THEGOVLAB

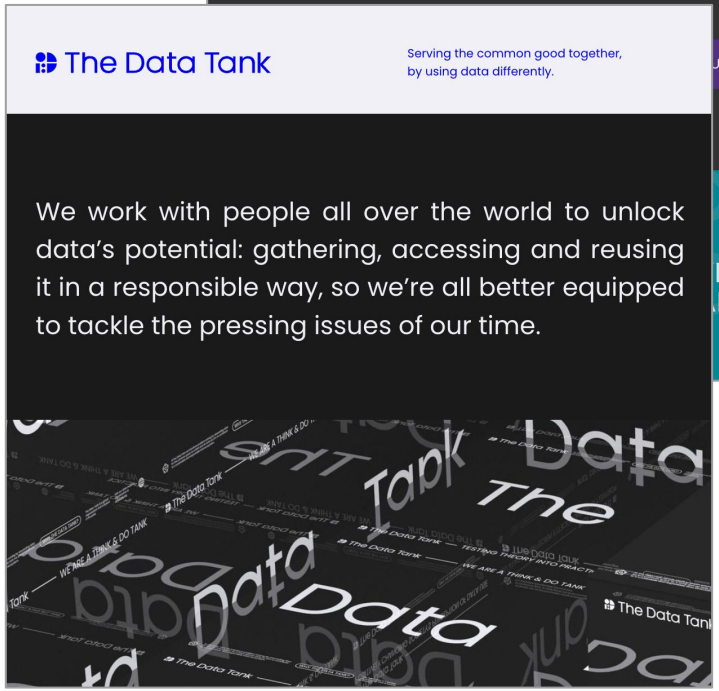
Toward a New Era & Paradigm of Data Stewardship

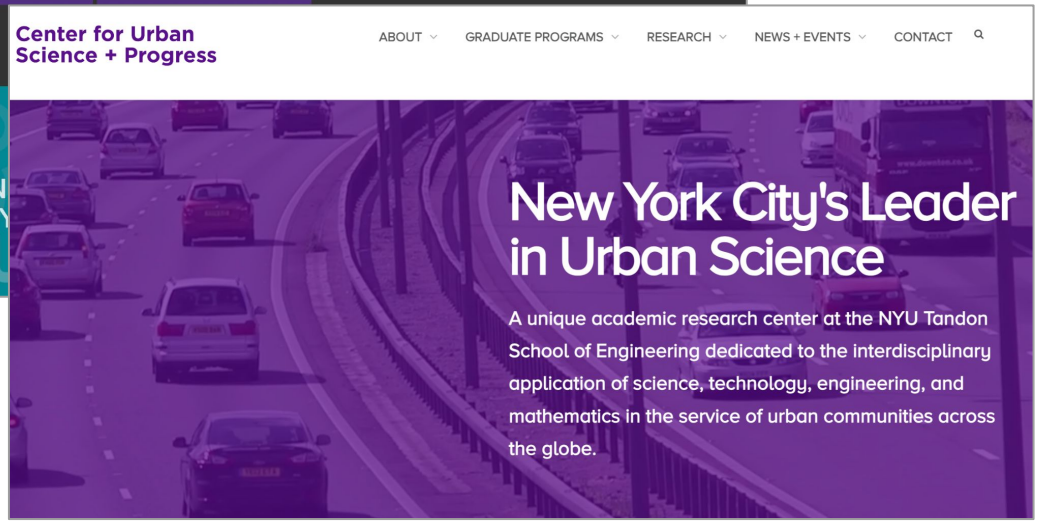
MET2023

Dr. Stefaan Verhulst

July 4, 2023

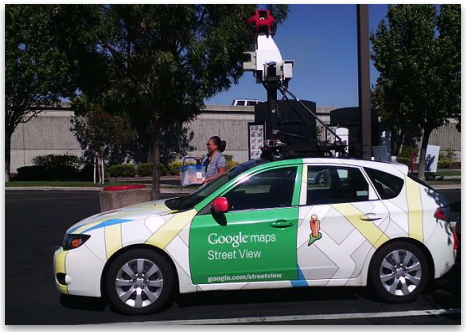
THEGOVLAB
DEEPENING OUR UNDERSTANDING OF HOW TO GOVERN
MORE EFFECTIVELY AND LEGITIMATELY THROUGH TECHNOLOGY
UT MORE
EXPLORE THE BLOG


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.

Introduction

TODAY'S PUBLIC PROBLEMS REQUIRE INNOVATION



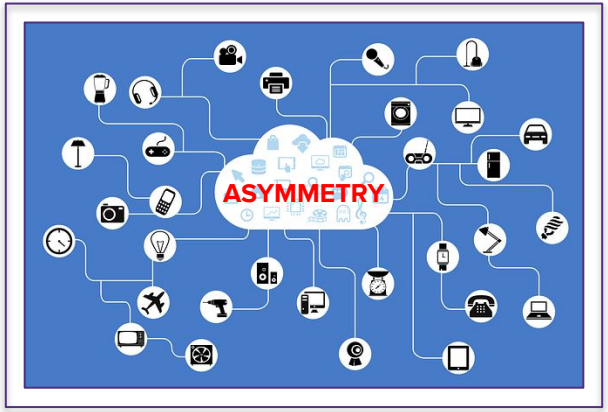
DATAFICATION: THE MIS-USE & RE-USE OF DATA

Volume

MIS-USE

Veracity

DEMAND



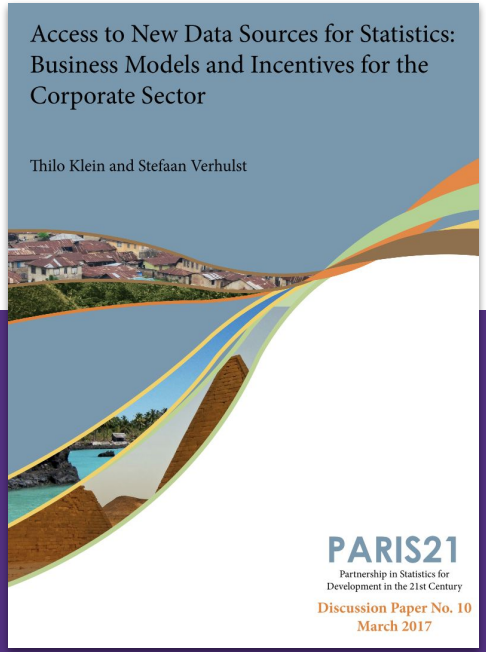
SUPPLY

Velocity

RE-USE

Variety

NEW DATA SOURCES FOR STATISTICS



https://bit.ly/SVParis21

bit.ly/3PxtPZh

B2G DATA SHARING

bit.ly/3i5eEW2

Home > Press corner > Data Act: political agreement

Available languages: English

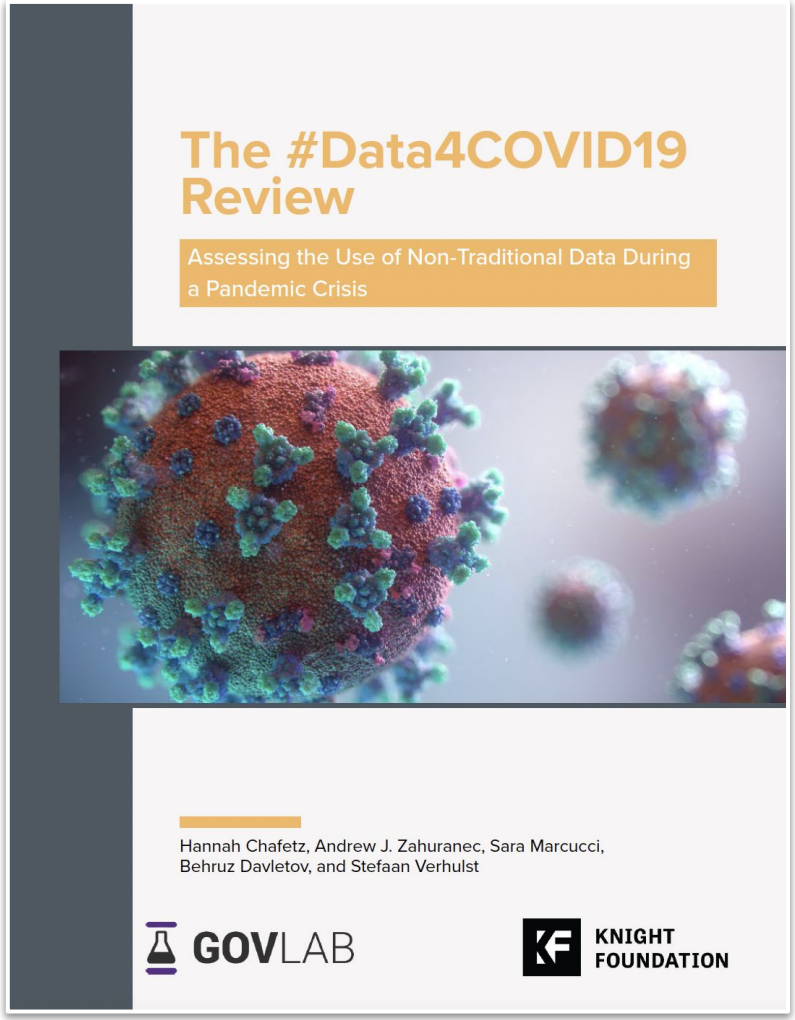
Press release | 28 June 2023 | Brussels

Data Act: Commission welcomes political agreement on rules for a fair and innovative data economy

Concretely, the Data Act includes:

- **Measures that enable users of connected devices to access the data generated by these devices and by services related to these devices**. Users will be able to share such data with third parties, boosting aftermarket services and innovation. Simultaneously, manufacturers remain incentivised to invest in high-quality data generation while their trade secrets remain protected.
- **Measures to provide protection from unfair contractual terms that are unilaterally imposed.** These aim to safeguard EU companies from unjust agreements, fostering fair negotiations and enabling SMEs to participate more confidently in the digital marketplace.
- **Mechanisms for public sector bodies to access and use data held by the private sector** in cases of public emergencies such as floods and wildfires, or when implementing a legal mandate where the required data is not readily available through other means.

bit.ly/3JyM4tq



DATA FOR COVID-19

Long Lead Times

Frequently, there are long timelines from data collection to dissemination due to bureaucracy and resource costs.

Scale

Traditional data may not exist at scale in certain low-resource contexts, leading to an inaccurate evidence base.

Quality Challenges

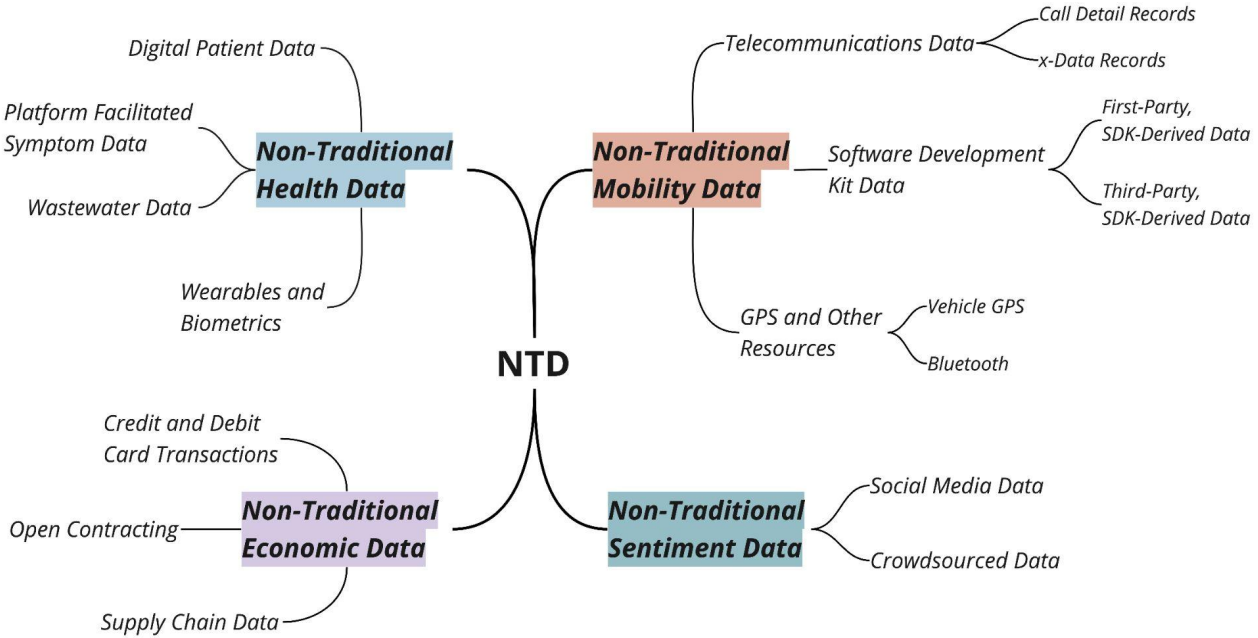
Traditional data efforts -such as surveys- are often prone to bias and error (in part because of decreasing response rates).

review.data4covid19.org

DATA FOR COVID-19: TRADITIONAL & NON-TRADITIONAL DATA

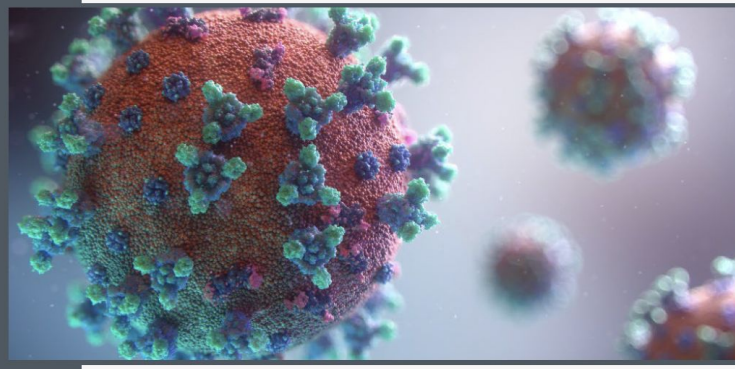
	Traditional Data	NTD
What it includes	Data collected about a topic	Proxies for the data we intend to measure
Where it is located	Analog + digital systems	Digital systems
Data Type	“Small”, Structured	“Big”, Often Unstructured
Why it is generated	Data collected for a specific purpose	Data that has been repurposed for a new use case
Examples	*Surveys, data collected by National Statistics Offices*	*Social media data, crowdsourced data, telecoms data*

Types of NTD used during COVID-19



The #Data4COVID19 Review

Assessing the Use of Non-Traditional Data During a Pandemic Crisis

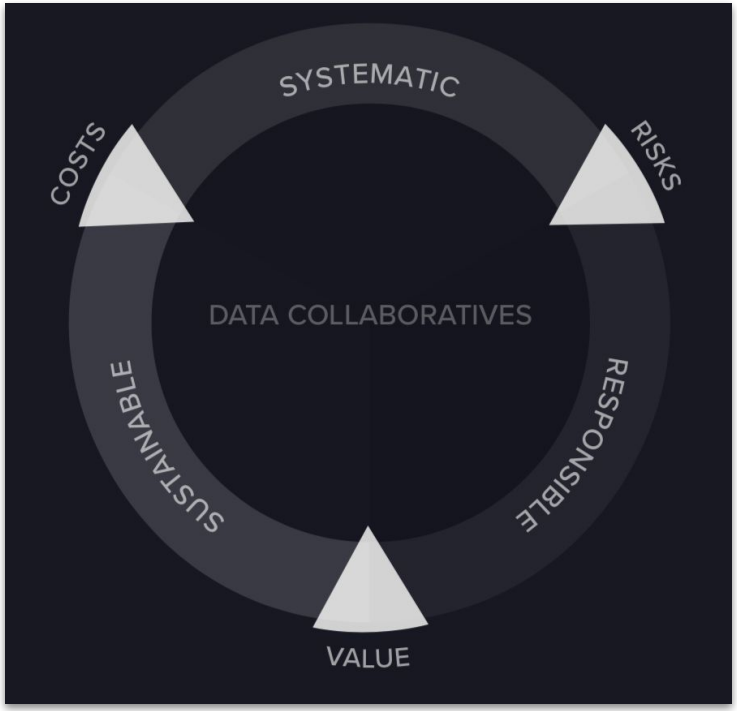


Hannah Chafetz, Andrew J. Zahuranec, Sara Marcucci, Behruz Davletov, and Stefaan Verhulst

GOVLAB

KNIGHT FOUNDATION

DATA FOR COVID-19



review.data4covid19.org

REQUIRES A NEW ERA & PARADIGM OF DATA STEWARDSHIP:

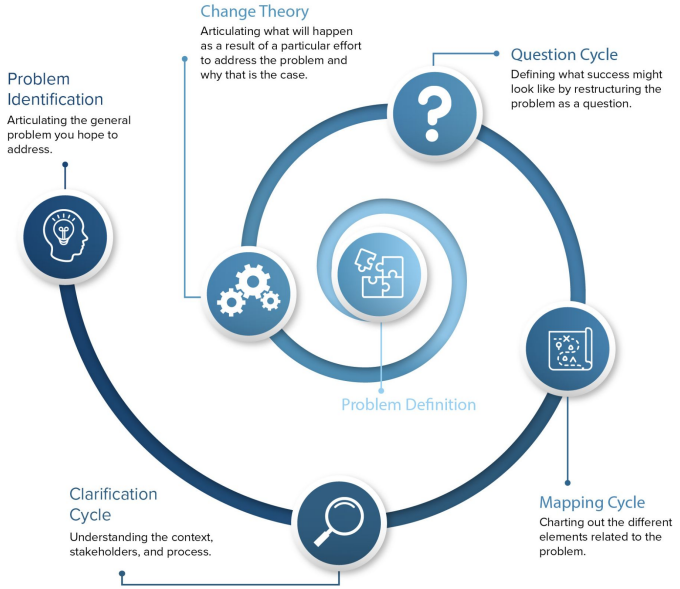
Wanted: Data Stewards — Drafting the Job Specs for A Re-imagined Data Stewardship Role

Stefaan G. Verhulst · Follow
Published in Data Stewards Network · 5 min read · Mar 13

10 Steps Toward a New Era & Paradigm of Data Stewardship

1. STRENGTHEN DEMAND FOR (Non Traditional) DATA

the100questions.org



bit.ly/ODPLProbTool

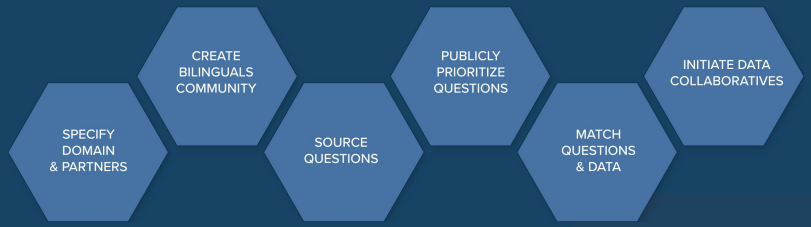
IMPORTANCE OF THE 100 QUESTIONS

What are we trying to accomplish?

THE 100 QUESTIONS

WHAT IS THE 100 QUESTIONS INITIATIVE?

The 100 Questions Initiative seeks to map the world's 100 most pressing, high-impact questions that could be answered if relevant datasets were leveraged in a responsible manner.

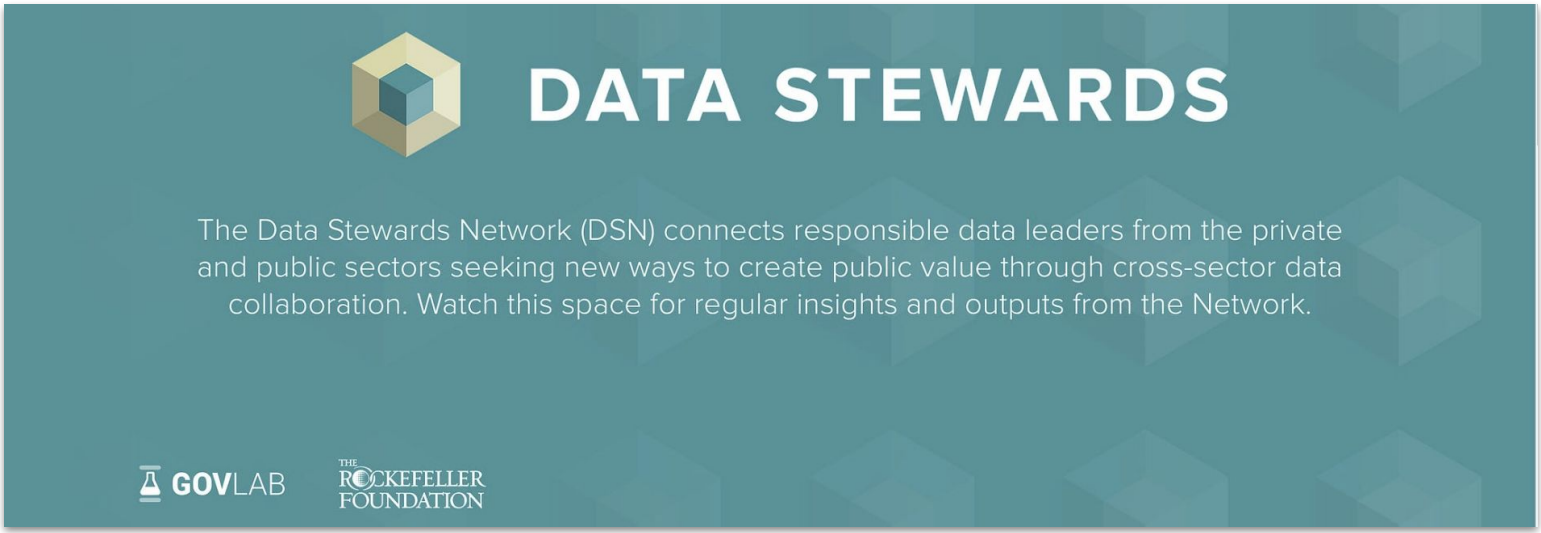


WHY DO WE NEED 100 QUESTIONS?

The complexity of challenges facing our world today -- from climate change to the reinvention of work to economic inequality to mass migration -- is unprecedented. At the same time, our policymaking and problem-solving approaches are insufficient. We need not only innovative solutions but also innovation in how we develop solutions.

Data has been widely recognized as a possible source of innovation, and data collaboratives offer a particularly promising avenue.

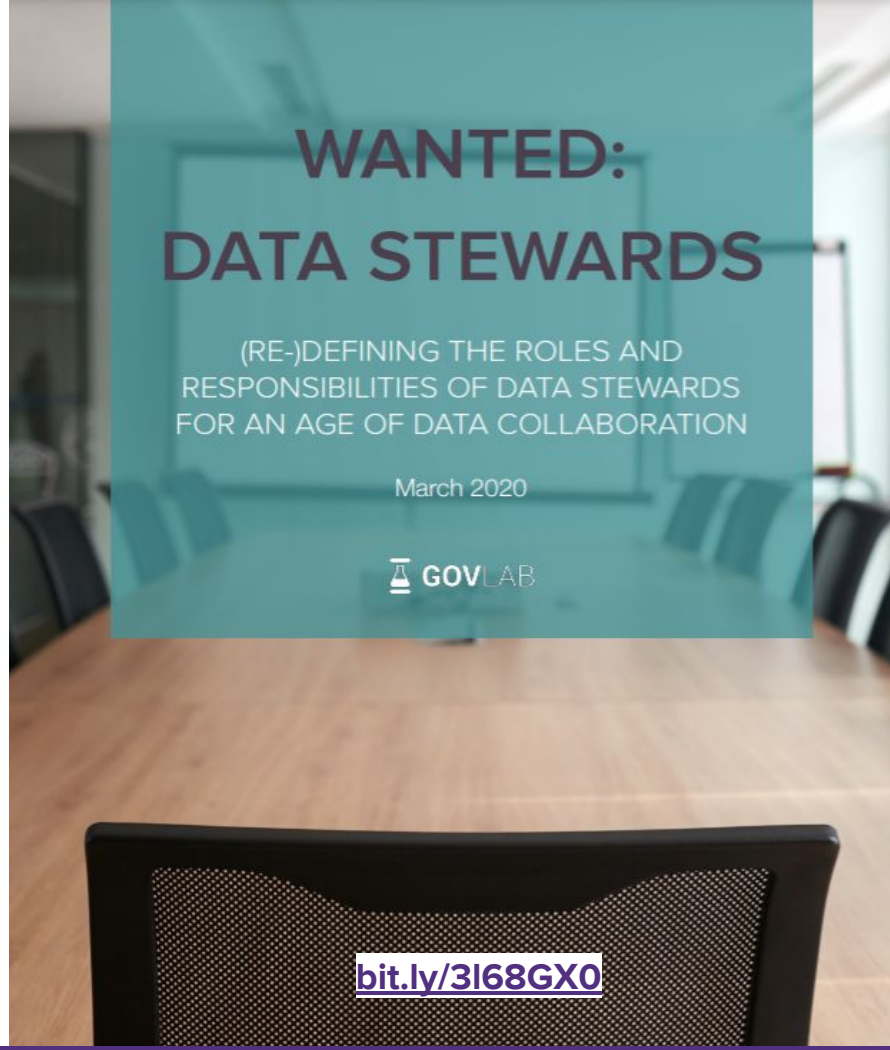
2. BUILD THE HUMAN INFRASTRUCTURE FOR DATA COLLABORATION: DATA STEWARDS



https://datastewards.net/

2. WANTED: DATA STEWARDSHIP

DATA STEWARDSHIP: the functions and roles that enable the re-use of data for public benefit in a systematic, sustainable and responsible way through data collaboration.



2. RE-IMAGINING DATA STEWARDSHIP

Data Stewardship within a Scientific and Library Context

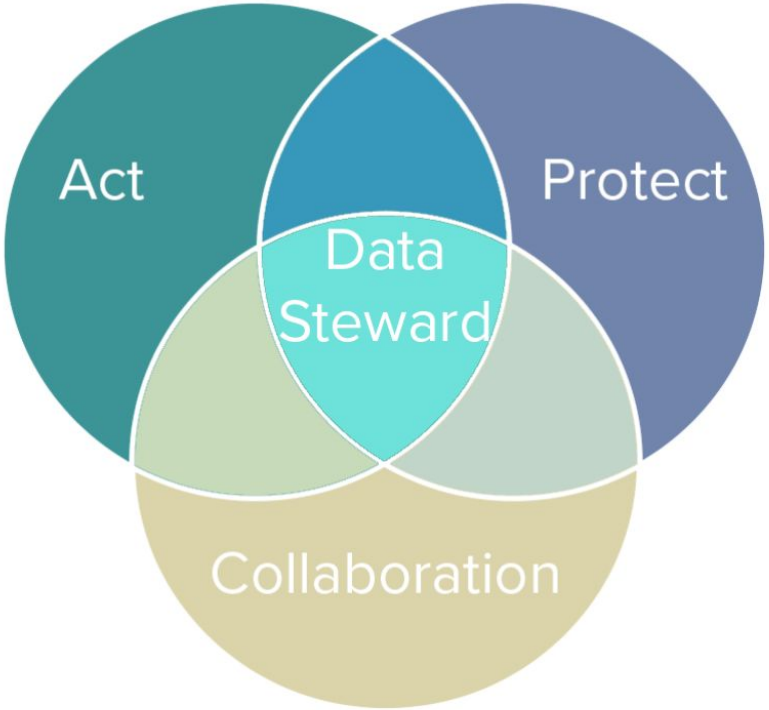
Data Stewardship within a Corporate Data Governance Context

Integrity of the Data

Security of the Data

Re-Use of the Data?

2. RE-IMAGINING DATA STEWARDSHIP

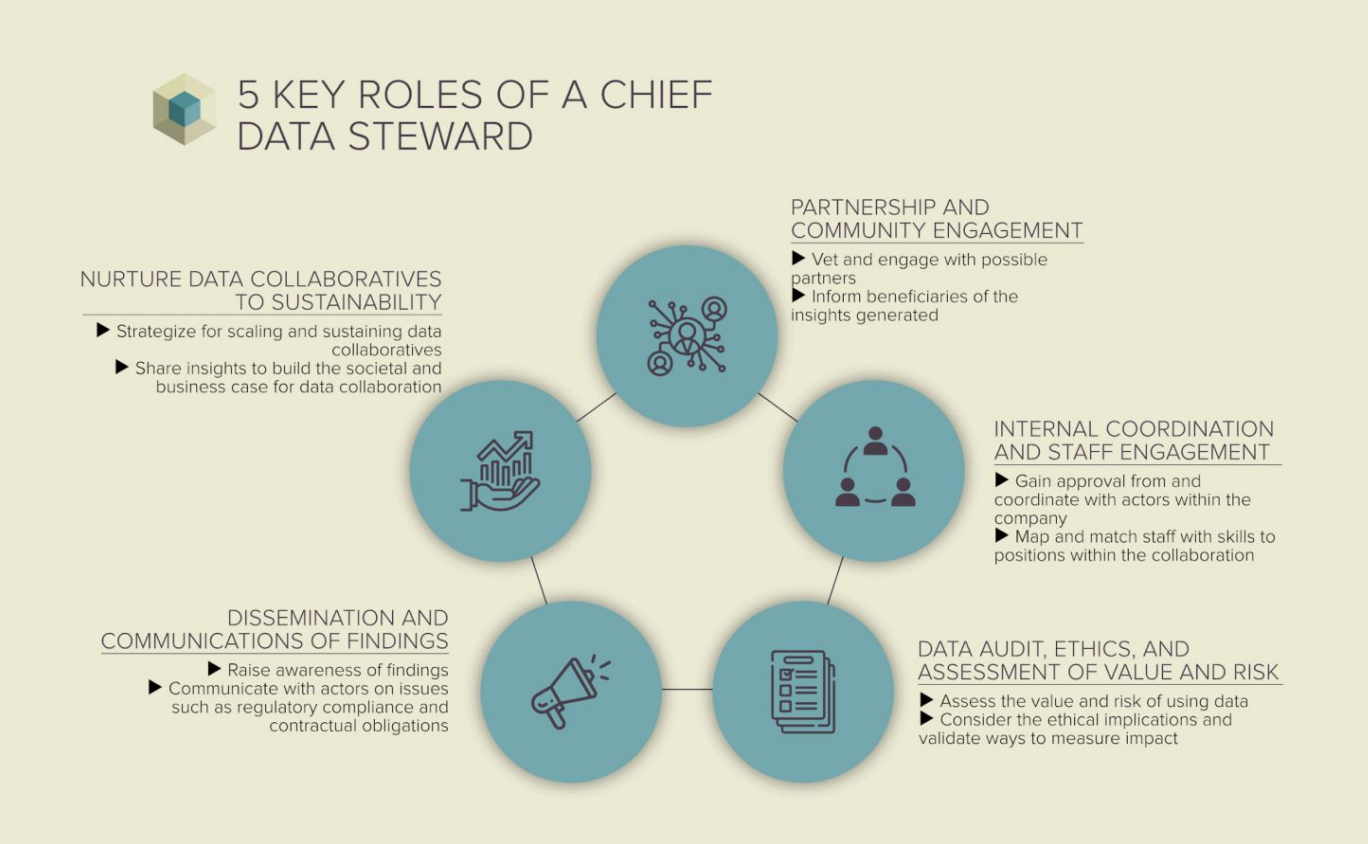


Collaborate:
Working with others to unlock the inherent value of data when it serves the public good.

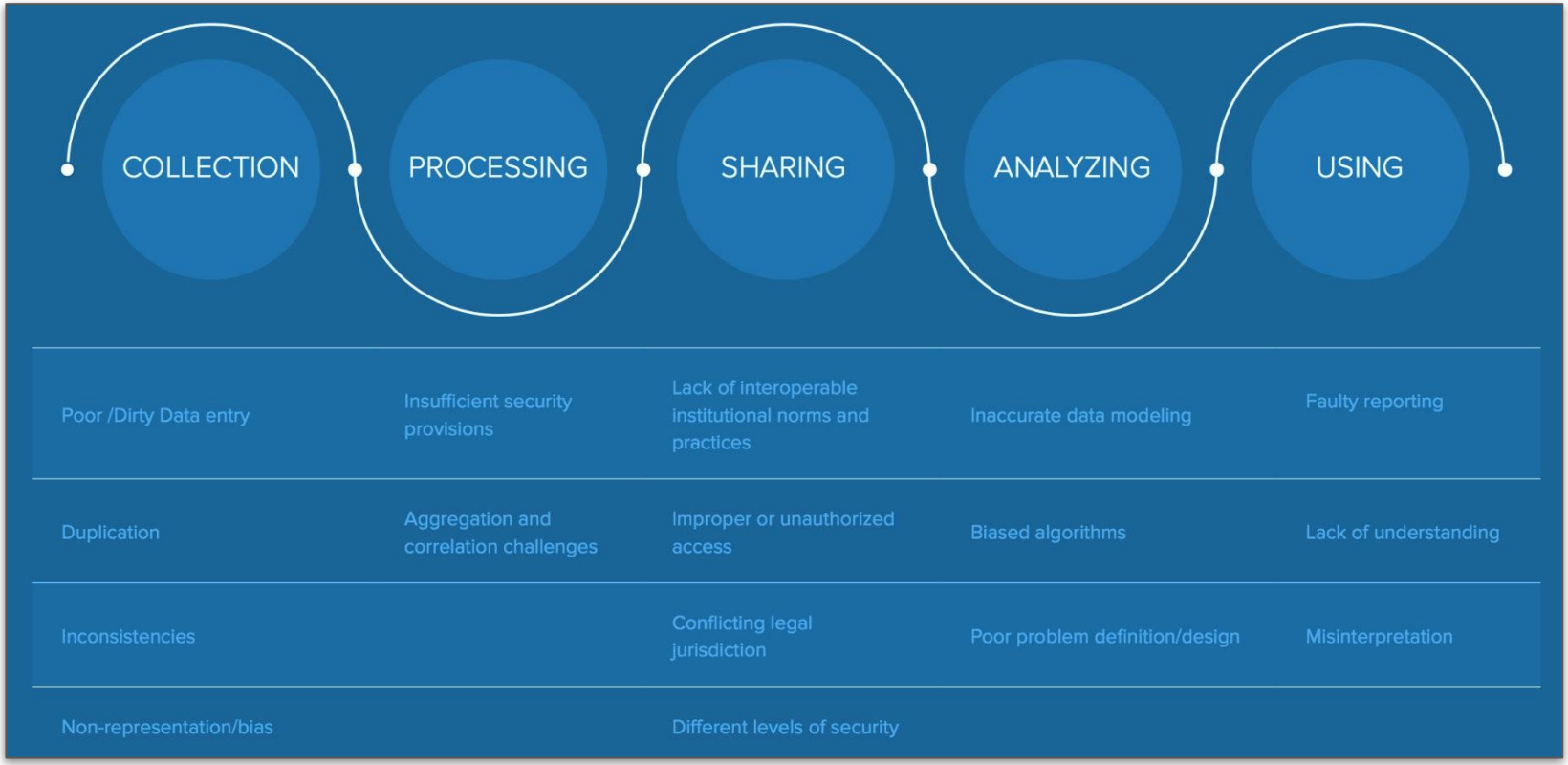
Protect:
Managing data ethically and preventing harm to all whose data may be shared.

Act:
Proactively identifying partners who can unlock value and insights.

2. THE FUNCTIONS OF A DATA STEWARD



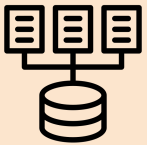
3. ASSESS AND MITIGATE RISKS ACROSS THE DATA LIFECYCLE



3. ASSESS AND MITIGATE RISK ACROSS THE DATA LIFECYCLE

https://dataresponsibilityjourney.org

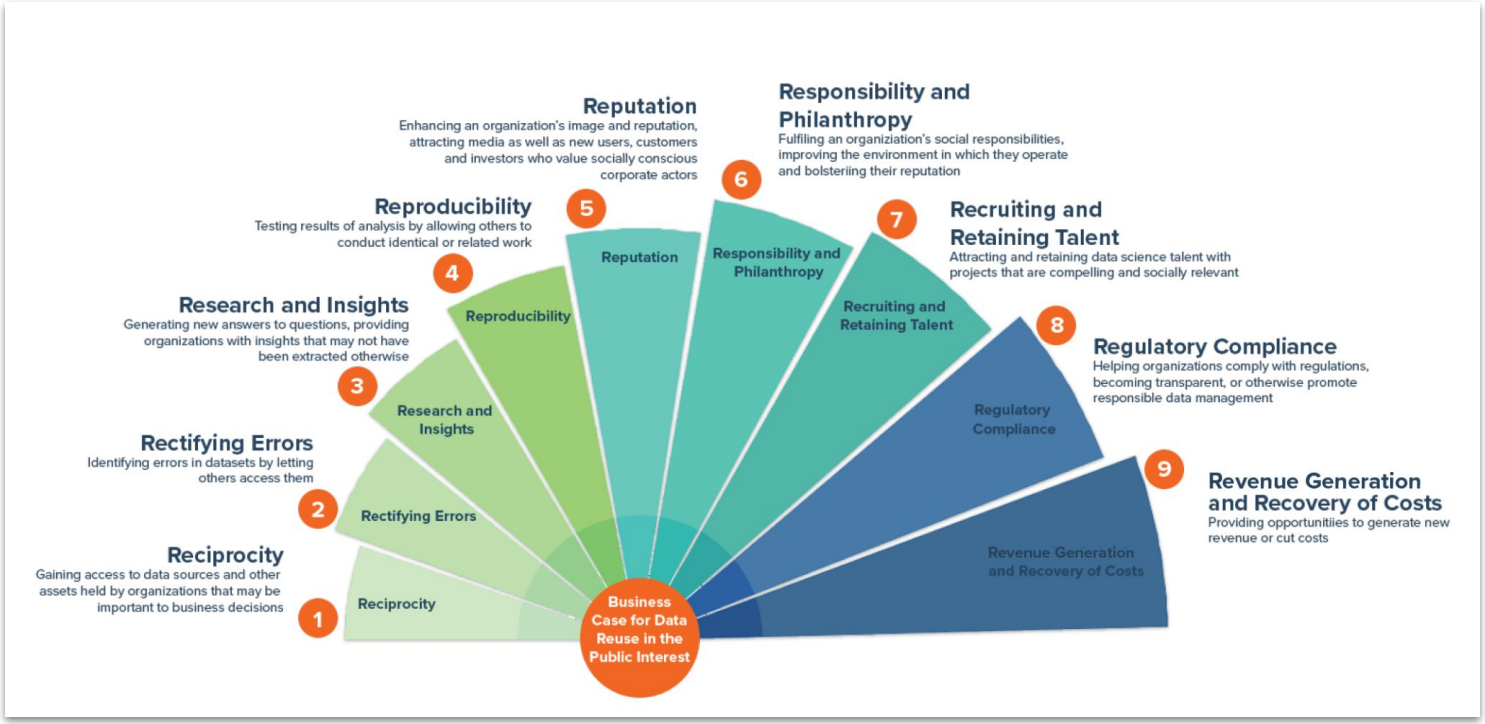
3. ASSESS AND MITIGATE RISK ACROSS THE DATA LIFECYCLE

Stage	Risks
Planning	Unclear purpose and goals of the project
	Financial resources and stakeholder partnerships hard to find and maintain
 Collecting	Data subjects and sources and their consent not being protected
	Poor quality (duplication or inconsistencies) or dirty data, data bias/non-representation
Processing	Poor anonymization and privacy of data subjects and sources
	Insufficient security provisions, inaccessibility, aggregation and correlations challenges

3. ASSESS AND MITIGATE RISK ACROSS THE DATA LIFECYCLE

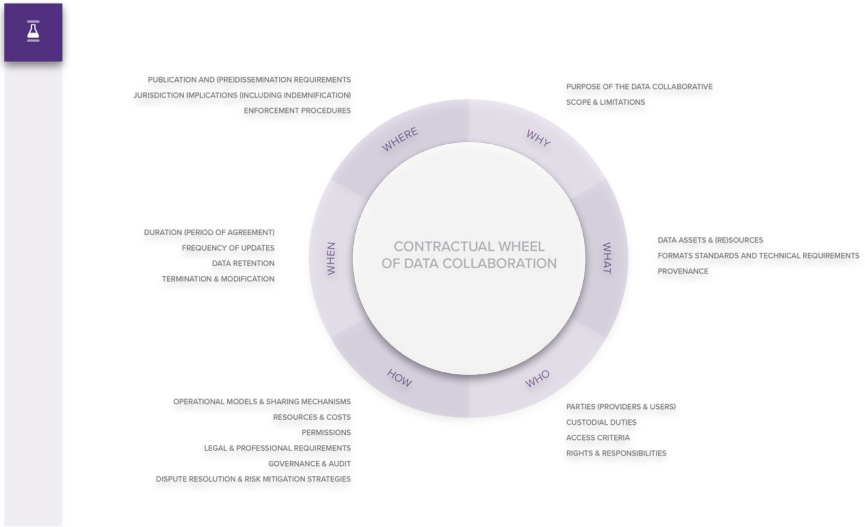
Stage	Risks
Sharing	Lack of trust and communication among partners, conflicting jurisdictions, different levels of security, data compatibility with different IT systems
	Poor balance between interoperability and context-rich, relevant data
Analyzing	Poor focus and rigour in the data analysis, inaccurate data modeling, poor definitions and problem design
	Black boxes, blind spots, unexplainable automated decision-making, inequity production and reproduction (eg. biased algorithms)
Using	Faulty reporting (eg. misinterpretation), malicious actors using findings for unforeseen and/or harmful purposes
	Poor communication strategy and external use and repurpose of findings

4. BUILD THE BUSINESS CASE FOR DATA COLLABORATION



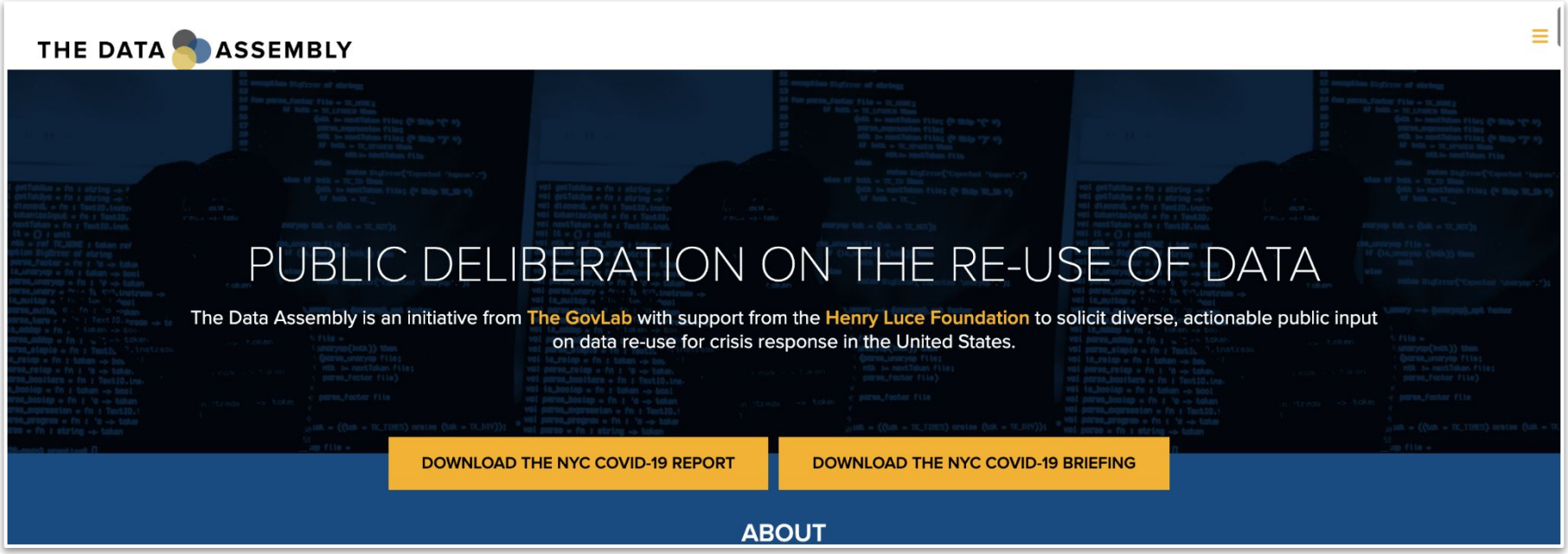
https://businesscase.opendatapolicylab.org/

5. STREAMLINE DATA SHARING AGREEMENTS



https://contractsfordatacollaboration.org/

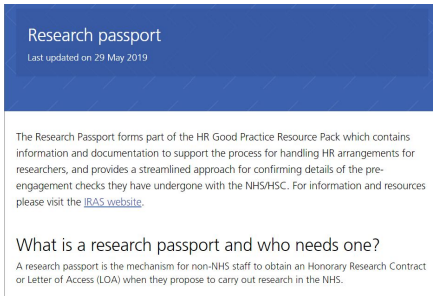
6. ACQUIRE A SOCIAL LICENSE FOR DATA RE-USE



https://thedataassembly.org/

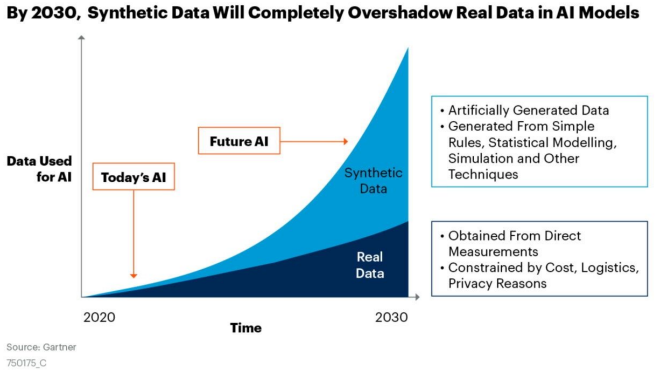
7. FOCUS ON FIT FOR PURPOSE INFRASTRUCTURE

ACCESS CONTROLS



https://bit.ly/3r3iS7r

DATA PREPARATION



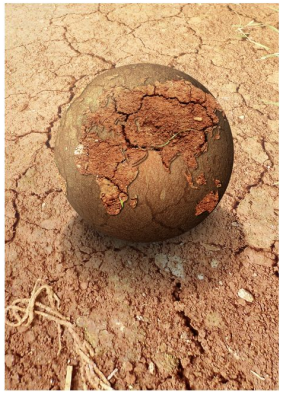
bit.ly/3XubmPg

DATA TRANSFER/STORAGE

metis-sandbox.europeana.eu/

8. FOCUS ON THE LAST MILE: INVEST & INNOVATE IN COMMUNICATIONS

dt.asu.edu/



Case studies

Data drives media coverage of climate refugees

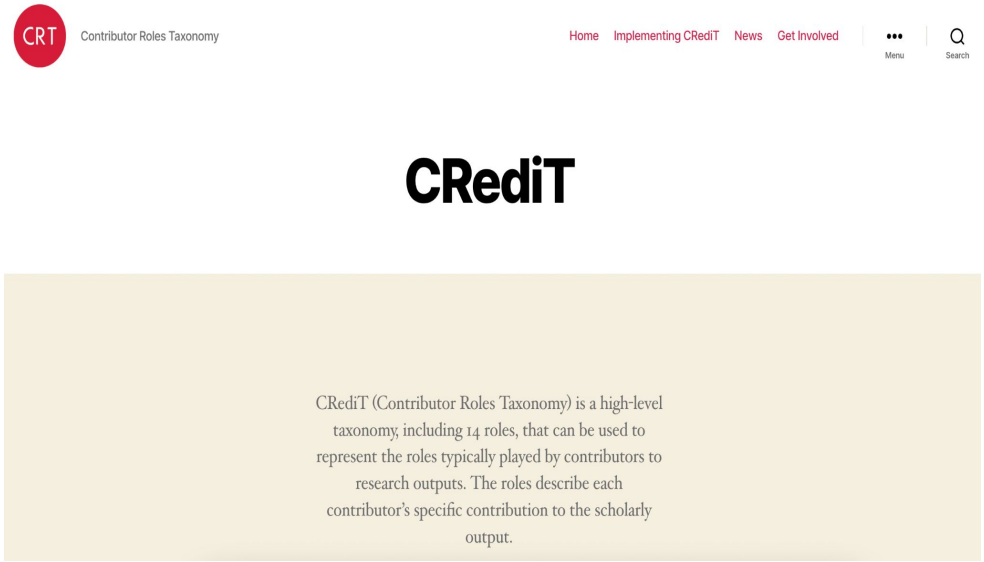
Giving voice to “world’s forgotten victims”

Newsrooms around the world are stepping up their climate coverage by investing in the resources needed to grow and support their teams. What are some good examples of climate reporting from around the world that we should look at? This article looks at data journalism unique position in analysing the global crisis.

READ ONLINE

datajournalism.com/

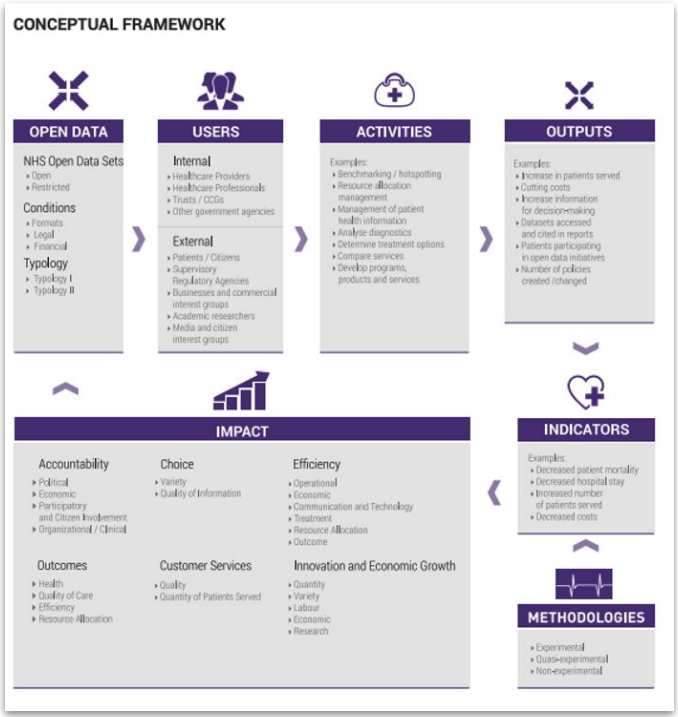
9. NAME AND FAME: FOCUS ON CULTURE CHANGE

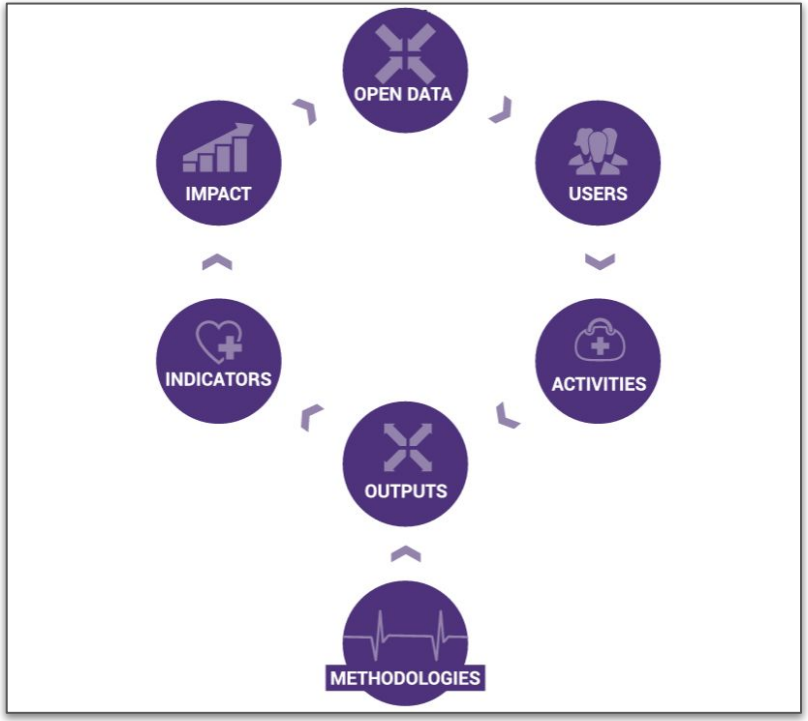


https://credit.niso.org/

https://incubator.opendatapolicylab.org/

10. BECOME DATA-DRIVEN ABOUT THE USE OF DATA





https://thegovlab.org/static/files/publications/nhs-health-report.pdf

Final Reflections

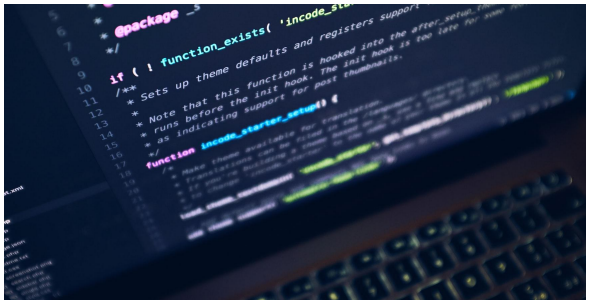
TODAY'S ENVIRONMENT

POLYCRISIS

Officials responding to the multitude and complexity of modern policy challenges need decision support.

DIGITAL TRANSFORMATION

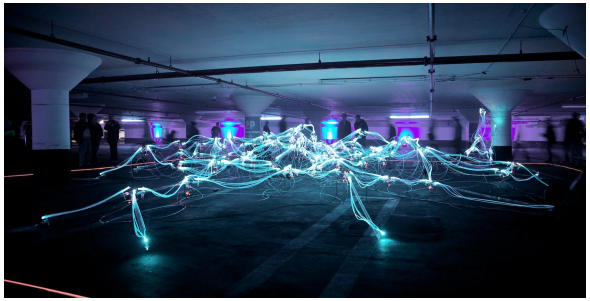
The continued datafication of our world is the result of an ongoing digital transformation, which we must reckon with.



DISINFORMATION

Alongside digital transformation, the risk of false or misleading information spread to cause harm is a key challenge.

TODAY'S ENVIRONMENT



TECHNOLOGICAL ADVANCES

New tools such as Generative AI and PETs could present new opportunities to harness data.

DECLINE IN TRUST AND EFFECTIVENESS

Our modern era is defined by declining trust in institutions and methods.. This decline both complicates efforts and calls us to use data to address new challenges.

BUDGET CHALLENGES

While open data projects can often be launched cheaply, budget challenges pose a threat to their long term sustainability.

A NEW ERA & PARADIGM OF STEWARDSHIP

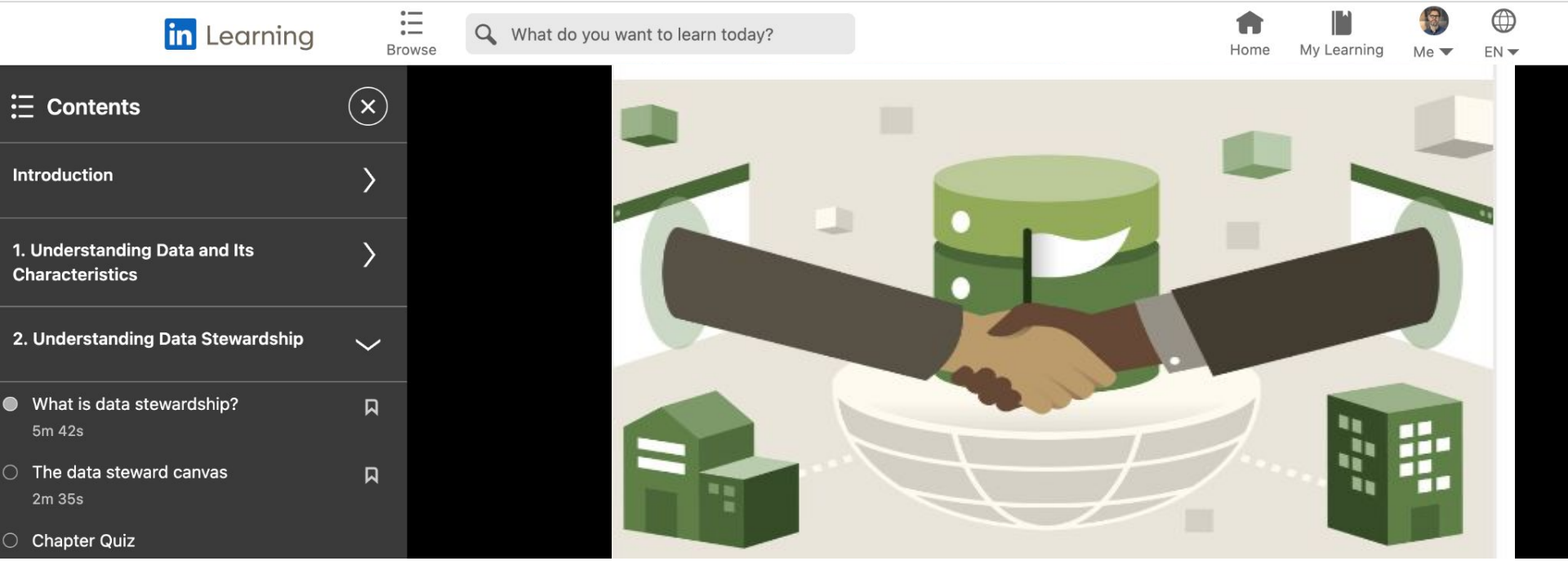
https://datacollaboratives.org/

https://datastewards.net/

10 STEPS TOWARD A NEW ERA OF DATA STEWARDSHIP

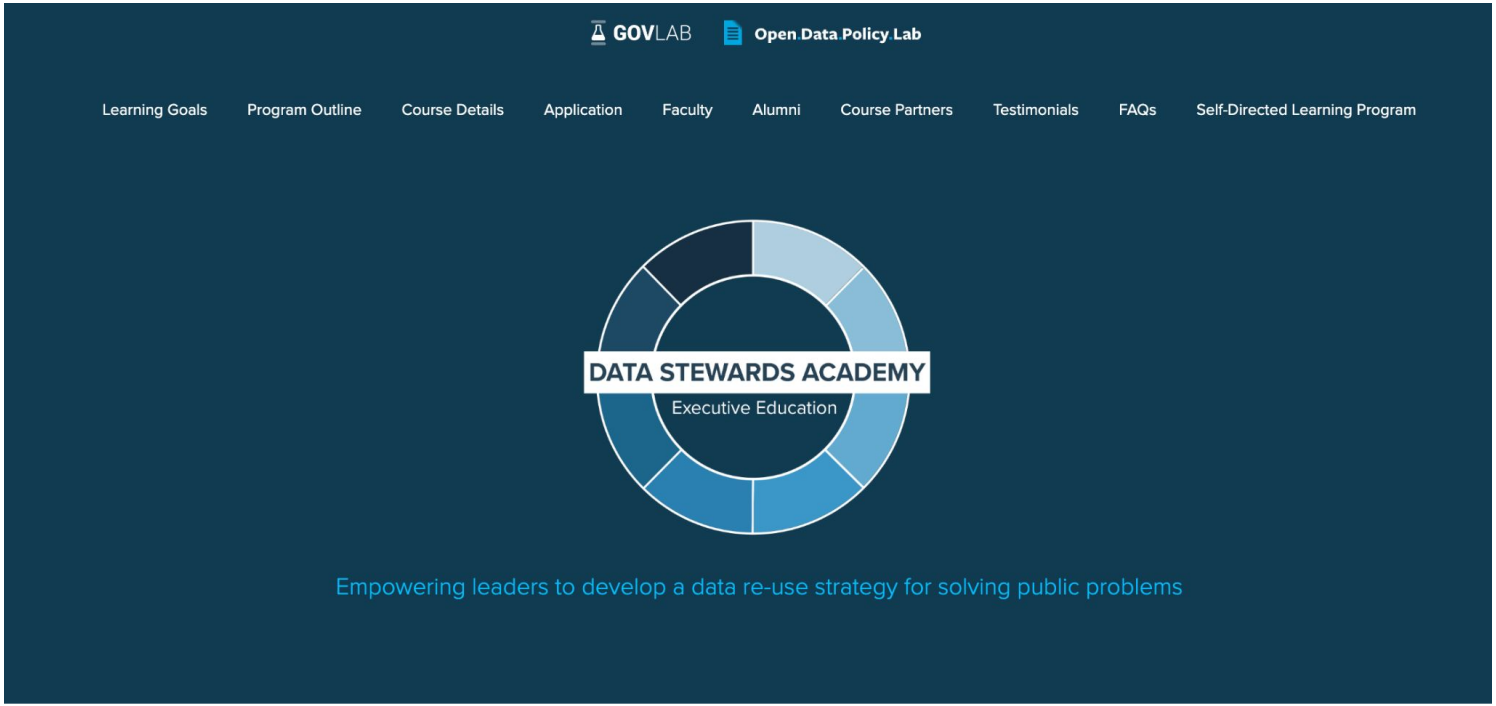
- Strengthen **Demand** For Data
- Build The **Human Infrastructure** For Data Collaboration: **Data Stewards**
- Assess And Mitigate Risks Across **The Data Lifecycle**
- Build The **Business Case** for Data Collaboration
- Streamline **Data Sharing Agreements**
- Acquire a **Social License** For Data Re-Use
- Focus On **Fit For Purpose** Technical Infrastructure
- Focus on the Last Mile: Invest & Innovate In **Communications**
- Name And Fame: Focus on **Culture Change**
- Become **Data-Driven About The Use Of Data**

Global Data Stewardship



https://www.linkedin.com/learning/global-data-stewardship/

Data Stewards Academy



https://course.opendatapolicylab.org/

www.thegovlab.org