

Toward a New Era & Paradigm of Data Stewardship

MET2023

Dr. Stefaan Verhulst

July 4, 2023



DEEPENING OUR UNDERSTANDING OF HOW TO GOVERN
MORE EFFECTIVELY AND LEGITIMATELY THROUGH TECHNOLOGY

 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.



LEARN MORE →

EXPLORE THE BLOG →

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

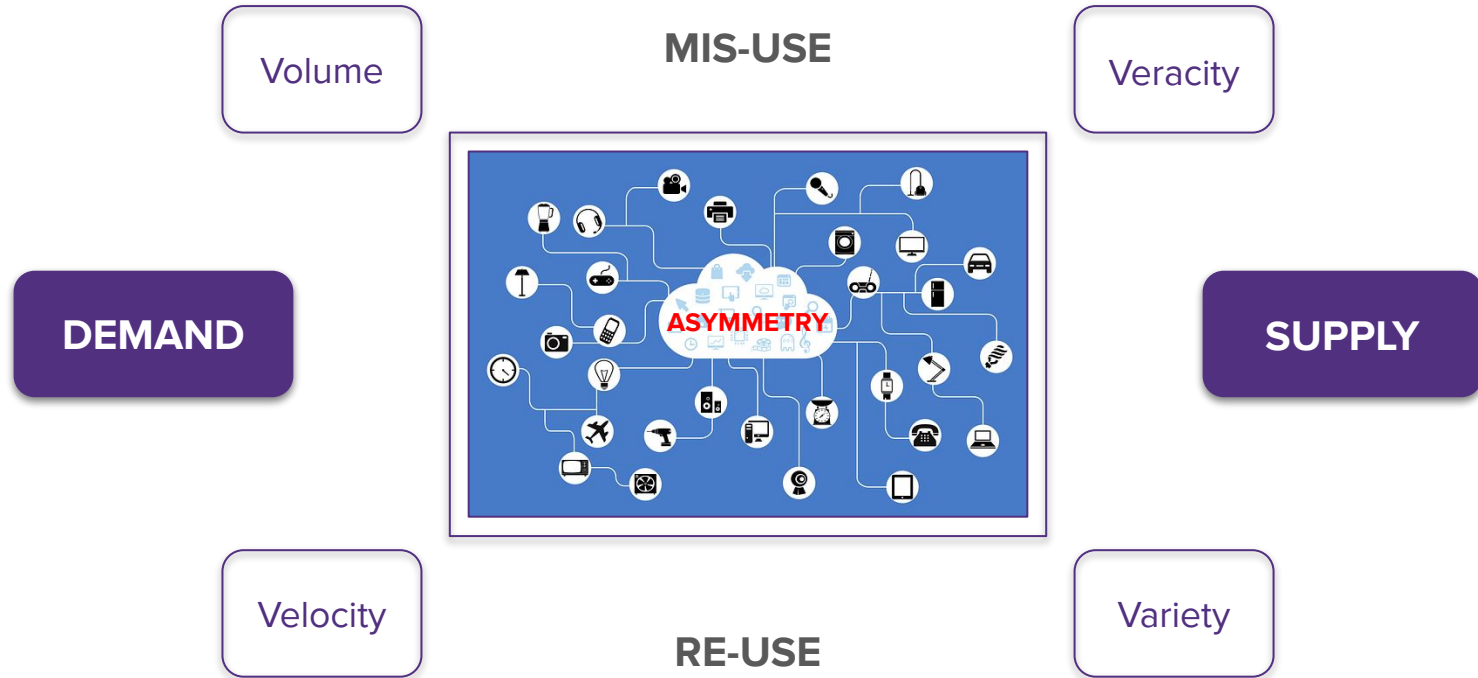


TODAY'S PUBLIC PROBLEMS REQUIRE INNOVATION



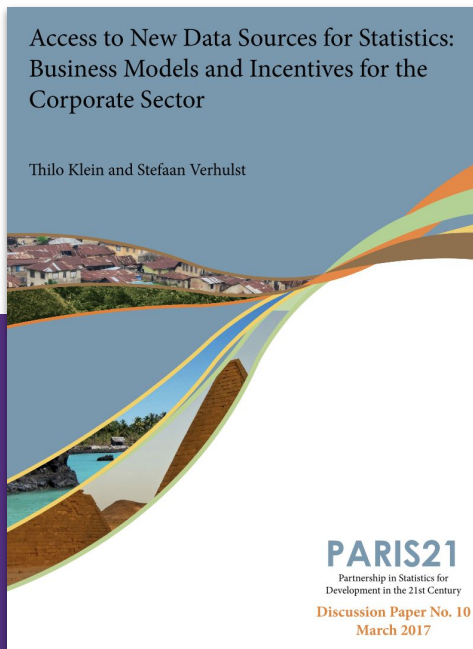


DATAFICATION: THE MIS-USE & RE-USE OF DATA





NEW DATA SOURCES FOR STATISTICS



<https://bit.ly/SVParis21>



bit.ly/3PxtPZh



B2G DATA SHARING



bit.ly/3i5eEW2



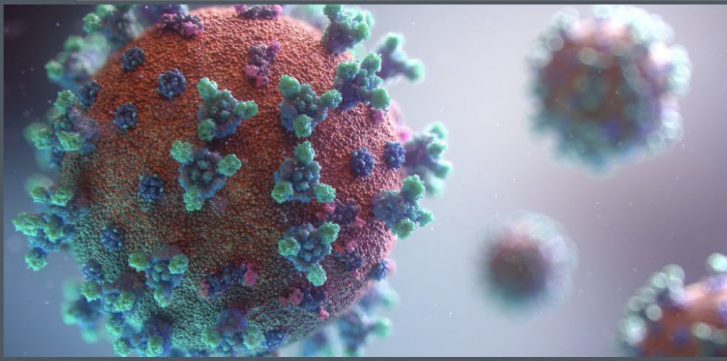
Concretely, the Data Act includes:

- **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

The #Data4COVID19 Review

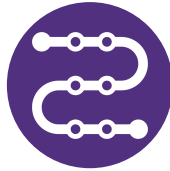
Assessing the Use of Non-Traditional Data During a Pandemic Crisis



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



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).

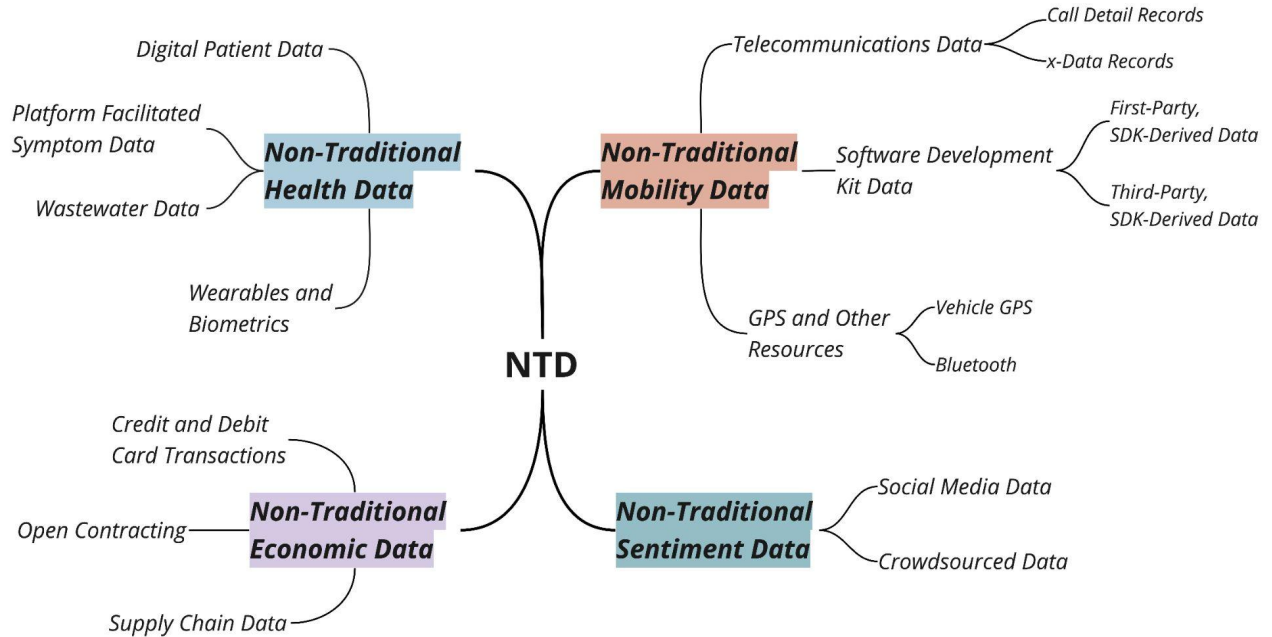
review.data4covid19.org



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

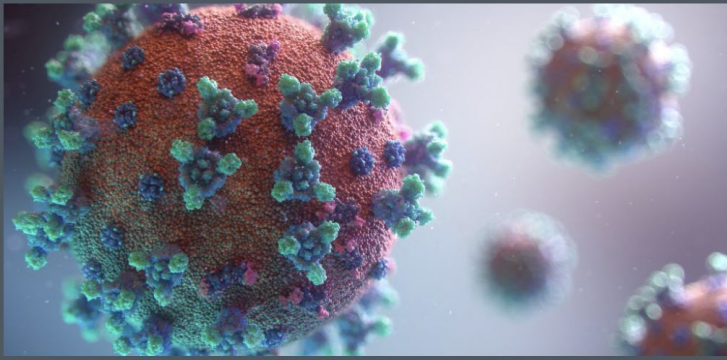
	Traditional Data	NTD
What it includes	Data collected about a topic	Proxies for the data we intend to measure
Where it is located	Analog + digital systems	Digital systems
Data Type	“Small”, Structured	“Big”, Often Unstructured
Why it is generated	Data collected for a specific purpose	Data that has been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

Types of NTD used during COVID-19



The #Data4COVID19 Review

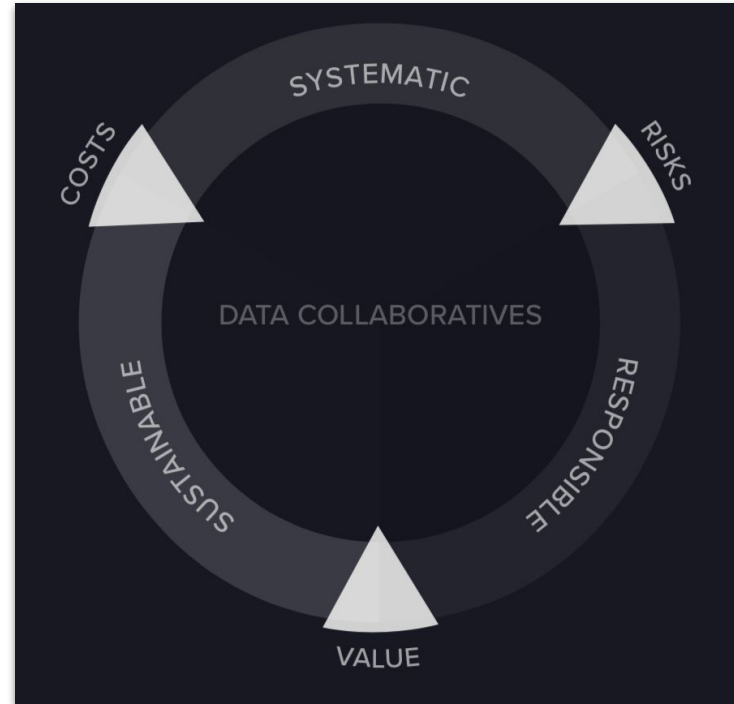
Assessing the Use of Non-Traditional Data During a Pandemic Crisis



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



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

ANNOUNCING: First of its Kind Executive Course on Data Stewardship - Focused on Data Re-Use in the Public Interest

Learn how to initiate a data strategy in the public interest from key players in the field

18 November 2020



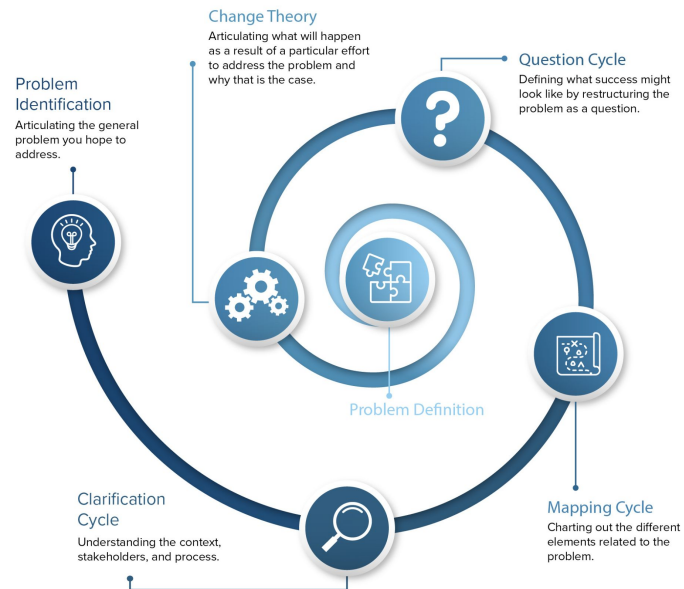
10 Steps Toward a New Era & Paradigm of Data Stewardship



1. STRENGTHEN DEMAND FOR (Non Traditional) DATA



the100questions.org



bit.ly/ODPLProbTool

IMPORTANCE OF THE 100 QUESTIONS

What are we trying to accomplish?

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



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.



<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.

WANTED: DATA STEWARDS

(RE-)DEFINING THE ROLES AND
RESPONSIBILITIES OF DATA STEWARDS
FOR AN AGE OF DATA COLLABORATION

March 2020



bit.ly/3I68GX0



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

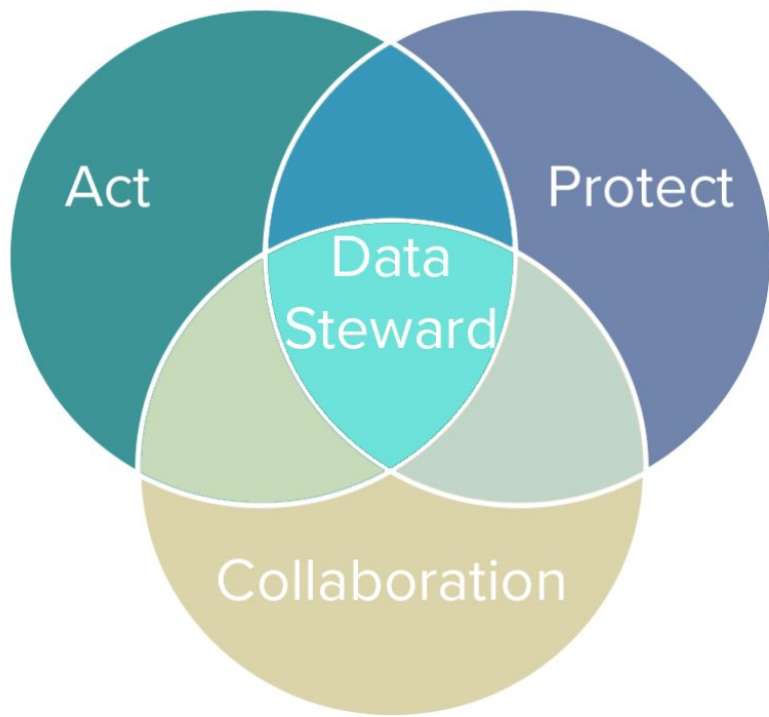
Security of the Data



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



5 KEY ROLES OF A CHIEF DATA STEWARD

NURTURE DATA COLLABORATIVES TO SUSTAINABILITY

- ▶ Strategize for scaling and sustaining data collaboratives
- ▶ Share insights to build the societal and business case for data collaboration



PARTNERSHIP AND COMMUNITY ENGAGEMENT

- ▶ Vet and engage with possible partners
- ▶ Inform beneficiaries of the insights generated

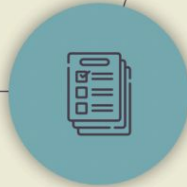


INTERNAL COORDINATION AND STAFF ENGAGEMENT

- ▶ Gain approval from and coordinate with actors within the company
- ▶ Map and match staff with skills to positions within the collaboration

DISSEMINATION AND COMMUNICATIONS OF FINDINGS

- ▶ Raise awareness of findings
- ▶ Communicate with actors on issues such as regulatory compliance and contractual obligations

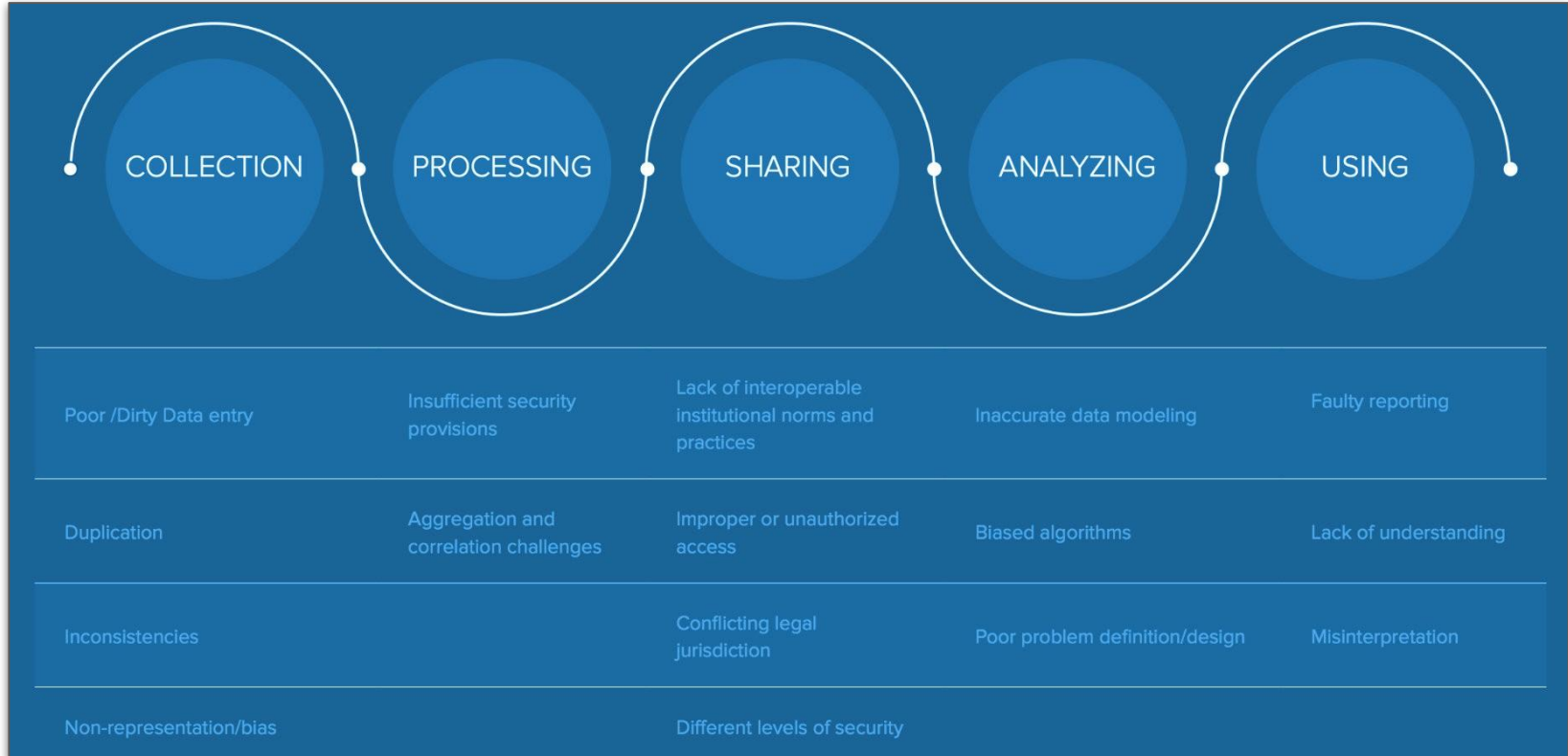


DATA AUDIT, ETHICS, AND ASSESSMENT OF VALUE AND RISK

- ▶ Assess the value and risk of using data
- ▶ Consider the ethical implications and validate ways to measure impact



3. ASSESS AND MITIGATE RISKS ACROSS THE DATA LIFECYCLE





3. ASSESS AND MITIGATE RISK ACROSS THE DATA LIFECYCLE

BETA v0.1

RESOURCES ABOUT

1 PLANNING 2 COLLECTING 3 PROCESSING 4 SHARING 5 ANALYZING 6 USING

 **GOVLAB**

DATA RESPONSIBILITY JOURNEY




Risks & Responsibilities Throughout the Data Lifecycle

The Data Responsibility Journey for Data Collaboratives is an assessment tool that outlines the opportunities and risks to consider at each stage of the data lifecycle when implementing a data collaborative.

<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

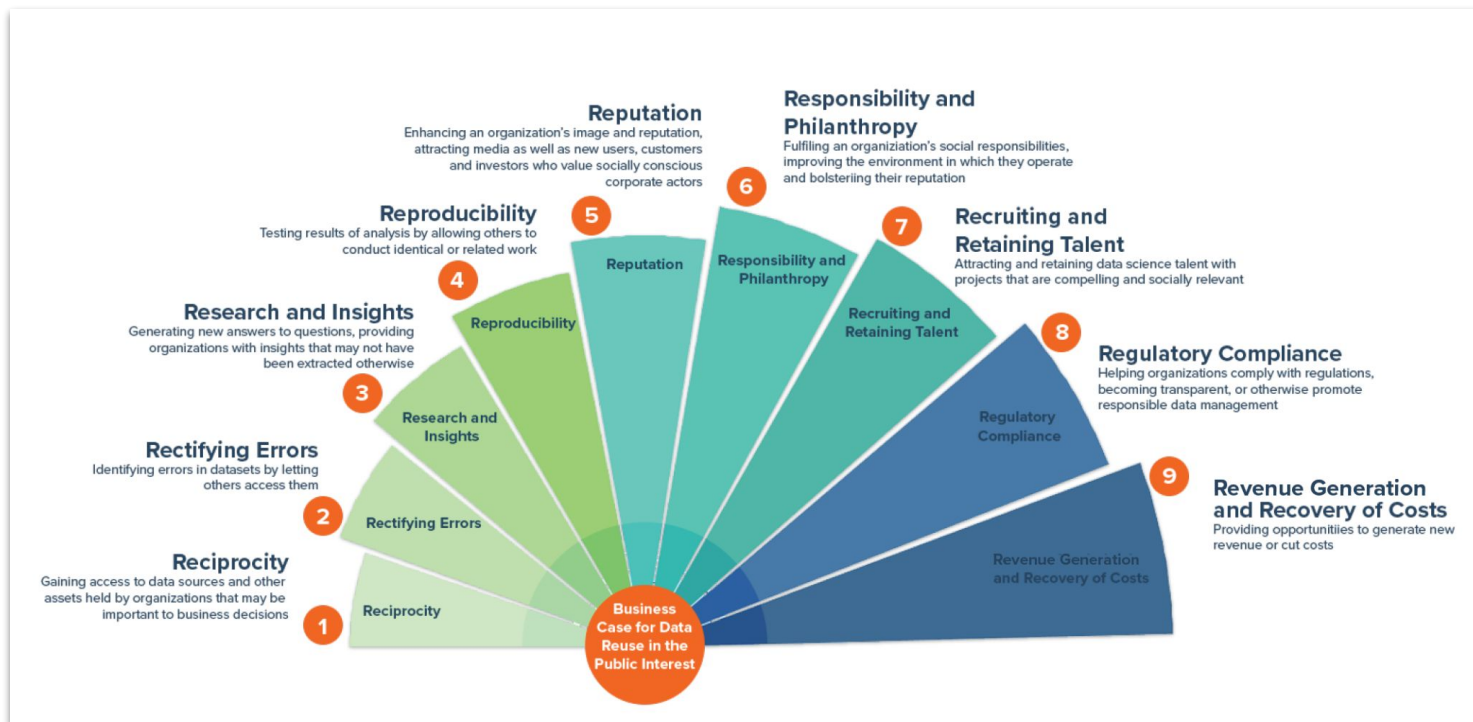


3. ASSESS AND MITIGATE RISK ACROSS THE DATA LIFECYCLE

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

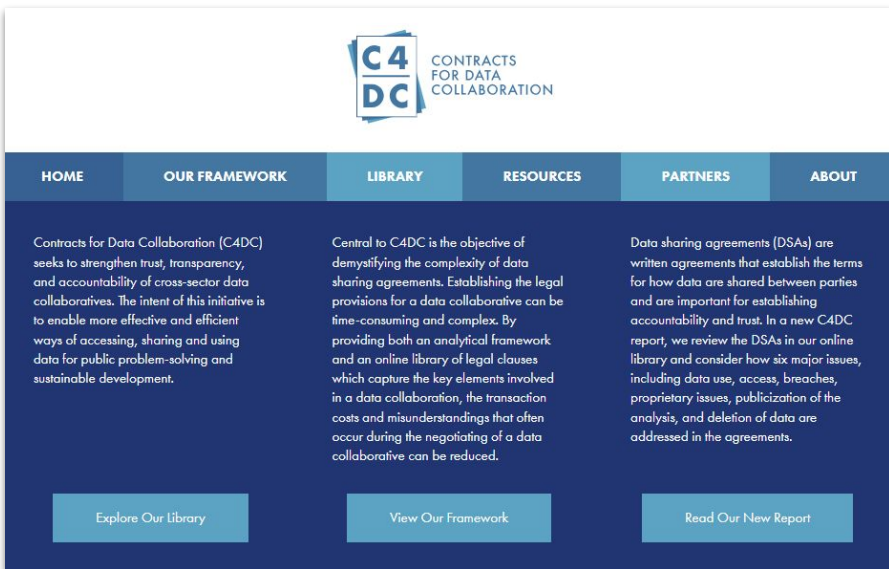


4. BUILD THE BUSINESS CASE FOR DATA COLLABORATION



<https://businesscase.opendatapolicylab.org/>

5. STREAMLINE DATA SHARING AGREEMENTS



C4 DC CONTRACTS FOR DATA COLLABORATION

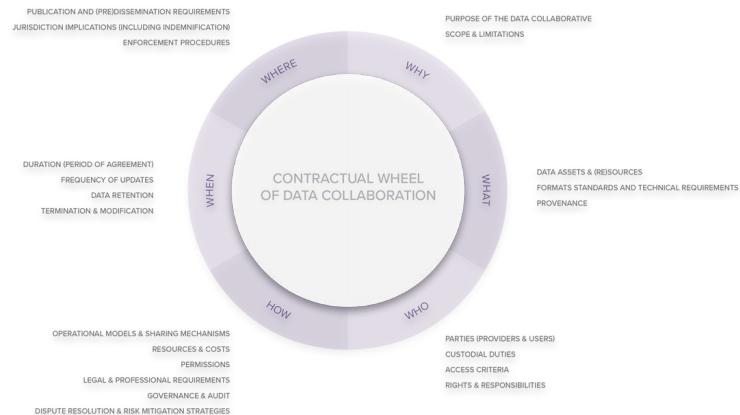
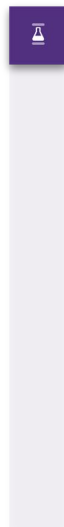
HOME **OUR FRAMEWORK** **LIBRARY** **RESOURCES** **PARTNERS** **ABOUT**

Contracts for Data Collaboration (C4DC) seeks to strengthen trust, transparency, and accountability of cross-sector data collaboratives. The intent of this initiative is to enable more effective and efficient ways of accessing, sharing and using data for public problem-solving and sustainable development.

Central to C4DC is the objective of demystifying the complexity of data sharing agreements. Establishing the legal provisions for a data collaborative can be time-consuming and complex. By providing both an analytical framework and an online library of legal clauses which capture the key elements involved in a data collaboration, the transaction costs and misunderstandings that often occur during the negotiating of a data collaborative can be reduced.

Data sharing agreements (DSAs) are written agreements that establish the terms for how data are shared between parties and are important for establishing accountability and trust. In a new C4DC report, we review the DSAs in our online library and consider how six major issues, including data use, access, breaches, proprietary issues, publicization of the analysis, and deletion of data are addressed in the agreements.

[Explore Our Library](#) [View Our Framework](#) [Read Our New Report](#)



<https://contractsfordatacollaboration.org/>



6. ACQUIRE A SOCIAL LICENSE FOR DATA RE-USE

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.

[DOWNLOAD THE NYC COVID-19 REPORT](#) [DOWNLOAD THE NYC COVID-19 BRIEFING](#)

ABOUT

<https://thedataassembly.org/>



7. FOCUS ON FIT FOR PURPOSE INFRASTRUCTURE

ACCESS CONTROLS

Research passport

Last updated on 29 May 2019

The Research Passport forms part of the HR Good Practice Resource Pack which contains information and documentation to support the process for handling HR arrangements for researchers, and provides a streamlined approach for confirming details of the pre-engagement checks they have undergone with the NHS/HSC. For information and resources please visit the [IRAS website](#).

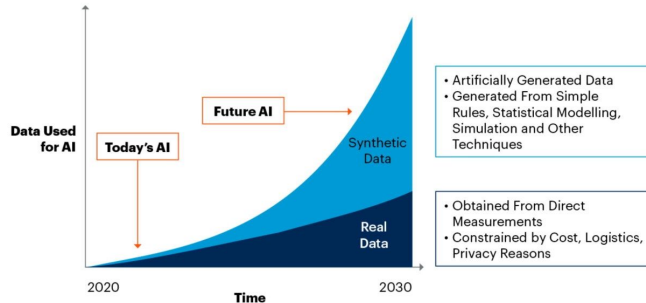
What is a research passport and who needs one?

A research passport is the mechanism for non-NHS staff to obtain an Honorary Research Contract or Letter of Access (LOA) when they propose to carry out research in the NHS.

<https://bit.ly/3r3iS7r>

DATA PREPARATION

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



Source: Gartner
750175_C

Gartner

bit.ly/3XubmPg

DATA TRANSFER/STORAGE

METIS SANDBOX
TEST YOUR DATA BEFORE INGESTION

Logo: europeana
www.europeana.eu
@EuropeanaEU

Co-Funded by the Connecting Europe Facility of the European Union

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



Contributor Roles Taxonomy

[Home](#) [Implementing CRedit](#) [News](#) [Get Involved](#)



Menu



Search

CRedit

CRedit (Contributor Roles Taxonomy) is a high-level taxonomy including 14 roles, that can be used to represent the roles typically played by contributors to research outputs. The roles describe each contributor's specific contribution to the scholarly output.

<https://credit.niso.org/>

2021-2022 INAUGURAL PROGRAM COHORT



Oliver Bjornsson

Policy Analyst, Mayor's Office of Data Analytics, City of New York, NY, United States of America



Jennifer Bodnarchuk

Senior Data Scientist, Innovation & Technology Department, City of Winnipeg, Canada



Emri Brickner

Smart City Department Manager, Municipal Division of IT & Innovation, Beer Sheva, Israel



Andrea Calderon

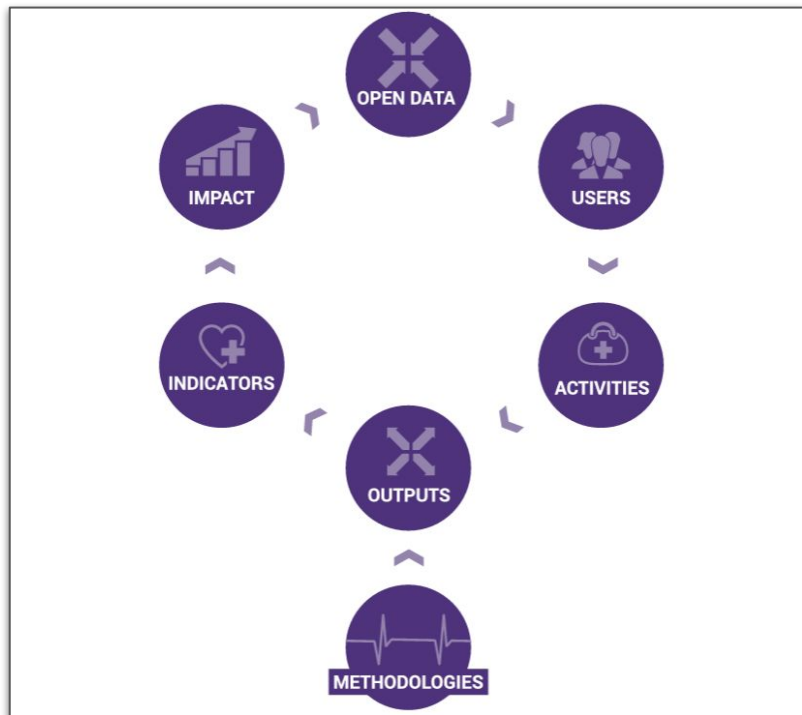
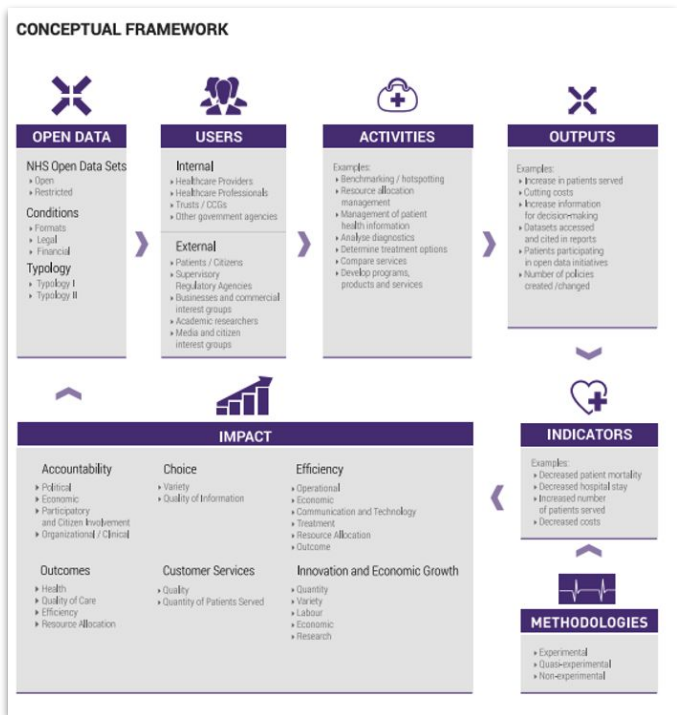
Race & Equity Data Analyst, Office of Equity & Inclusion, City of Albuquerque, New Mexico, United States of America



<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

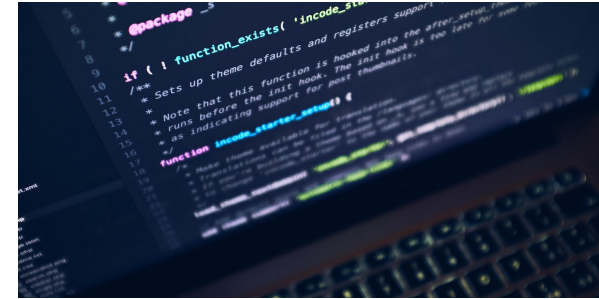


POLYCRISIS

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

DIGITAL TRANSFORMATION

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

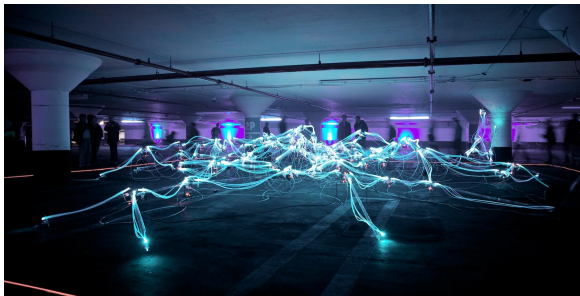


DISINFORMATION

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



TODAY'S ENVIRONMENT



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.



TECHNOLOGICAL ADVANCES

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

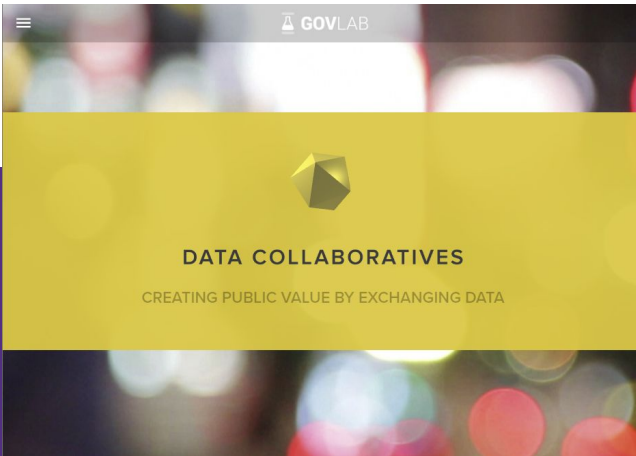


BUDGET CHALLENGES

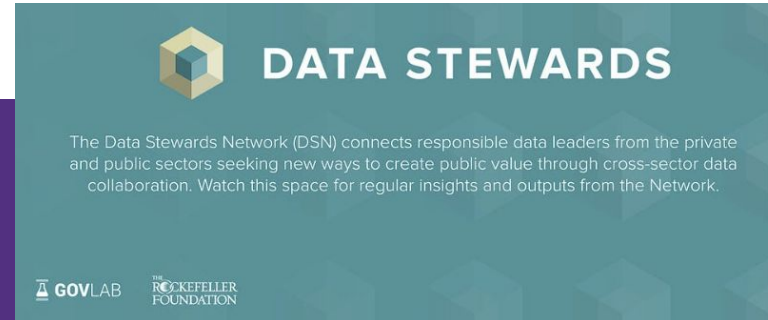
While open data projects can often be launched cheaply, budget challenges pose a threat to their long term sustainability.



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

in Learning

Browse

What do you want to learn today?

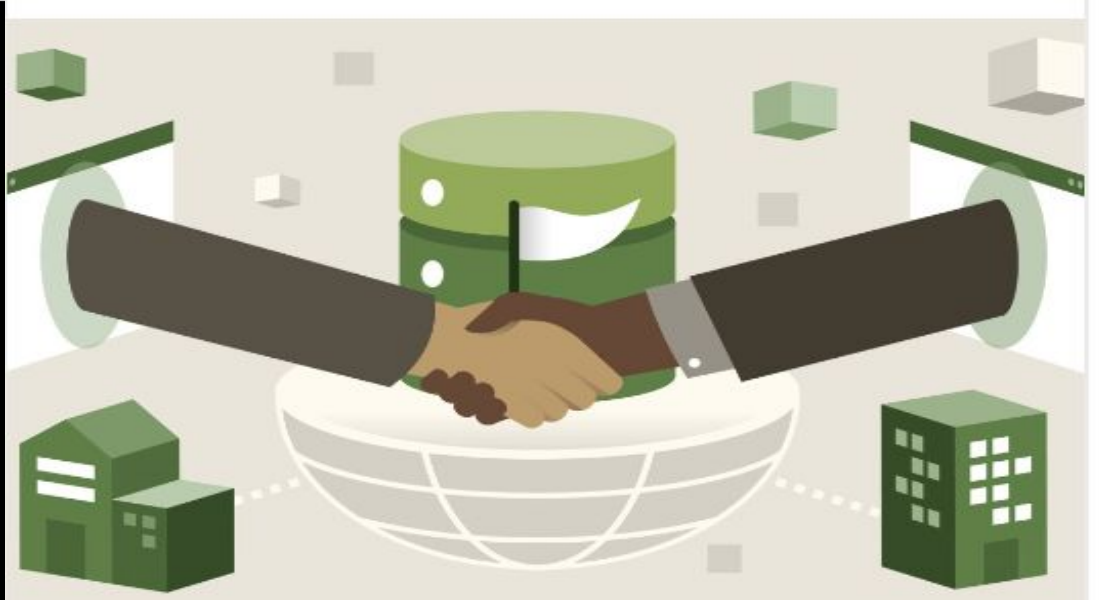
Home

My Learning

Me

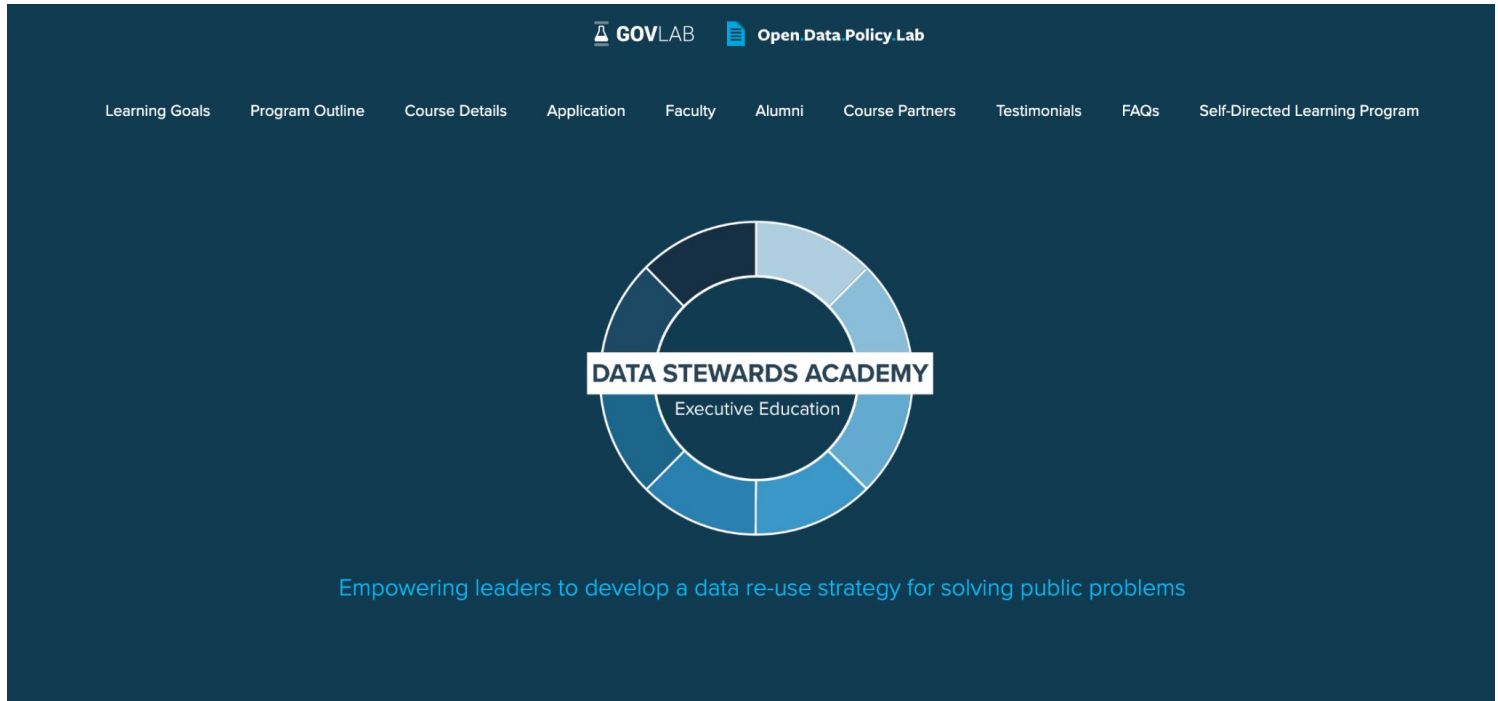
EN

- Contents
- Introduction
- 1. Understanding Data and Its Characteristics
- 2. Understanding Data Stewardship
 - What is data stewardship? 5m 42s
 - The data steward canvas 2m 35s
 - Chapter Quiz



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

Data Stewards Academy



The image shows a screenshot of the Data Stewards Academy website. At the top, there are logos for GOVLAB and Open Data Policy Lab. Below the logos is a navigation menu with the following items: Learning Goals, Program Outline, Course Details, Application, Faculty, Alumni, Course Partners, Testimonials, FAQs, and Self-Directed Learning Program. The main content area features a circular graphic composed of several segments in shades of blue and white. A white banner across the center of the circle contains the text "DATA STEWARDS ACADEMY" in bold, uppercase letters, with "Executive Education" written below it in a smaller font. Below the circular graphic, the text "Empowering leaders to develop a data re-use strategy for solving public problems" is displayed in a light blue color.

<https://course.opendatapolicylab.org/>



www.thegovlab.org