THEGOVLAB

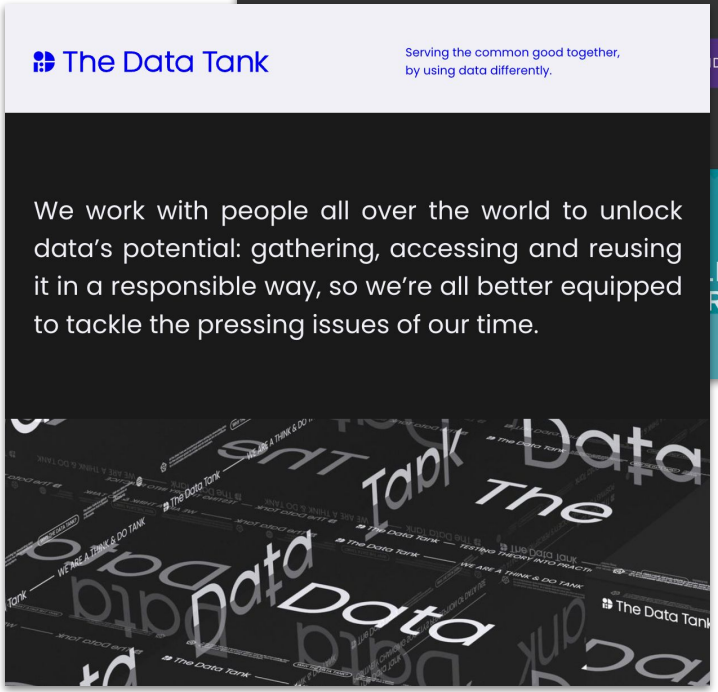
Toward a New Science of Questions for Unlocking the True Value of Data

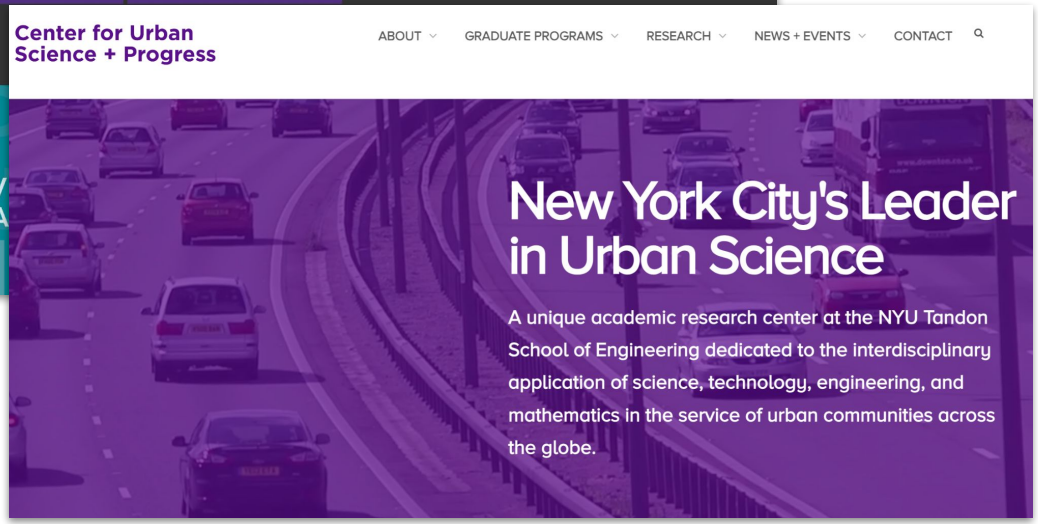
ADV Data Excellence Konferenz 2025

Stefaan G. Verhulst

April 11, 2025

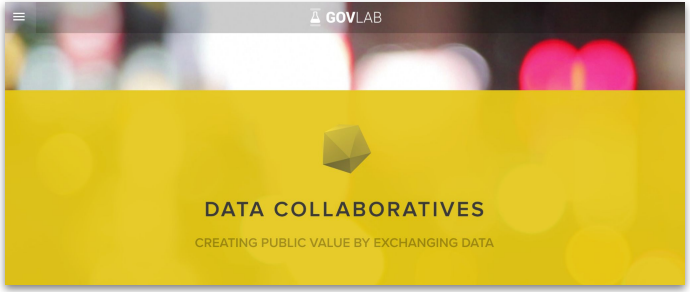

THEGOVLAB
DEEPENING OUR UNDERSTANDING OF HOW TO GOVERN
MORE EFFECTIVELY AND LEGITIMATELY THROUGH TECHNOLOGY
D OUT MORE
EXPLORE THE BLOG


The Data Tank
Serving the common good together, by using data differently.
We work with people all over the world to unlock data's potential: gathering, accessing and reusing it in a responsible way, so we're all better equipped to tackle the pressing issues of our time.

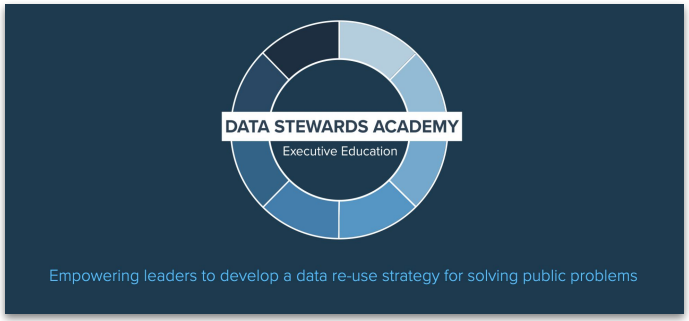

Center for Urban Science + Progress
ABOUT
GRADUATE PROGRAMS
RESEARCH
NEWS + EVENTS
CONTACT
New York City's Leader in Urban Science
A unique academic research center at the NYU Tandon School of Engineering dedicated to the interdisciplinary application of science, technology, engineering, and mathematics in the service of urban communities across the globe.

Improving how we make decisions....

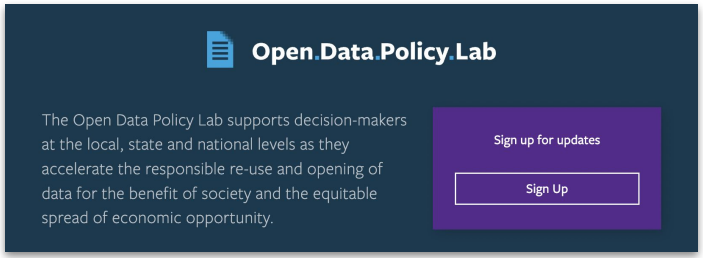
PEOPLE	*DATA*
Collective Intelligence	*Data Intelligence*
Crowd-sourcing	*Artificial Intelligence*
Citizen Science	*Behavioral Science*

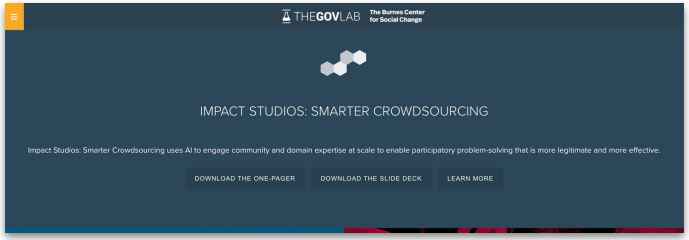

GOVLAB
DATA COLLABORATIVES
CREATING PUBLIC VALUE BY EXCHANGING DATA


DATA4Philanthropy
About
Methods & Case Studies
Data-Driven Grant Making
Updates
Join
Submit
DATA4Philanthropy
Accelerating data-driven innovation across the philanthropic sector
Learn more


DATA STEWARDS ACADEMY
Executive Education
Empowering leaders to develop a data re-use strategy for solving public problems


BIG DATA FOR MIGRATION
ABOUT
INITIATIVES
BLOG
REPOSITORY
EVENTS
JOIN
BD4M
Big Data for Migration Alliance
Accelerating the responsible and ethical use of non-traditional data sources and methods to support migration policy on the global, national, and local levels


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

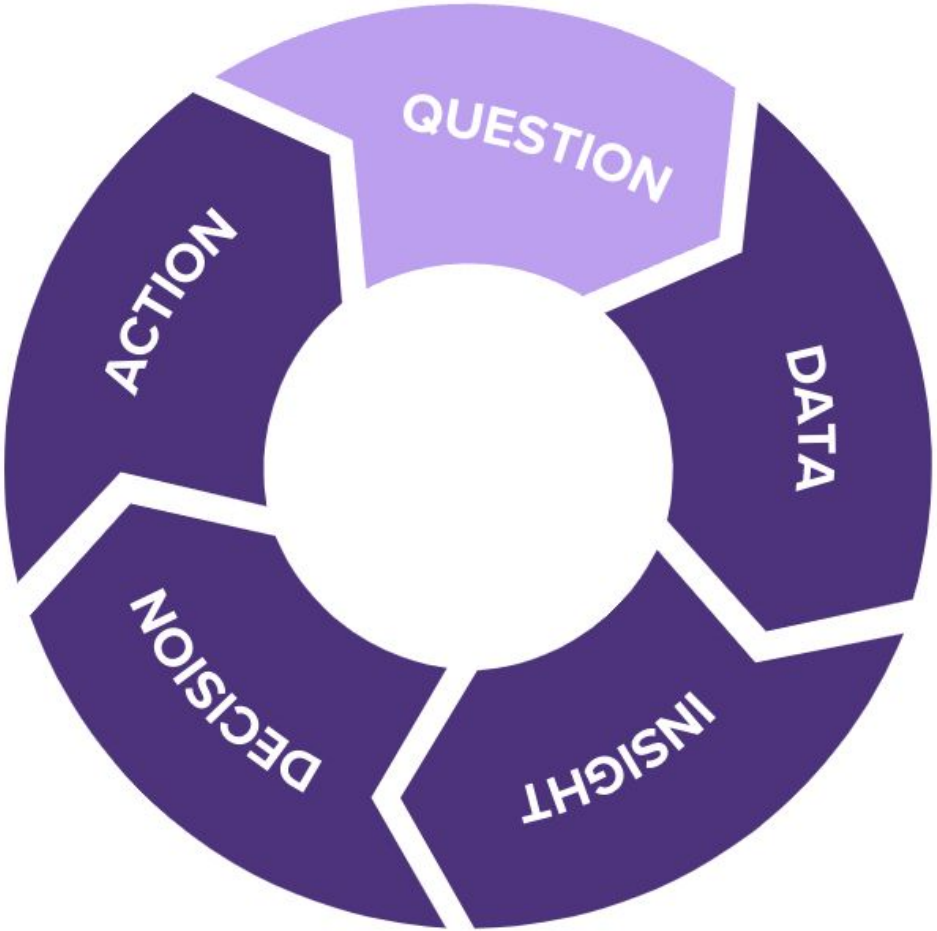

THEGOVLAB
The Burnes Center for Social Change
IMPACT STUDIOS: SMARTER CROWDSOURCING
Impact Studios: Smarter Crowdsourcing uses AI to engage community and domain expertise at scale to enable participatory problem-solving that is more legitimate and more effective.
DOWNLOAD THE ONE-PAGER
DOWNLOAD THE SLIDE DECK
LEARN MORE

Observation 1:
The value of data & AI depends directly on the questions we ask.

? > !

DX = QX

From question to action



The value of questions

> *If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.*
>
> **– Albert Einstein**

> *Judge a man by his questions rather than his answers.*
>
> **– Voltaire**

> *You can tell whether a man is clever by his answers. You can tell whether a man is wise by his questions.*
>
> **– Naguib Mahfouz**

Observation 2: Questions enable data responsibility.

DR = QX

Questions as a Device for Data Responsibility: Toward a New Science of Questions to Steer and Complement the Use of Data Science for the Public Good in a Polycentric Way

in: Aguerre, C., Campbell-Verduyn, M., & Scholte, J. A., Global Digital Data Governance: Polycentric Perspectives, Properties and Controversies. Routledge, Forthcoming

26 Pages • Posted: 10 Jul 2023

Stefaan Verhulst

New York University (NYU) ; Vrije Universiteit Brussel (VUB); The Data Tank; The GovLab

Date Written: February 28, 2023

The Data Stewardship Canvas

Designed by:	Date:	Version:

The Data Stewardship Canvas is a step by step process that maps a data steward's journey when building a data collaborative to support data re-use—whether the data steward is requesting or providing access to data. The steps of the canvas seek to create a systematic and responsible approach to effectively re-using data for positive social and economic outcomes.

1. Defining the Demand for Data

- **Framing the problem**: What societal, economic, or organizational challenge are you addressing?
- **Decision mapping:** What decisions do you seek to inform and when in the decision lifecycle is the data needed?
- **Identifying stakeholders:** Who needs to make decisions? Who will act on the data?
- **Question formulation:** What (type of) question, if answered, will inform the decision or address the problem?
- **Bringing all together**: What is your theory of change?

2. Defining the Supply of Data

- **Determining the Minimal (Viable) Data Needed**: What are the data elements that are required to answer the priority questions?
- **Data scouting and cataloguing**: What data sources or products exist that match the requirements?
- **Data wrangling and preparation**: How to prepare data to make it reusable?
- **Data audit and tagging:** Assessing and categorizing data in terms of quality, timeliness, interoperability, and ethical or legal considerations.
- **Data ops:** What are the expertise and capacity needs for this project?
- **Addressing data barriers:** Is there a role for synthetic data, proxies, and modeling?

3. Making a Value Proposition

- **Defining the Value to Society:** What societal problem does this data project address?
- **Identify Beneficiaries:** Who are the primary beneficiaries, and how will they benefit?
- **Making the Business Case for Data Holders:** What incentives do data holders have to participate?
- **Developing a Cost/Benefit estimate:** What are the direct and indirect costs of implementing this project? What are the expected short-term and long-term benefits?

4. Assessing the Risk

- **Assessing Risks Across the Data Lifecycle:** What are the risks at each stage of the data lifecycle?
- **Due Diligence of Possible Partners:** What are the risks of providing or receiving access to data for particular stakeholders?
- **Assessing the Risks of Not Having Access to Data**: What critical decisions cannot be made without this data?
- **Externalities Assessment:** What are the intended and unintended consequences of this data initiative?

5. Matching Demand & Supply: Operational Models

- **Assessing the level of Conditionality for Data Access:** What level of openness or control is appropriate for this project? What mechanisms (e.g. tiered access, data licensing, APIs) will enable controlled access?
- **Designing a Fit-for-Purpose Collaborative Model:** What is the best collaboration model for this data initiative?

6. Matching Demand & Supply: Governance

- **Defining the 4 Ps:** How do Purpose, Principles, Processes, and Practices guide governance?
- **Decision Provenance:** Who is responsible and accountable across the data life cycle?
- **Establishing a Social License:** How do we build public trust and legitimacy in data re-use? What engagement methods e.g., consultations, co-design workshops) will foster inclusivity?
- **Compliance Assessment:** What key legal frameworks apply?
- **Operationalization:** How do we translate governance principles into real-world decision-making and enforcement, such as data sharing agreements?

7. Matching Demand & Supply: Tech Infrastructure

- **Infrastructure Requirements:** What are the options and requirements regarding Data Transfer, Storage, and Access?
- **Preparing AI-ready data:** What data formats, labeling techniques, and governance are essential for AI applications?

8. Moving from Insight to Action: Decision Intelligence

- **Decision Intelligence:** How do we transform data insights into actionable policy decisions?
- **Lived Experience:** How do we design effective feedback loops?
- **Decision Legitimacy**: How do we embed trust in translation (such as simulations and visualizations) and decision intelligence systems?

9. Measuring Impact

- **KPIs, Impact and Evaluation:** How will you capture the impact and success of this project?
- **Exit Strategies:** How will you know when to end this project? What indicators signal completion, scale-up, or pivoting? How do we define data project sustainability?

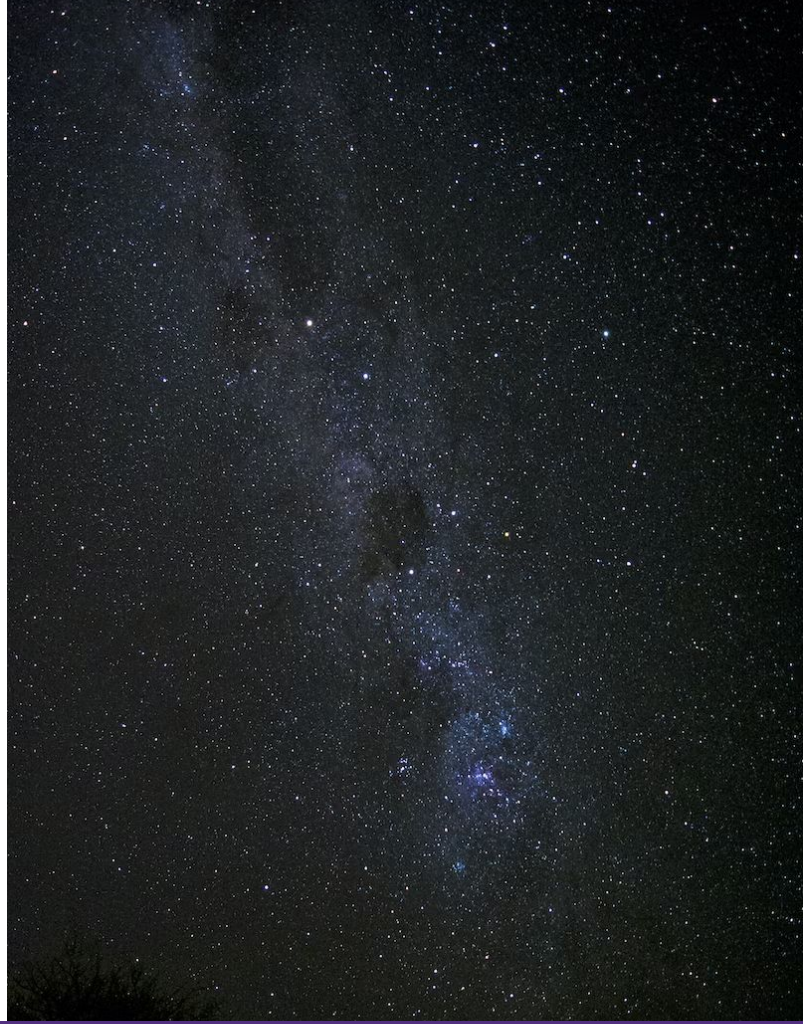
Questions as a tool for data minimization and proportionality

Questions can enable actors to:

- Determine the purpose for data collection and re-use
- Determine data retention policies
- Develop an overall data strategy

Minimum viable data point

- Your **minimum viable datapoint** is the most minimal amount of data that you need to make progress in answering your question.
- By focusing on a minimum viable datapoint, we ensure that we are using data in a **targeted and responsible manner** and **avoiding inefficiency** by only using relevant data that is proportional to the question you need to answer.



Questions as a tool for participation (democratization)

Questions can enable actors to:

- Offer a more sophisticated way for researchers, policy makers, and data holders to engage with the public
- Foster a more exclusive public debate
- Enable data users to acquire a social license for re-using data beyond initial consent

Questions as a tool for accountability

Questions can enable actors to:

- Identify and engage with key stakeholders
- Create feedback loops to fine-tune and iterate on initial versions of projects
- Create incentives for data holders to share data and participate in data sharing activities
- Enhance accountability by helping project holders anticipate and measure impact and risk

Observation 3:
We are data rich but question poor

? < !

Today's data practice

Scientific findings and answers...

Science News — *from research organizations*

NASA's Webb reveals new features in heart of Milky Way

Date: November 21, 2023

Source: NASA/

Bonobos, like humans, cooperate with unrelated members of other groups

By Jake Buehler • November 16, 2023

NOVEMBER 24, 2023 | 3 MI

Air-Conditioning Discovery Eliminates Harmful Gases

Heat pumps are ubiquitous in the form of air conditioners. Scientists just invented one that avoids harmful refrigerant gases

BY DAVIDE CASTELVECCHI & NATURE MAGAZINE

Doctors encouraged by early-stage trial of MS stem cell therapy

njecting stem cells into patients' brains found to be safe and could top further damage from the disease

More

Science News — *from research organizations*

A study unveils the link between musical preferences and our inner moral compass

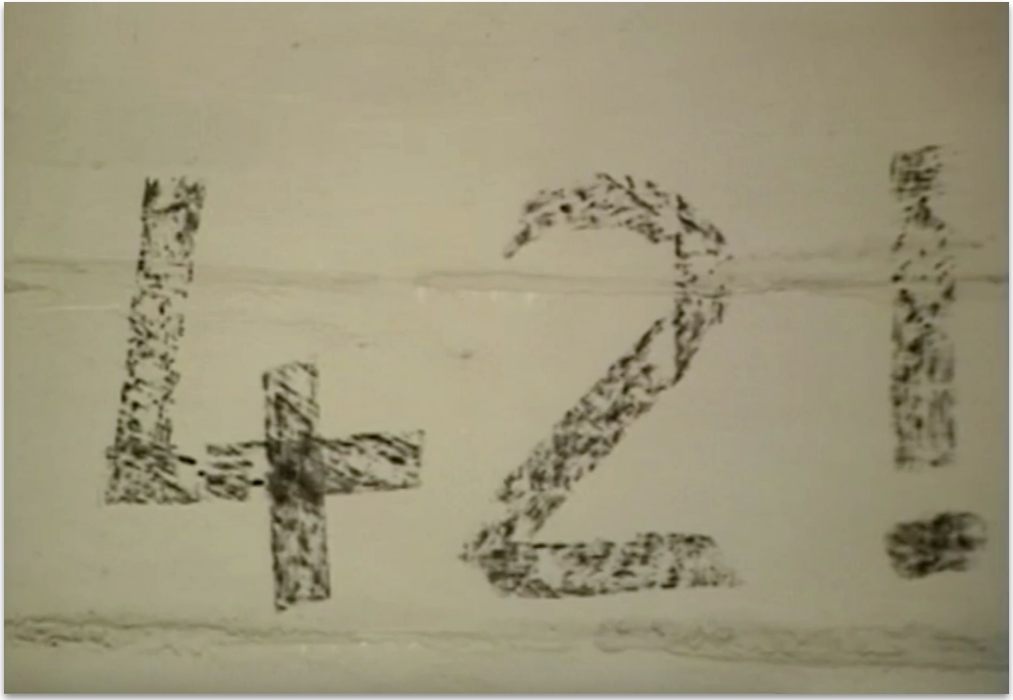
Date: November 29, 2023

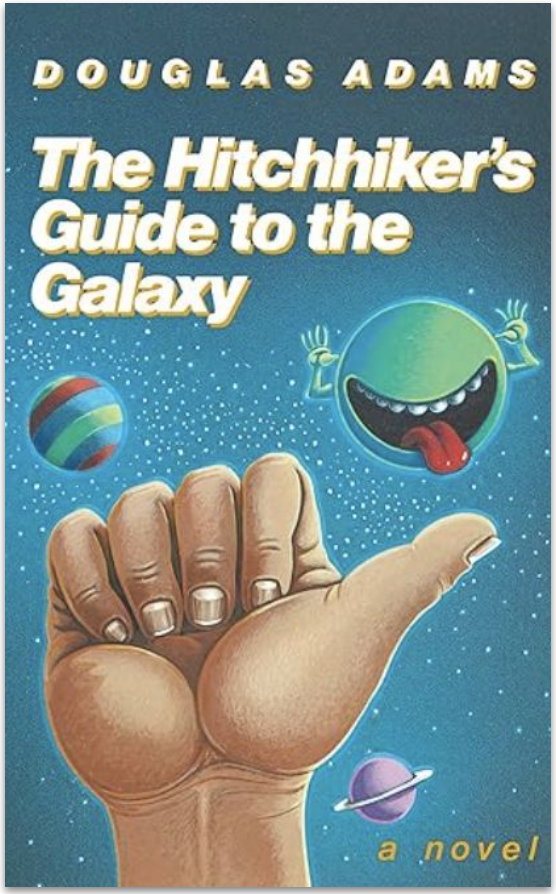
Source: Queen Mary University of London

Questions and schools

- Preschool children ask their parents an average of **100 questions a day**. By middle school, this number drops significantly, and they basically stop asking questions.
- Another study found that children ask their parents approximately **73 questions per day**.
- Some studies indicate that 4-year-olds can ask as many as **200 to 300 questions a day**, with the average being 40,000 questions between the ages of 2 and 5.

42! But what is the question





Debate: ChatGPT reminds us why good questions matter

Published: February 7, 2023 7.34pm GMT

shutterstock.

Email
Twitter 11
Facebook 12
LinkedIn

Over 100 million people used ChatGPT in January alone, according to one estimate, making it the fastest-growing consumer application in history. By producing resumes, essays, jokes and even poetry in response to prompts, the software brings into focus not just language models' arresting power, but

Author

Stefaan G. Verhulst
Co-Founder and Chief Research and Development Officer of the Governance Laboratory (GovLab), New York University

Democracy suffers a deficit in questions

Questions In-equity

Civic Engagement

The Critical Role of Questions in Building Resilient Democracies

Asking questions in new and participatory ways can complement advancements in data science and AI while enabling more inclusive and more adaptive democracies.

SHARE COMMENT PRINT ORDER REPRINTS

By Stefaan G. Verhulst, Hannah Chafetz & Alex Fischer | Oct. 22, 2024

(Illustration by iStock/AlexSecret)

Observation 4:
Questions are a critical yet poorly understood device in society

However, we are seeing an emerging field in questions

METAPHILOSOPHY

ORIGINAL ARTICLE | Open Access

Can a good philosophical contribution be made just by asking a question?[1]

Joshua Habgood-Coote, Lani Watson, Dennis Whitcomb

First published: 05 December 2022 | https://doi.org/10.1111/meta.12599

[1] The authors are listed in alphabetical order. Joshua Habgood-Coote and Dennis Whitcomb conceptualised this paper. All three authors contributed to writing it and signed off on its final version. Joshua Habgood-Coote's research on this project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 818633).

PDF TOOLS SHARE

ELSEVIER

Research Policy

Volume 51, Issue 4, May 2022, 104491

Crowdsourcing research questions in science

Susanne Beck [a b], Tiare-Maria Brasseur [a b], Marion Poetz [a b], Henry Sauermann [c]

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.respol.2022.104491

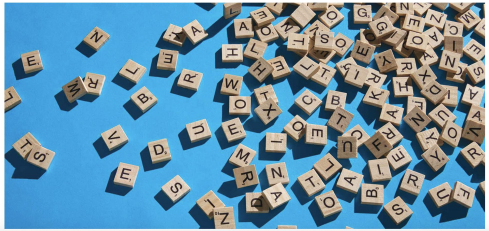
Get rights and content

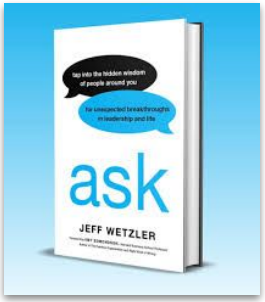
Under a Creative Commons license

open access

The Surprising Power of Questions

It goes far beyond exchanging information. by Alison Wood Brooks and Leslie K. John





RQI Right Question Institute

A Catalyst for Microdemocracy

Questions, Agency and Democracy

Dan Rothstein · Follow

5 min read · Feb 25, 2017

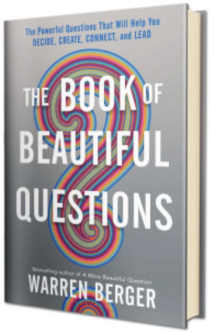
69

What do data portals do? Tracing the politics of online devices for making data public

Published online by Cambridge University Press: **30 March 2023**

Jonathan W.Y. Gray

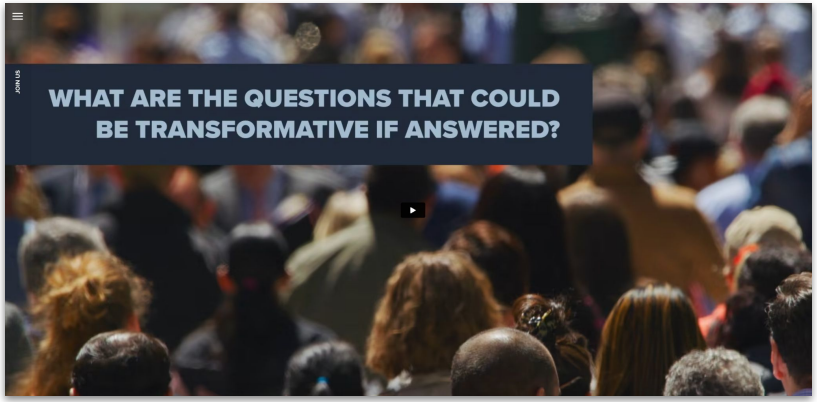
Show author details



The need for a new science of questions?

Our quest so far

- A polycentric approach to questioning
- Opens up who *asks* and *provides input* on questions
- Seeks to foster questions equity
- Harnesses crowdsourcing and participatory approaches to discover, formulate, prioritize, and implement shared questions for society



https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4494553
https://the100questions.org/

Domains

MIGRATION

GENDER

AIR QUALITY

FUTURE OF WORK

GOVERNANCE

DISINFORMATION

FOOD SYSTEMS SUSTAINABILITY

URBAN MOBILITY

Five stages of a new science of questions

Pre-questioning

Developing a topic map

Participatory questioning

Identifying and engaging with bilinguals

Using a Taxonomy of Data-Actionable Questions

Post-questioning

Clustering Questions

Prioritizing Big, Data-Actionable Questions

Public Voting

Answering

Non-traditional data sources

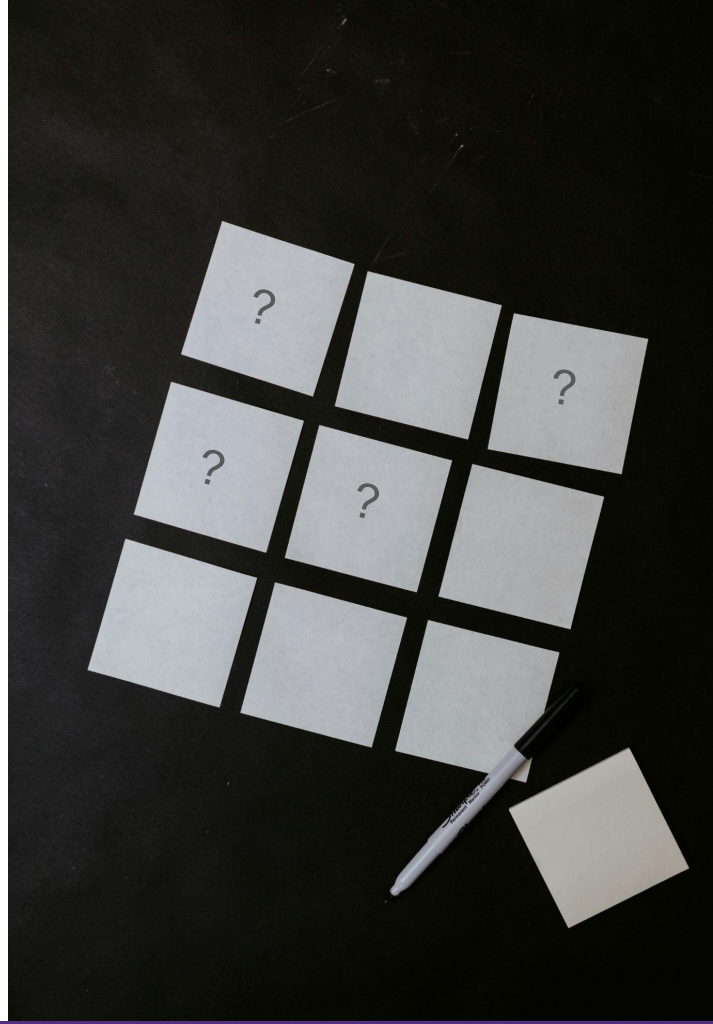
Data collaboratives

Feedback & adjustment

https://ssir.org/articles/entry/resilient-democracy-asking-better-questions

Stage 1: Pre-Questioning

- Explore and frame an area of interest
- Gain a baseline understanding of an issue area
- How? Topic Mapping, Living Evidence Reviews, Systems Thinking



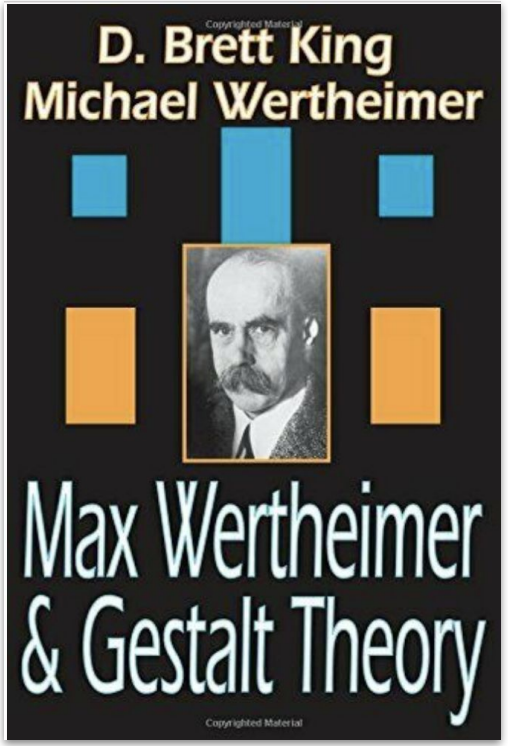
https://www.sciencedirect.com/science/article/pii/S1054139X22004955

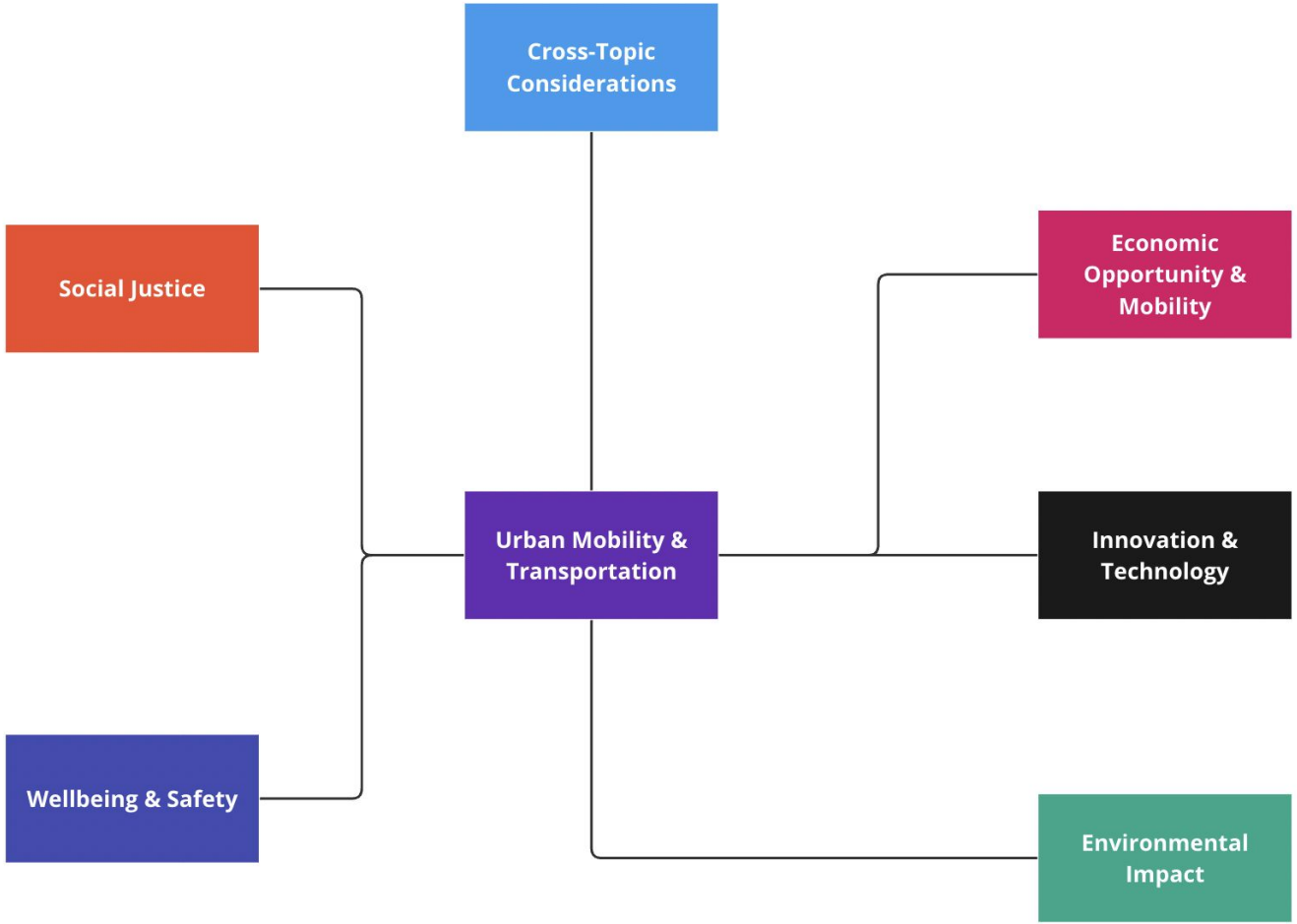
Topic Mapping

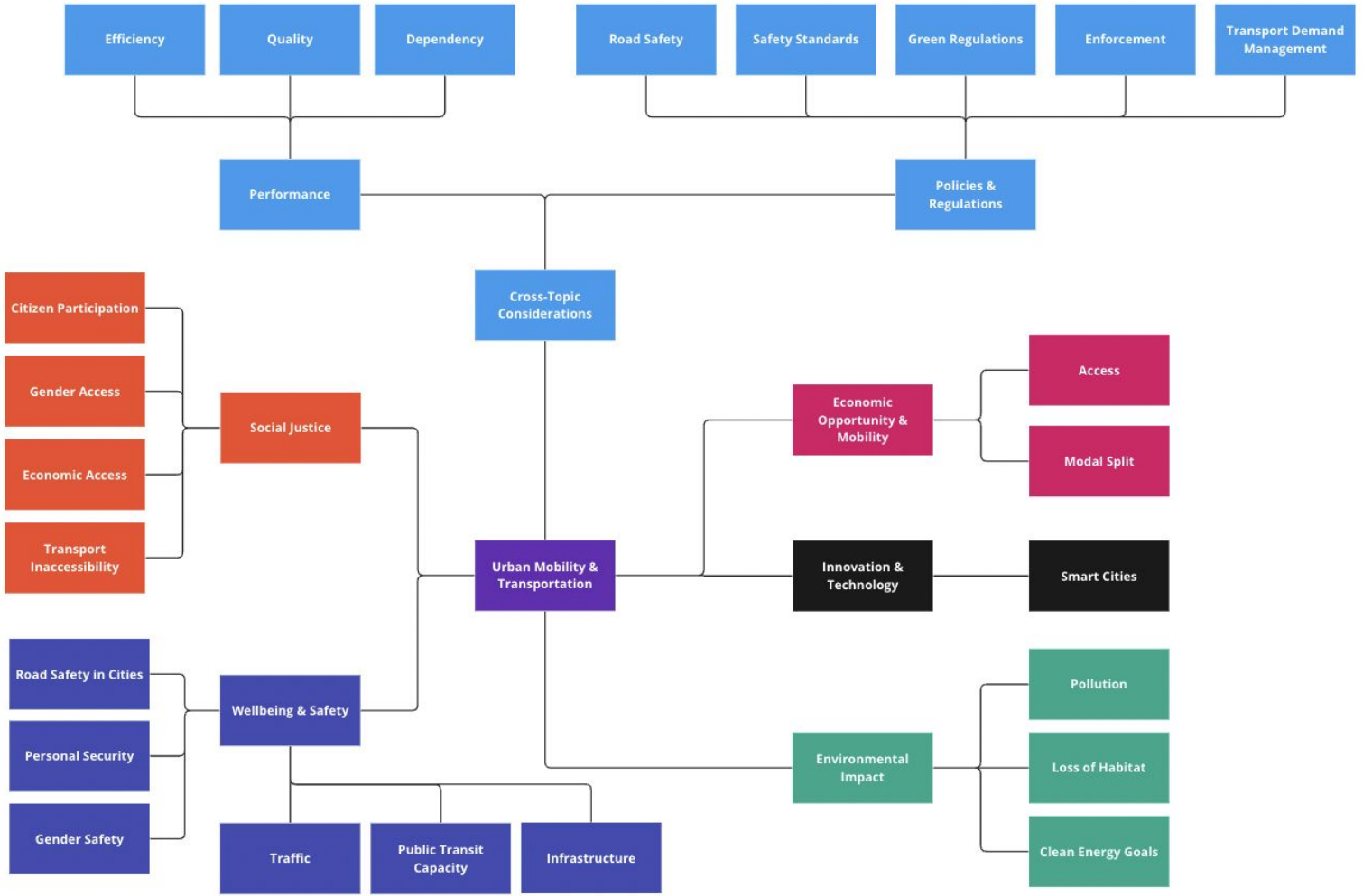
- Builds a “gestalt” of a problem field
  - ***Gestalt***: “Seeing things as a whole”

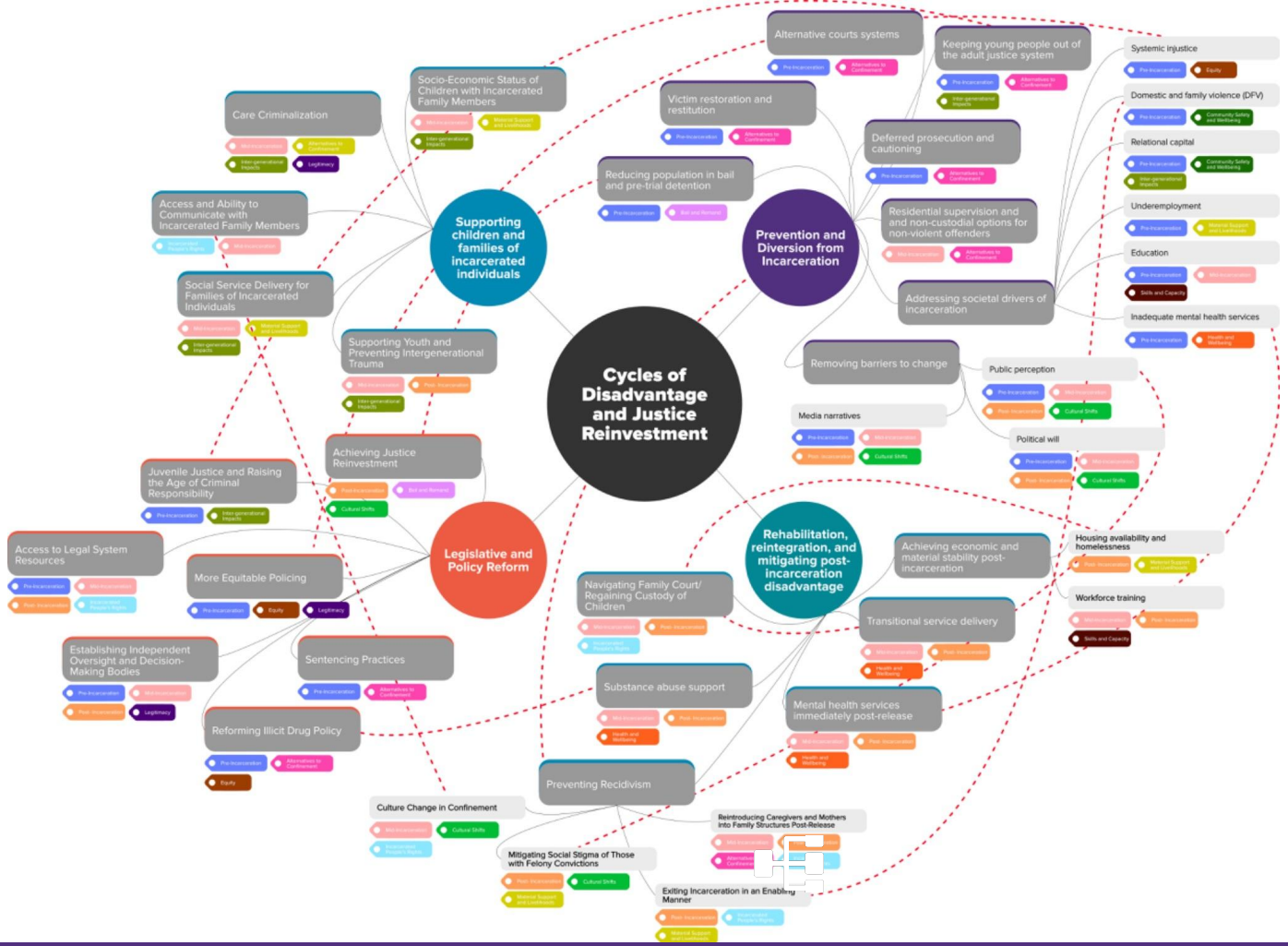
- “Grouping” or “**Laws of Organization**”
  - Proximity
  - Similarity
  - Closure
  - Simplicity

- Leveraging **R-Search Methodology** (Rapid Re-search)




Cross-Topic Considerations
Social Justice
Economic Opportunity & Mobility
Urban Mobility & Transportation
Innovation & Technology
Wellbeing & Safety
Environmental Impact


Efficiency
Quality
Dependency
Road Safety
Safety Standards
Green Regulations
Enforcement
Transport Demand Management
Performance
Policies & Regulations
Cross-Topic Considerations
Citizen Participation
Gender Access
Economic Access
Transport Inaccessibility
Social Justice
Economic Opportunity & Mobility
Access
Modal Split
Urban Mobility & Transportation
Innovation & Technology
Smart Cities
Road Safety in Cities
Personal Security
Gender Safety
Wellbeing & Safety
Traffic
Public Transit Capacity
Infrastructure
Environmental Impact
Pollution
Loss of Habitat
Clean Energy Goals



A new science of questions?

Journal of Adolescent Health xxx (2022) 1–7

ELSEVIER

JOURNAL OF ADOLESCENT HEALTH
www.jahonline.org

Original Article

Toward a Demand-Driven, Collaborative Data Agenda for Adolescent Mental Health

Stefaan Verhulst, Ph.D. [a,*], Constanza M. Vidal Bustamante [a], Liliana Carvajal, M.Sc. [b], Fiona Cece, M.Sc. [a], Jennifer Harris Requejo, Ph.D. [b], Alexandra Shaw, M.A. [a], Michelle Winowatan, M.P.A. [a], Andrew Young, M.A. [a], and Andrew J. Zahuranec, M.A. [a]

[a] *The Governance Lab, Brooklyn, New York*
[b] *UNICEF, New York, New York*

Article history: Received September 28, 2021; Accepted May 29, 2022
Keywords: Adolescents; Mental health; Data science; Crowdsourcing; Research; Agenda setting; Questions; Iinternational

A B S T R A C T

Purpose: Existing datasets and research in the field of adolescent mental health do not always meet the needs of practitioners, policymakers, and program implementers, particularly in the context of vulnerable populations. Here, we introduce a collaborative, demand-driven methodology for the development of a strategic adolescent mental health research agenda. Ultimately, this agenda aims to guide future data sharing and collection efforts that meet the most pressing data needs of key stakeholders.

Methods: We conducted a rapid literature search to summarize common themes in adolescent mental health research into a "topic map". We then hosted two virtual workshops with a range of international experts to discuss the topic map and identify shared priorities for future collabora-

IMPLICATIONS AND CONTRIBUTION

The present methodology delivers a collaborative, demand-driven, and strategic research agenda that better reflects the needs of key stakeholders in the field of adolescent mental health. Adoption of this

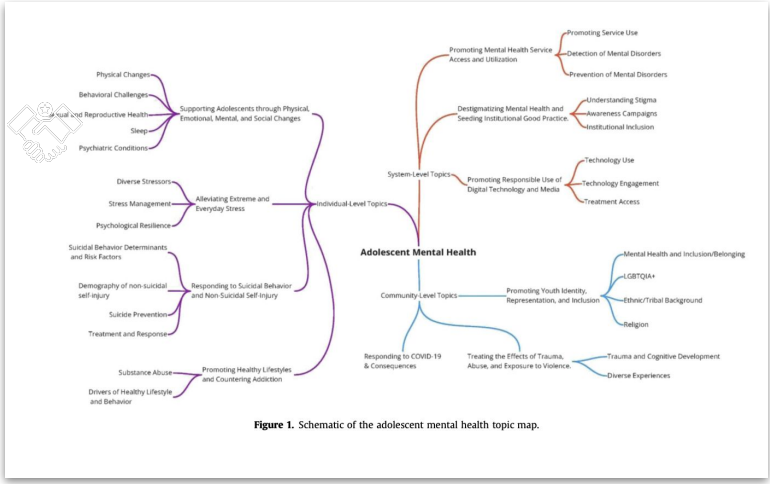
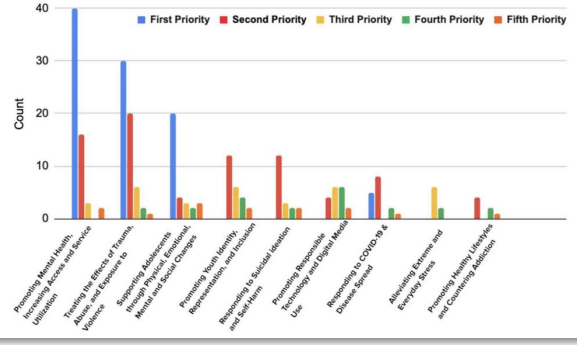


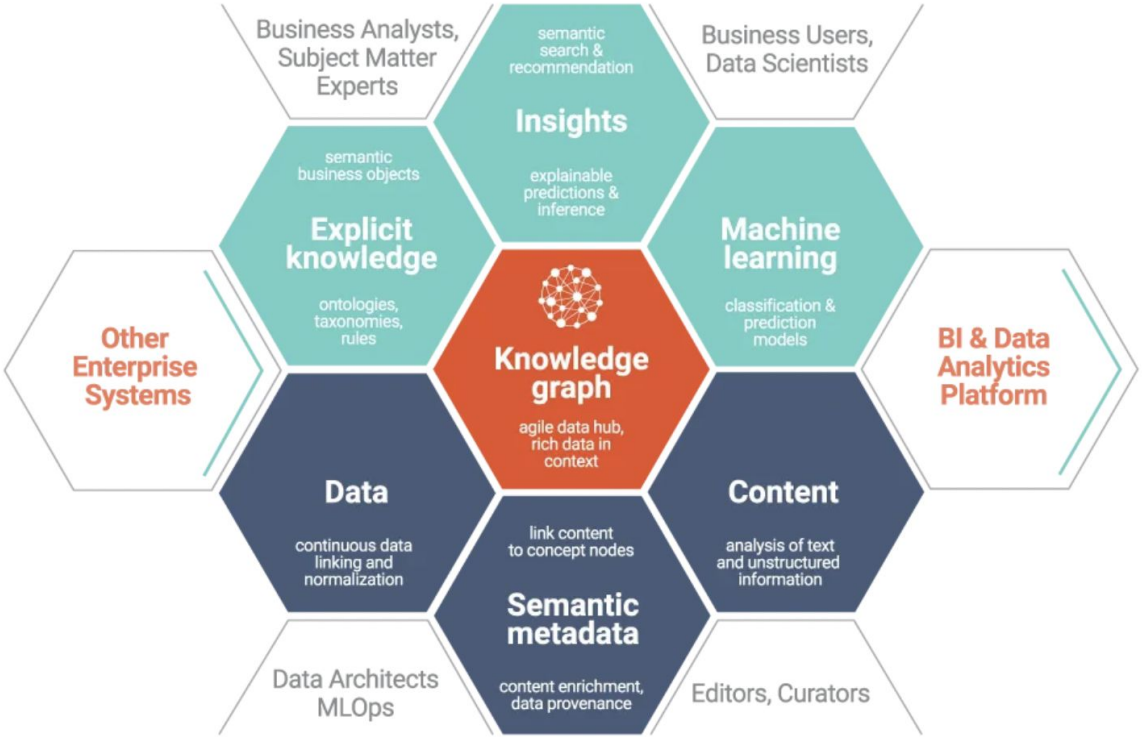
Figure 1. Schematic of the adolescent mental health topic map.



Topic Mapping allows for

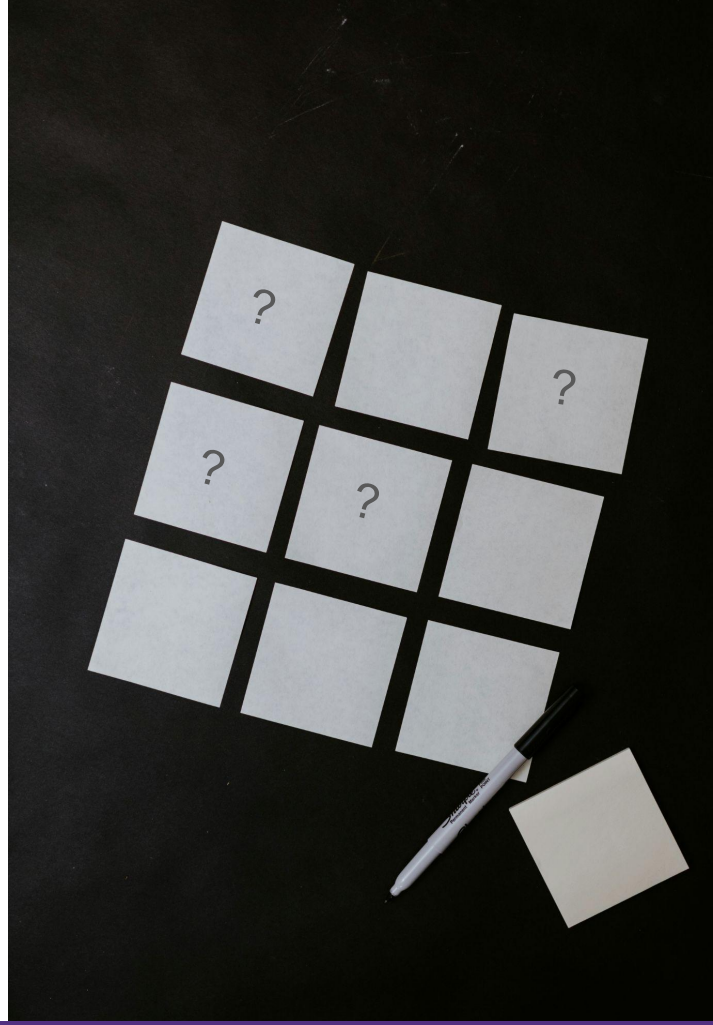
- DEVELOP “ACTOR MAPS”
- ENVIRONMENTAL SCANNING
- PUBLIC, INCLUSIVE & INTERACTIVE ENGAGEMENT
- ASSESSING THE STATE OF THE FIELD & KPI’S
- MORE RESPONSIBLE DATA & AI USE

Knowledge graphs?



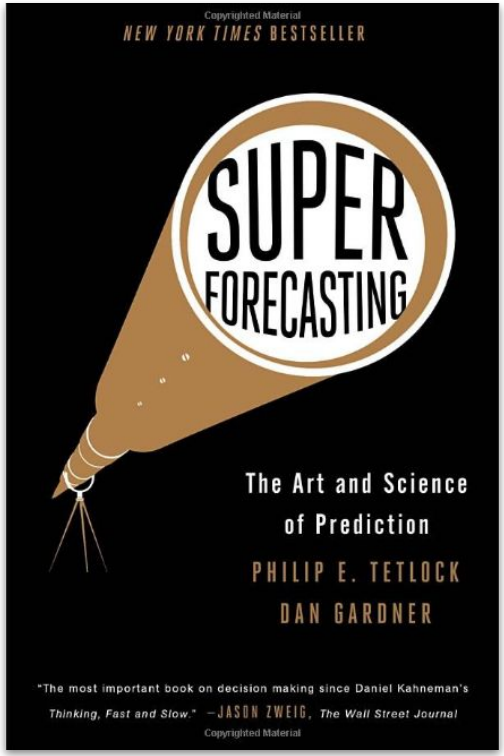
Stage 2: Participatory Questioning

- Aims to source and formulate questions
- Involves individuals with lived and subject expertise
- How? Engaging cohorts of “bilinguals,” taxonomy of data-actionable questions, question formulation techniques.



https://ssir.org/articles/entry/resilient-democracy-asking-better-questions

Bilinguals – Super Questioners



The Mind of a Superforecaster

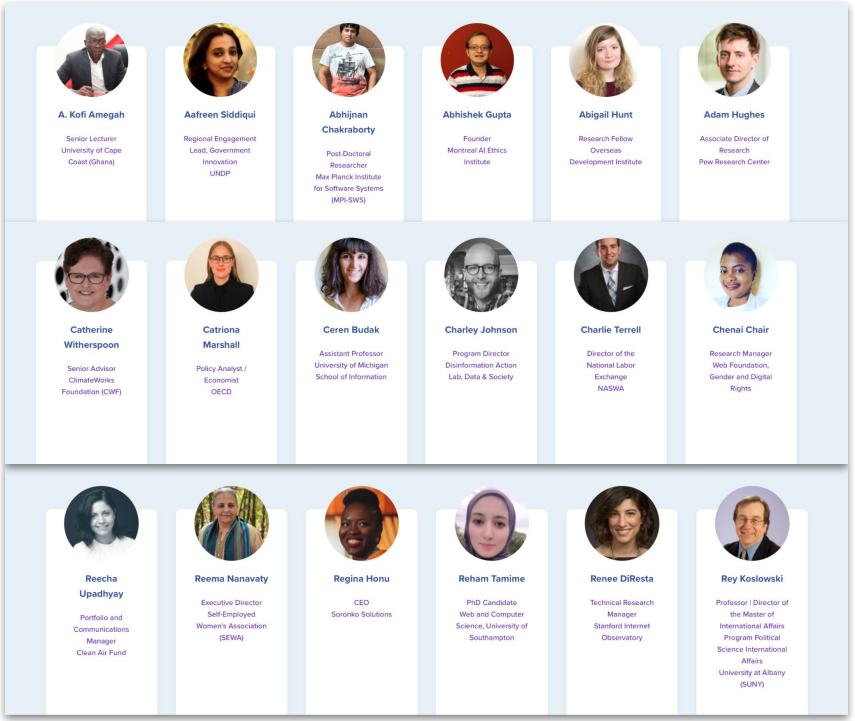
The Good Judgment Project identified the best of its volunteers as "Superforecasters" because of their consistently accurate predictions. Superforecasters differ but tend to possess the following traits:

- High intelligence—but not necessarily off the board
- Broad domain knowledge, especially of politics
- High scores on a test of actively open-minded thinking
- Willingness to seek and consider information contrary to their previous point of view
- Tendency to enjoy thinking and forecasting
- Belief that forecasting skill can be cultivated and is not just innate ability or blind luck
- Scientific worldview
- Not much faith in fate or luck

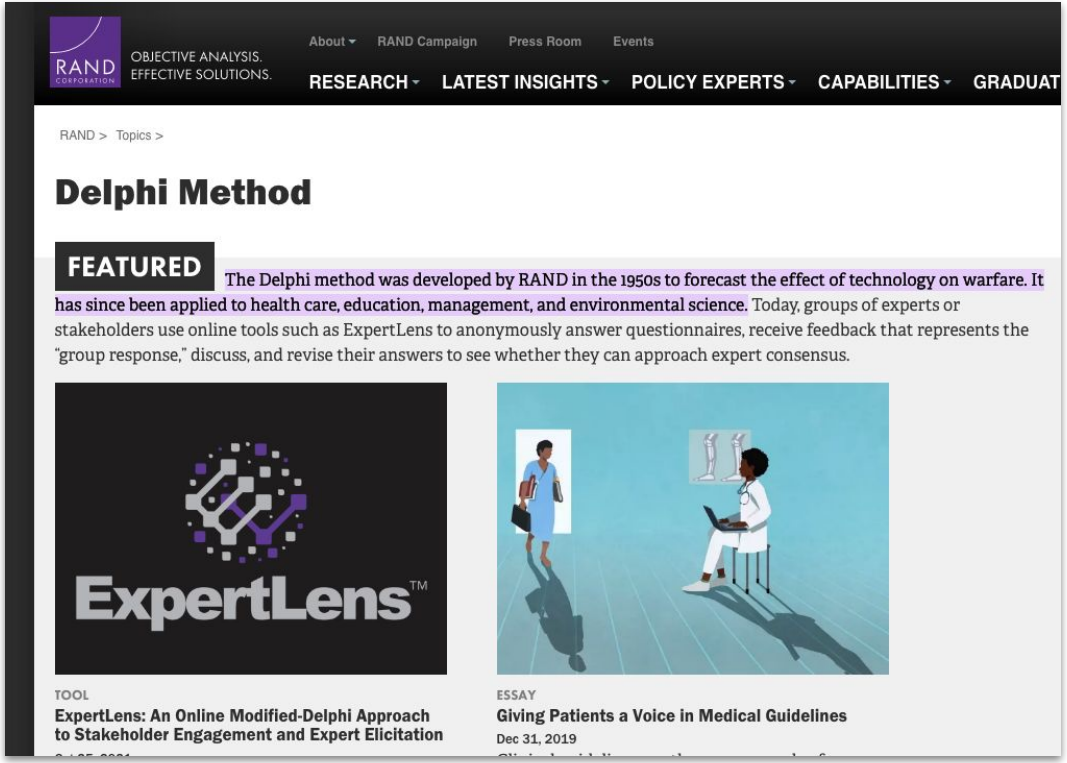
Identifying and Engaging with Bilinguals

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.

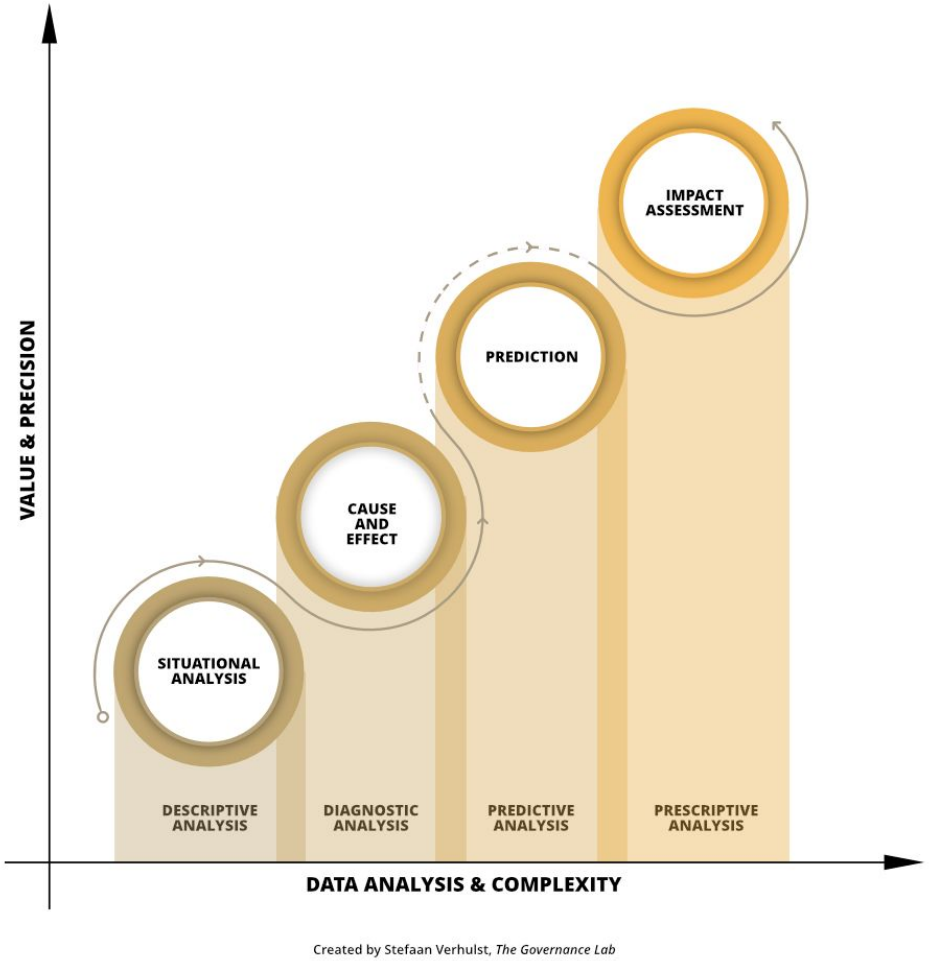


Utilizing a “modified” delphi method



Using a taxonomy of data-actionable questions

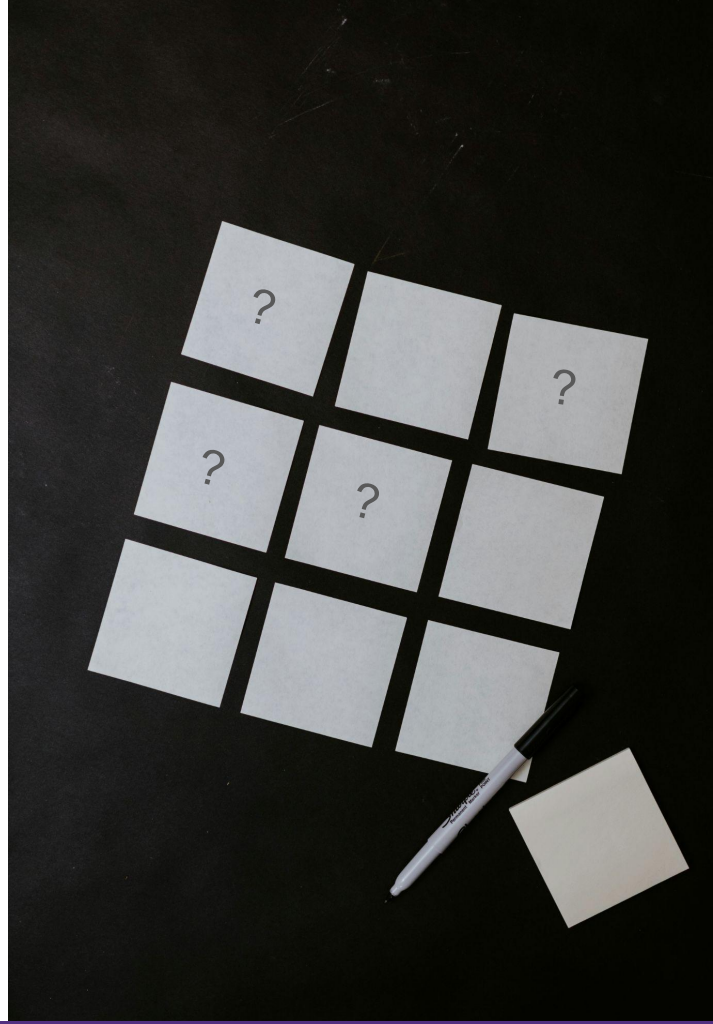
Hierarchy of questions



Created by Stefaan Verhulst, *The Governance Lab*

Stage 3: Post-Questioning

- Prioritize the questions that matter most
- How? Prioritizing, clustering data actionable questions, public voting.



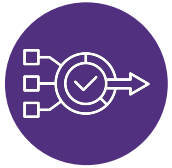
Clustering of Questions That Matter

What makes a good (data) question?

Practical and/or
Scientific Impact

Quality

Novelty



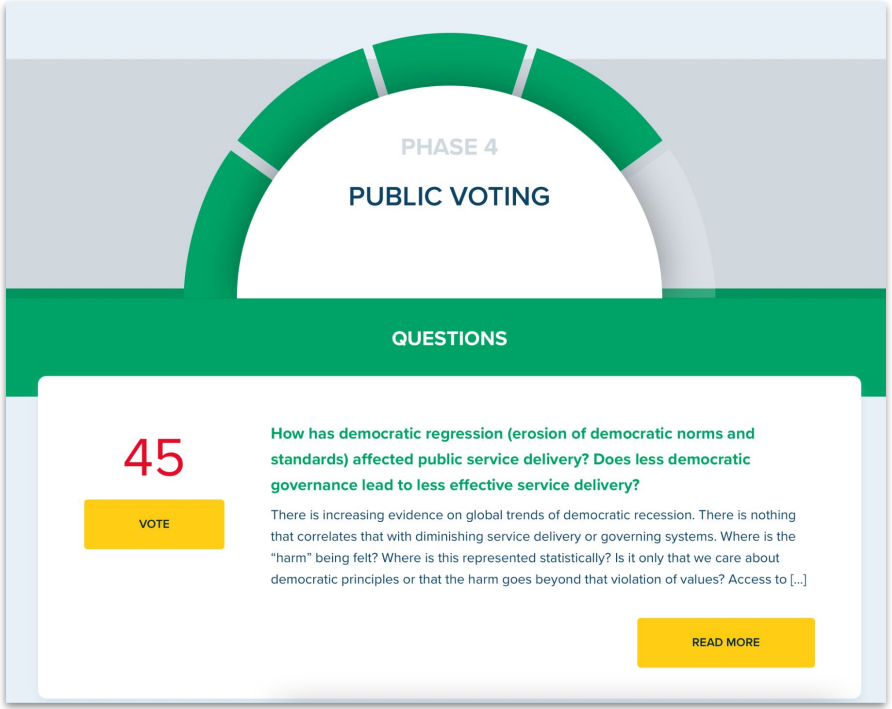
Feasibility and
Actionability

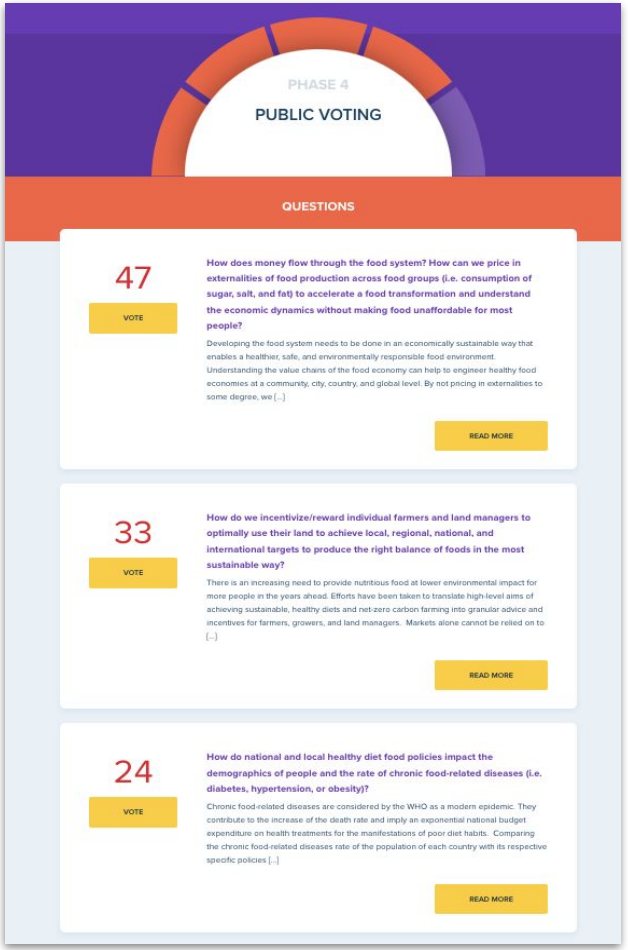
QUEST-ioning: Architecting the Inquiry

One question always leads to another question. Some things are better to wonder about.

— Christopher Pike —

Public Voting





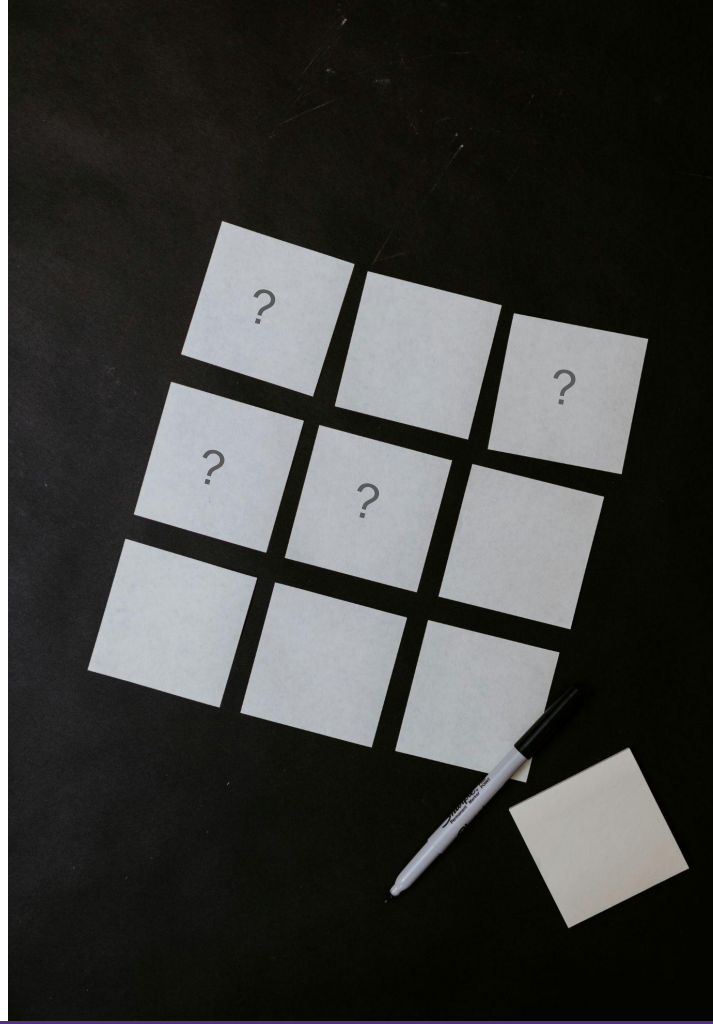
Stage 4: Identifying data

data not found

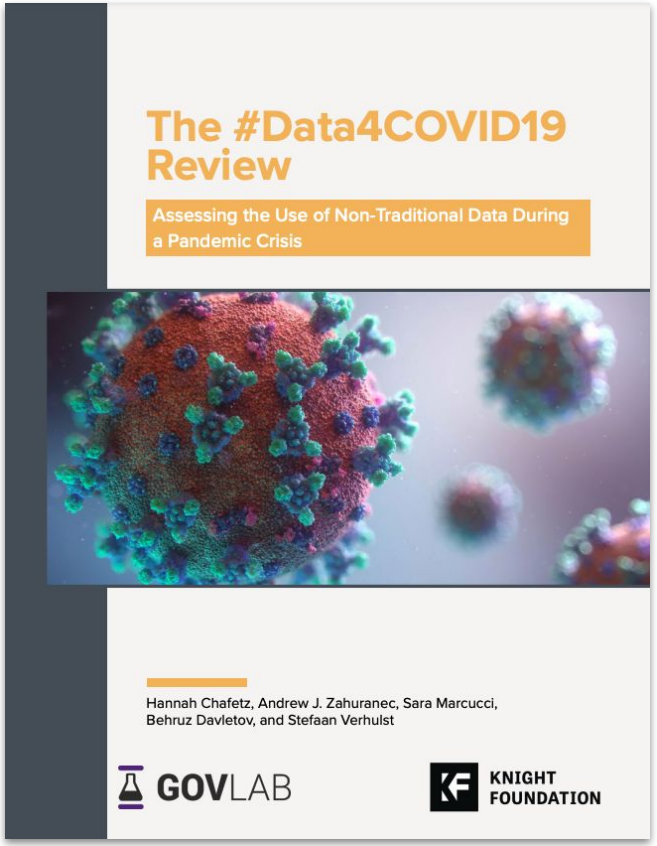
data not found is a dataset of datasets that were sought but not found on data portals around the world (read more).

Search:

data not found	date	portal	country code	url
accidents resulting in chemical release in the North Sea	2020-06-22	data.overheid.nl	NL	https://data.overheid.nl/en/community/datarequest/accidents-resulting-in-chemical-release-north-sea-ongevallen-met-chemische-lozing-noordzee
accidents with injuries as a result of an activity with a horse in the Netherlands	2021-08-12	data.overheid.nl	NL	https://data.overheid.nl/en/community/datarequest/aantal-ongevallen-met-letsel-in-nederland-als-gevolg-van-een-activiteit-met-een-paard-zoals-paardrij
actors and acting talents in Netherlands	2018-06-21	data.overheid.nl	NL	https://data.overheid.nl/en/community/datarequest/data-van-acteurs-en-acteertalenten-in-nl
acts of ecocide	2013-08-09	open.canada.ca	CA	https://search.open.canada.ca/en/sd/id/61945cc5-70b4-4c63-8c72-dfbc921000dc
addresses and times of ambulance visits	2018-09-05	data.overheid.nl	NL	https://data.overheid.nl/en/community/datarequest/ambulancebezoek
addresses where a firearm is available	2018-07-18	data.overheid.nl	NL	https://data.overheid.nl/en/community/datarequest/adressen-waar-een-vuurwapen-voorhanden-is
all deceased men 60 plus in January 2021 in the Netherlands	2022-01-31	data.overheid.nl	NL	https://data.overheid.nl/en/community/datarequest/alle-overleden-mannen-60-plus-in-januari-2021-in-nederland



Unlocking new data initiatives

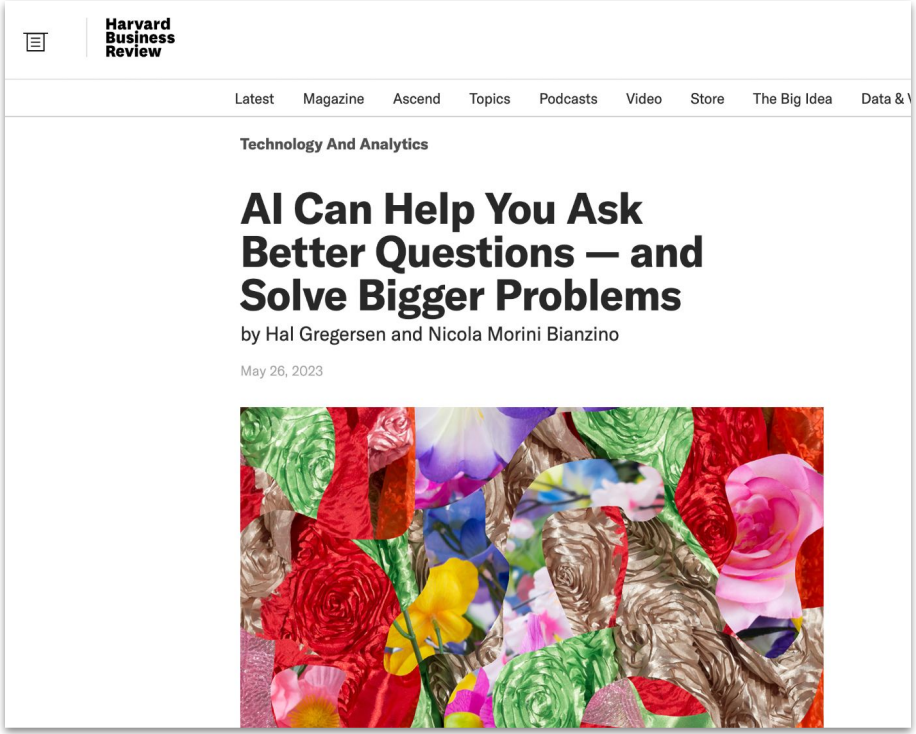


https://review.data4covid19.org/
https://datacollaboratives.org/

Our quest so far

The use of AI to formulate questions?

AI and questions?



Harvard Business Review

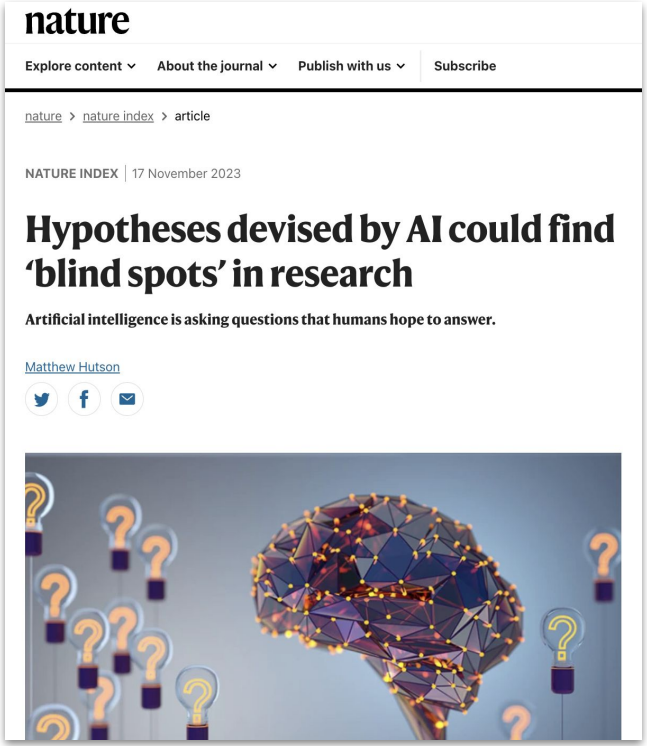
Latest Magazine Ascend Topics Podcasts Video Store The Big Idea Data &

Technology And Analytics

AI Can Help You Ask Better Questions — and Solve Bigger Problems

by Hal Gregersen and Nicola Morini Bianzino

May 26, 2023



nature

Explore content About the journal Publish with us Subscribe

nature > nature index > article

NATURE INDEX | 17 November 2023

Hypotheses devised by AI could find 'blind spots' in research

Artificial intelligence is asking questions that humans hope to answer.

Matthew Hutson

Takeaways

Invest in Questions Literacy

Beyond Answers Presented by AI: Unlocking Innovation and Problem Solving Through A New Science of Questions

10 Pages • Posted: 19 Mar 2025

Stefaan Verhulst
New York University (NYU) ; Vrije Universiteit Brussel (VUB); The Data Tank; The GovLab

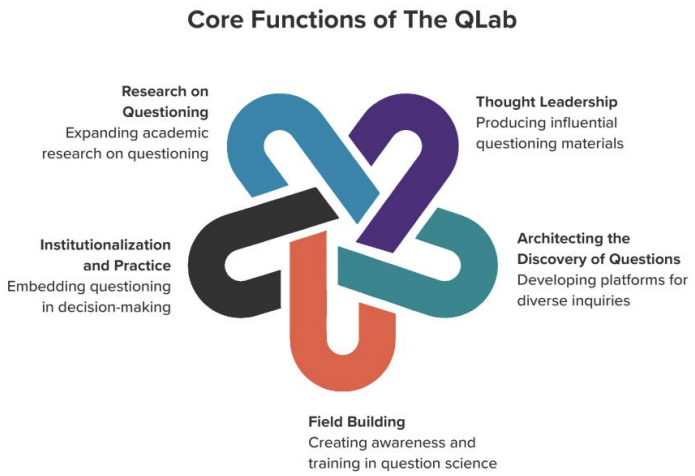
Hannah Chafetz
New York University (NYU) - The GovLab

Date Written: February 28, 2025

Abstract

Today's global crises–from climate change to inequality–have demonstrated the need for a broader conceptual transformation in how to approach societal issues. Focusing on the questions can transform our understanding of today's problems and unlock new discoveries and innovations that make a meaningful difference. Yet, how decision-makers go about asking questions remains an underexplored topic.

Much of our recent work has focused on advancing a new science of questions that uses participatory approaches to define and prioritize the questions that matter most. As part of this work, we convened an Interdisciplinary Committee on Establishing and Democratizing the Science of Questions to discuss why



Define High Value Questions

PRESS RELEASE | Publication 20 January 2023

Commission defines high-value datasets to be made available for re-use

Today, the Commission has published a list of high-value datasets that public sector bodies will have to make available for re-use, free of charge, within 16 months.

Certain public sector data, such as meteorological or air quality data are particularly interesting for creators of value-added services and applications and have **important benefits** for society, the environment and the economy – which is why they should be made available to the public.

Margrethe **Vestager,** Executive Vice-President for a Europe Fit for the Digital Age, said:

> Making high-value datasets available will benefit both the economy and society, for example by helping to combat climate change, reducing urban air pollution and improving transport infrastructure. This is a practical step towards achieving a successful Digital Decade and building a more prosperous digital future.

iStock photo Getty images plus

Related topics

Big data | Data policy | Data value chain | Open data

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?

Develop Learning agendas

Learning Agendas, or evidence-building plans, are systematic plans for identifying and addressing priority questions relevant to the programs, policies, and regulations of an agency. They identify, prioritize, and establish strategies to develop evidence to answer important short- and long-term questions including:

- **Strategic questions** about how the agency meets its mission(s), including about how programs, policies, and regulations function, and
- **Operational questions** about the agency's operations like human resources, grant-making procedures, financial systems and tracking, and internal processes.

Areas of research interest

Areas of Research Interest (ARIs) are specific topics or issues that the government is interested in.

They are a way for government organisations to indicate that they're keen to hear research evidence or insights related to that area.

ARIs are wide-ranging and can cover anything from economic and social policy to international relations and national security. By identifying ARIs, the civil service can ensure that research efforts are targeted where evidence is most needed to inform policy decisions and improve government performance.

Celebrate Questions Excellence


THEGOVLAB