obesity among New York State students. [Working paper.]