

Toward a Purpose and Question-Driven Open Data Ecosystem

Keynote Address

National Open Data Conference 2021:
The Importance of Data Quality

Stefaan G. Verhulst

2021



MISSION: TO IMPROVE THE WAY WE MAKE DECISIONS IN THE PUBLIC INTEREST ...



GOVLAB

ABOUT THE GOVLAB

The GovLab's mission is to improve people's lives by changing the way we govern.

Our goal is to strengthen the ability of institutions – including but not limited to governments – and people to work more openly, collaboratively, effectively and legitimately to make better decisions and solve public problems.

We believe that increased availability and use of **data**, new ways to leverage the capacity,

intelligence, and expertise of **people** in the problem-solving process, combined with new advances in **technology and science** can transform governance.

We approach each challenge and opportunity in an **interdisciplinary, collaborative** way, irrespective of the problem, sector, geography and level of government.

WHAT WE DO

BUILD

We develop and test prototypes for new platforms to enable open and collaborative problem-solving inside a variety of organizations.

LEARN

We assess and disseminate information and analysis about governance innovations (successes and failures) globally and their impact on people's lives.

CONNECT

We build networks of innovators from across disciplines to gain deeper insights and develop new frameworks for improving governance.

TRAIN

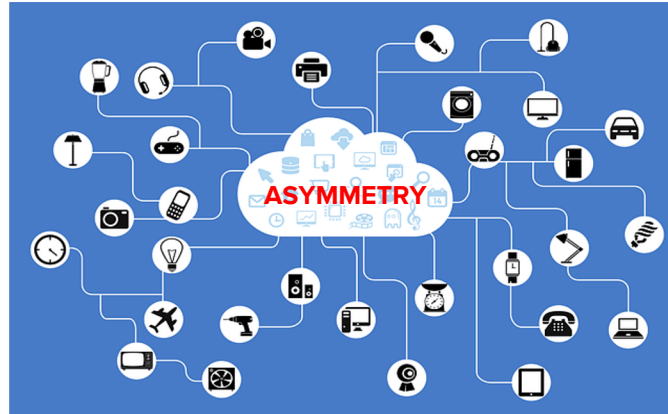
We train entrepreneurs to tackle public interest challenges using open innovation to take projects from idea to implementation.



CHALLENGE OF OUR DATA AGE: DATAFICATION AND DATA ASYMMETRIES

Digital Transformation

DEMAND



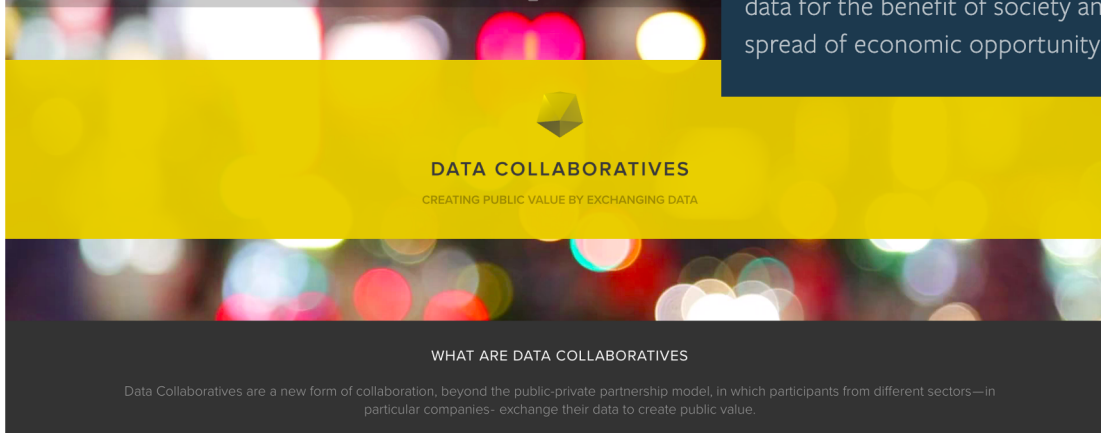
SUPPLY

Mis-Use / Re-Use



ACCELERATING ACCESS TO DATA FOR RESPONSIBLE RE-USE

<https://datacollaboratives.org/>



NEW | Open Data Policy Lab City Incubator



Open.Data.Policy.Lab

The Open Data Policy Lab supports decision-makers at the local, state and national levels as they accelerate the responsible re-use and opening of data for the benefit of society and the equitable spread of economic opportunity.

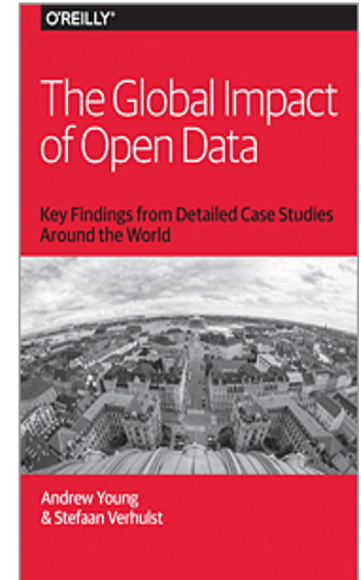
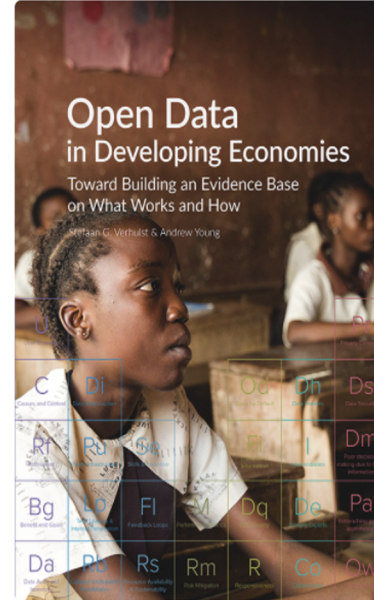
Sign up for updates

Sign Up

<https://opendatapolicylab.org/>



MEASURING THE IMPACT OF OPEN DATA





IMPACT DEPENDS ON... CLEAR PROBLEM DEFINITION

Problem and Demand Definition	Capacity and Culture		Governance		Partnerships	Risks
U User Research	Di Data Infrastructure		Od Open by Default		Pr Privacy Concerns	
C Causes and Context					Ds Data Security	
Rf Refinement	Pu Public Infrastructure	Se Skills & Expertise			Fi Freedom of Information	I Intermediaries
Bg Benefit and Goals	Lp Tech Literacy & Internet Penetration	Fl Feedback Loops	M Performance Metrics	Dq Data Quality	De Domain Experts	Pa Entrenching power asymmetries
Da Data Audit and Inventory	Rb Cultural/ Institutional Roadblocks	Rs Resource Availability & Sustainability	Rm Risk Mitigation	R Responsiveness	Co Collaborators	Ow Open washing

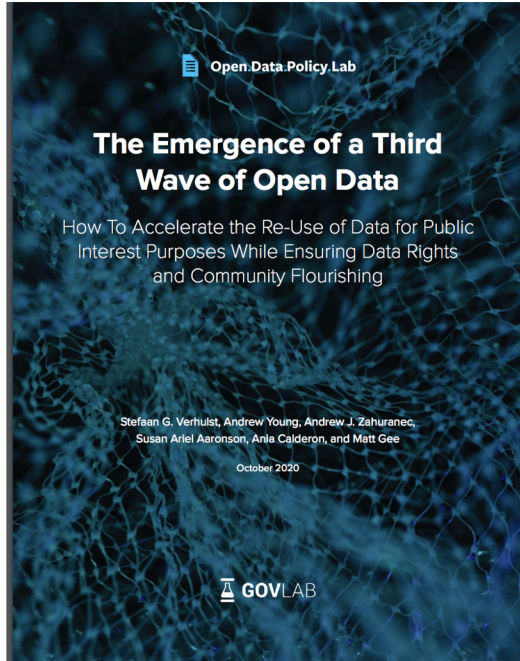
THE CHALLENGE OF BEING DATA-SUPPLY FOCUSED

- Today, too many data initiatives start from the data supply, NOT the **questions** one seek to answer...
- Similarly too many AI initiatives fail to specify which societal objectives they seek to optimize for.





TOWARD A THIRD WAVE OF OPEN DATA



A New Science of Questioning

THE 100 QUESTIONS

UNLESS WE DEFINE THE QUESTIONS WELL...
TO UNLOCK THE POTENTIAL OF DATA AND DATA SCIENCE
HOW CAN WE PROVIDE ANSWERS THAT MATTER?

100 Questions Methodology



The 100 Questions: Toward Demand-Driven Data Collaboration



Migration



Gender



Air Quality



Future of Work



Disinformation



Governance



Food Systems
Sustainability



I. TOPIC MAPPING & DEMAND ASSESSMENT

- A **topic map**: a conceptual understanding of a complex system for collective action purposes.

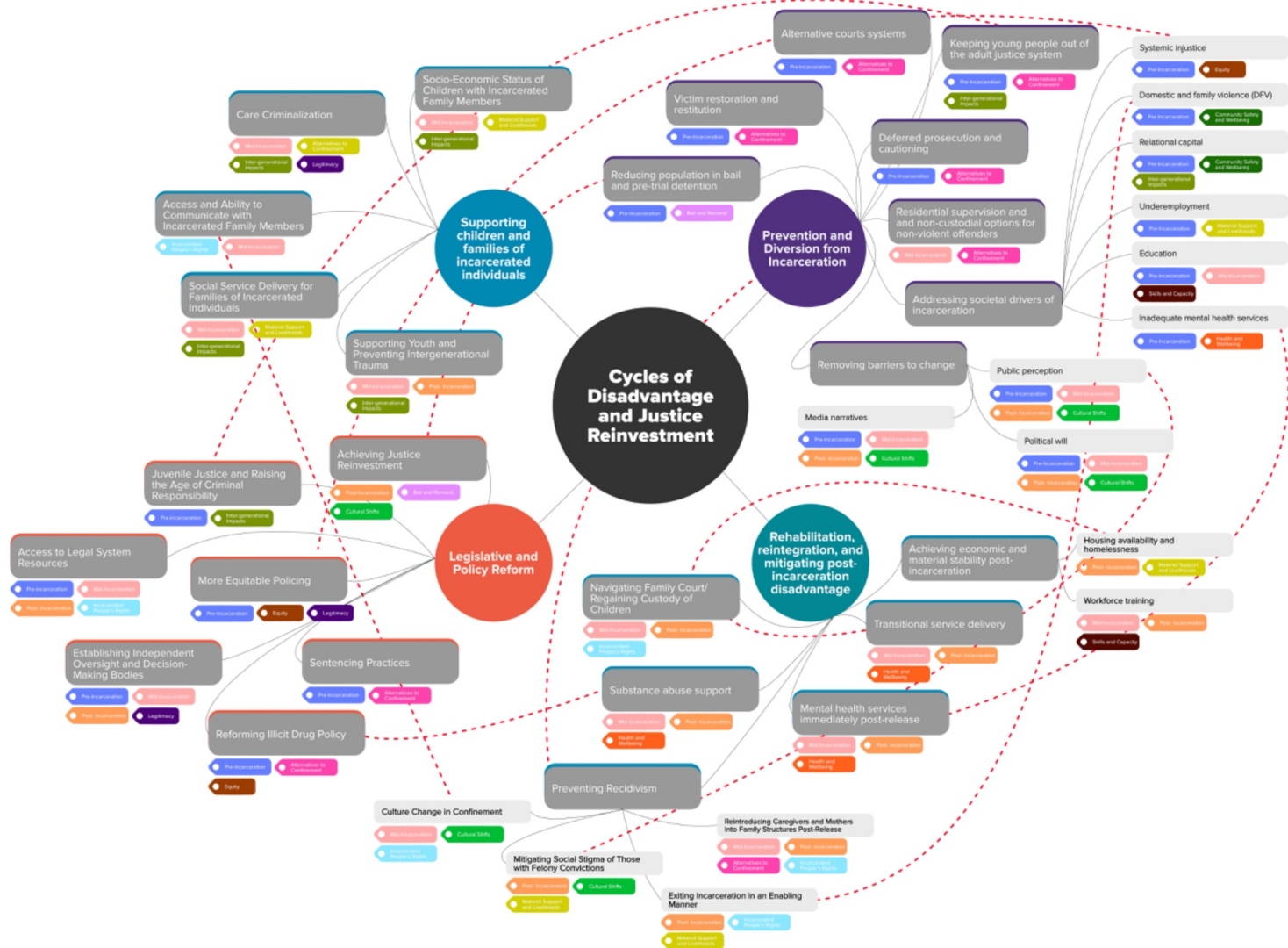


Adolescent mental health: Using a participatory mapping methodology to jointly identify key topics, questions, and priorities for future work and data collaboration

Data and Research Mental health

Share this story:

By Alexandra Shaw, Andrew J. Zahuranec, Andrew Young, Stefaan G. Verhulst, Jennifer Requejo, Liliana Carvajal





II. ACTOR MAPPING & BUILDING A BILINGUAL COMMUNITY

BILINGUAL COMMUNITY

"Bilinguals" are practitioners who possess both relevant domain knowledge and data science expertise. They are experts who understand the importance of data, are aware of the ways data can inform decision-making, and can provide new actionable insights to educate the community on where and how to leverage data responsibly.



A. Kofi Amegah
Senior Lecturer
University of Cape
Coast (Ghana)



Aafreen Siddiqui
Regional Engagement
Lead, Government
Innovation
UNDP



**Abhijnan
Chakraborty**
Post Doctoral
Researcher
Max Planck Institute
for Software Systems
(MPI-SWS)



Abhishek Gupta
Founder
Moral AI Ethics
Institute



Abigail Hunt
Research Fellow
Oversara
Development Institute



Adam Hughes
Associate Director of
Research
Pew Research Center



**Catherine
Witherspoon**
Senior Advisor
ClimateWorks
Foundation (CWF)



**Catriona
Marshall**
Policy Analyst /
Economist
OECD



Ceren Budak
Assistant Professor
University of Michigan
School of Information



Charley Johnson
Program Director
Disinformation Action
Lab, Data & Society



Charlie Terrell
Director of the
National Labor
Exchange
NLEWS



Chenai Chair
Research Manager
Web Foundation,
Gender and Digital
Rights



**Reecha
Upadhyay**
Portfolio and
Communications
Manager
Clean Air Fund



Reema Nanavaty
Executive Director
Self-Employed
Women's Association
(SEWA)



Regina Honu
CEO
Soromo Solutions



Reham Tamime
PhD Candidate
Web and Computer
Science, University of
Southampton



Renee DiResta
Technical Research
Manager
Stanford Internet
Observatory



Roy Koslowski
Professor / Director of
the Master of
International Affairs
Program Political
Science International
Affairs
University at Albany
(SUNY)



A GLOBAL BRAINTRUST OF BILINGUALS



Migration

Gender

Air Quality

Future of Work

Disinformation

Governance

Food Systems
Sustainability

70 bilinguals

93 bilinguals

80 bilinguals

90 bilinguals

139 bilinguals

134 bilinguals

60 bilinguals

666 total bilinguals



III. SOURCING QUESTIONS THAT MATTER

Types of Insight



SITUATIONAL AWARENESS

Initiatives that leverage data to enable stakeholders across sectors to better understand trends and geographic distribution.



CAUSE-AND-EFFECT INSIGHT

Initiatives that can help stakeholders better understand the key drivers and consequences of different phenomena.



PREDICTION

Initiatives that leverage data to enable new predictive capabilities and allow stakeholders to assess future risks, needs, and opportunities.



IMPACT ASSESSMENT

Initiatives that determine which, whether, and how various interventions meet certain objectives, including the Sustainable Development Goals. They can also seek to understand impediments to certain objectives or the success of particular interventions.



IV. PRIORITIZING QUESTIONS THAT MATTER

100Q seeks to source **questions that matter**. To do so, the following criteria is used when submitting high-impact questions:

What makes a good (data) question?

1. **Practical and/or Scientific Impact** — potential to transform our understanding of the problem, fuel innovation or improve how we address societal challenges;
2. **Novelty** — questions that have not been answered or analyzed before with answers that will confirm, refute, or extend previous findings;
3. **Feasibility** — likelihood that the question could be answered with existing data and methods;
4. **Quality** — overall alignment with the above criteria.

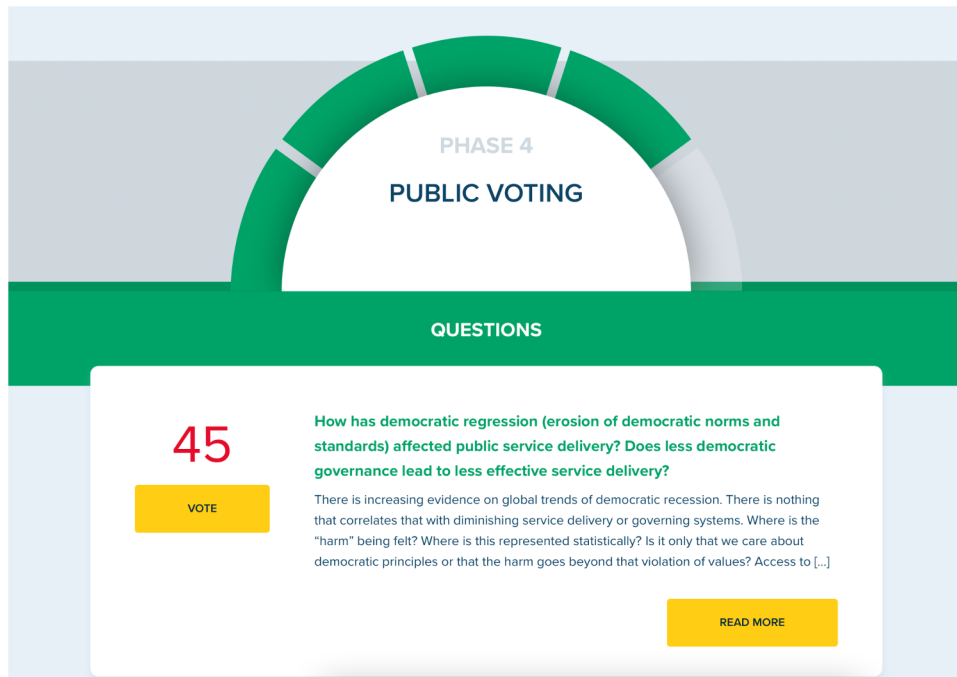
3 → [FOOD PRODUCTION AND PROCESSING]

Please select your top 10 questions across the 5 subgroups.

Choose as many as you like

- A** [Production Tools] How are the existing datasets helping agriculture players? Where are the gaps and limitations in harmonizing indicators for sustainability in food systems?
- B** [Production Tools] What is the main reason why modern biotechnology is not widely adopted?
- C** [Consumption Patterns] What are the health risks and benefits of the currently prevailing high-calorie diets and what are the health, socio-economic, and environmental costs and benefits of a global shift to healthier, nutritious, and diverse diets?
- D** [Smallholder Farmers] How do we incentivize/reward individual farmers and land managers to optimally use their land to achieve local, regional, national, and international targets to produce the right balance of foods in the most sustainable way?
- E** [Consumption Patterns] Can we build a detailed map of food consumption among the population globally?
- F** [Production Tools] What prevents us from enhancing water use efficiency, including multilateral cross-border initiatives to improve water use?

Bilingual Clustering & Prioritization



Public Voting Campaign



V. PURPOSE-DRIVEN OPEN DATA

NEW: | BD4M Studio Report on Migration Flows in West Africa



BD4M

Big Data for Migration Alliance

3 DATA HOLDERS

Who holds or has access to those types of data within the region?

2 DATASETS

What are the minimum viable datasets needed?
What are the required characteristics

- Location data? Time based data?
- Structured/Unstructured? PII?

4 DATA USERS

Who needs access to the data and/or can act upon the insights?

5 INCENTIVES AND CHALLENGES

What are incentives and/or challenges for data holders to engage on the questions at hand? (e.g., corporate social responsibility strategy, organization, reputation, etc.)

ENABLING CONDITIONS

What are enabling conditions for data holders toward establishing a collaboration? Who can act as enablers?

- What lessons have you learned from similar efforts in other contexts, if any?
- Internal capacity (data literacy, legal, etc.)
 - External support for data initiative

1 PROBLEMS TO BE ADDRESSED

What are the key questions?
What is the context?

7 BD4M DATA COLLABORATIVES

What type of Collaborative would work best given the context, partners and needs? What can the BD4M Alliance support with?

6 CAPACITY AND OTHER NEEDS

What are capacity and other needs to address through a collaborative? (expertise, resources, et al)

ENABLING CONDITIONS

What are enabling conditions for data users to establishing a collaboration? Who can act as enablers? What lessons have you learned from similar efforts in other contexts, if any?

- Build long-term data capabilities for responsible re-use
- Legal and ethical framework for the data used



ALLOWS FOR ENGAGEMENT AT #LEVELS

THE GOVLAB

Designing the 100 Questions for NYC: Panel Reflects on New Science of Questioning During Open Data Week

05 March 2020

THE DATA ASSEMBLY

PUBLIC DELIBERATION ON THE RE-USE OF DATA

The Data Assembly is an initiative from [The GovLab](#) with support from the [Henry Luce Foundation](#) to solicit diverse, actionable public input on data re-use for crisis response in the United States.



Focus on Questions also provides for more Responsible Data and AI use

Allows for a collaborative approach toward:

- Purpose specification;
- Objective prioritization.

That provides for

- A reimagined concept of data stewardship and quality;
- A social license for re-use;
- Appropriate limits on data collection, access and re-use (minimization);
- Guidance for machine learning toward beneficial outcomes;
- People-centric data collaboration and (re-)use.





DATA STEWARDS

The Data Stewards Network (DSN) connects responsible data leaders from the private and public sectors seeking new ways to create public value through cross-sector data collaboration. Watch this space for regular insights and outputs from the Network.

