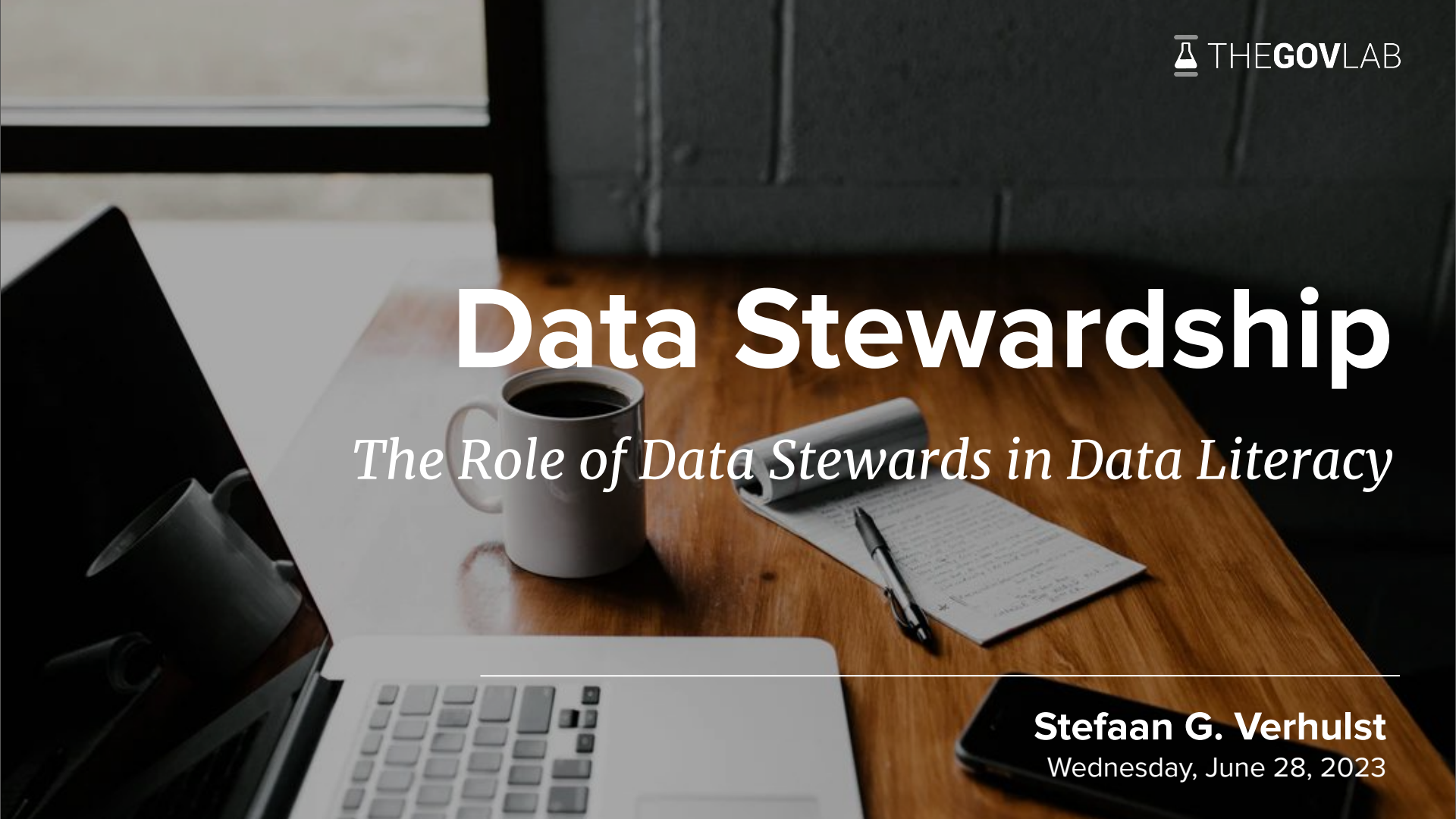


THEGOVLAB

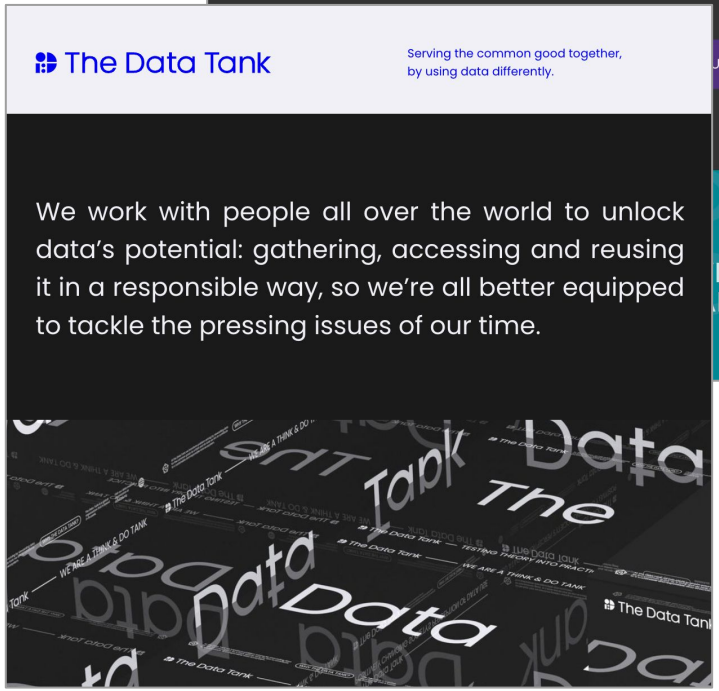
Data Stewardship

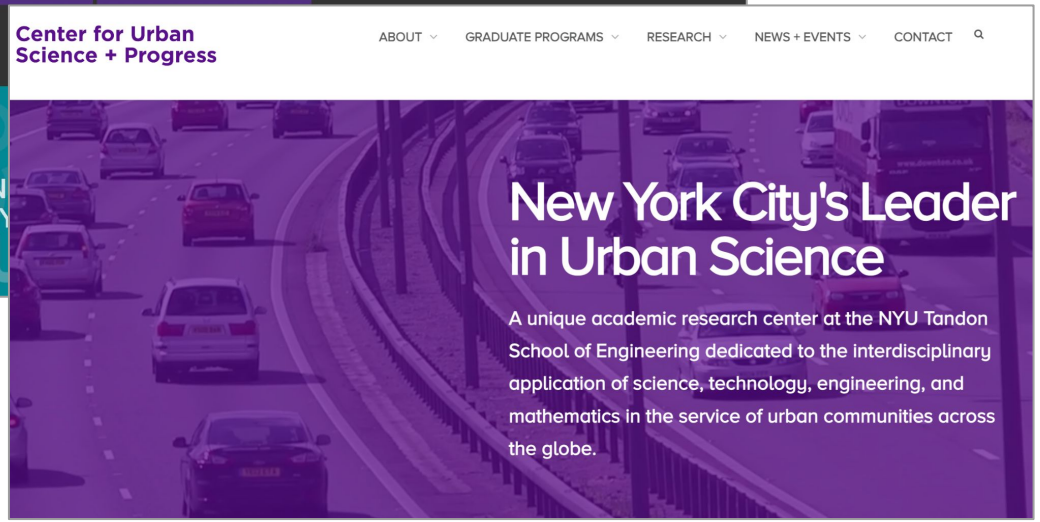
The Role of Data Stewards in Data Literacy

Stefaan G. Verhulst

Wednesday, June 28, 2023

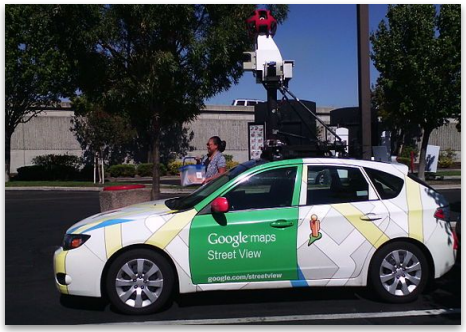

THEGOVLAB
DEEPENING OUR UNDERSTANDING OF HOW TO GOVERN
MORE EFFECTIVELY AND LEGITIMATELY THROUGH TECHNOLOGY
UT MORE
EXPLORE THE BLOG


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.

Introducing Data Stewardship

WE NEED TO INNOVATE HOW WE SOLVE PUBLIC PROBLEMS



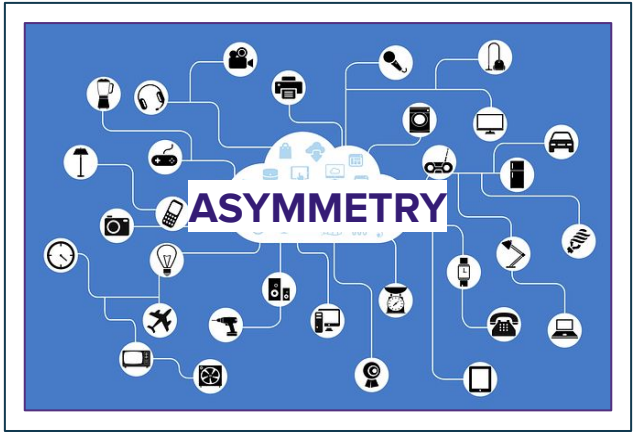
BRIDGING DATA ASYMMETRIES

Volume

MIS-USE

Veracity

DEMAND



SUPPLY

Velocity

RE-USE

Variety

THE POTENTIAL OF DATA (RE-)USE FOR SOCIAL GOOD

Sharing data can help advance the development of medicines and procedures. The Accelerating Medicines Partnership (AMP) is an example of effective data re-use to tackle diseases.

Data re-use can help us better understand the impact of human activity on our environment. It can also promote the implementation of environmental policies like the Global Fishing Watch.

Data can be invaluable during a crisis to help ensure aid and interventions are reaching those impacted. The UNHCR's Operational Data Portal: Refugee Situations is a key example.

DATA STEWARDSHIP: CREATING A NEW LITERACY

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

Stefaan G. Verhulst · Follow

Published in Data Stewards Netw

MATCHING DEMAND & SUPPLY

DEMAND SIDE

Demand side actors seek data to **understand the situation, identify cause and effect, make predictions,** and **solve public problems.**

SUPPLY SIDE

Supply side actors often have access to vast stores of **siloed data.** These sources can, when used responsibly, answer critical questions.

SCOPING DEMAND & SUPPLY

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK

Monitoring and assessing the value, potential, and risk of all data held within an organization.

What's the **question** we seek to answer?

DEMAND SIDE

What's the minimum **viable data** needed to answer the question? And what is available?

Who needs to know what?

SUPPLY SIDE

What **data** do we have that can answer the question?

What is the **value proposition** for us of sharing our data?

What's the **quality** and **interoperability** of our data?

ASSESSING THE RISKS

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK

Monitoring and assessing the value, potential, and risk of all data held within an organization.

What's the **quality** of the data? Is there bias in the data?

What's the **reputation** of the data holder?

DEMAND SIDE

Was the data acquired through **appropriate and ethical** means?

What may happen following a **breach**? Who may be harmed?

SUPPLY SIDE

Is it **legal** to share our data? Do we need consent for re-use?

How will **insights** be used?

Are there **confidentiality** concerns?

Is our data **sensitive**? How do we anonymize our data?

SCOPING OPERATIONAL MODELS

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK

Monitoring and assessing the value, potential, and risk of all data held within an organization.

What are our **requirements** for data access?

What is the data **methodology**?

API? Data pool or lake? **Secure Sandboxes**? Bi-lateral or multilateral?

DEMAND SIDE

What is **fit for purpose?**

SUPPLY SIDE

How can we provide **access** to our data?

What are the protocols for **secure transfer** of data insights?

How **open** is our data?

Is our data **re-usable** and how can we share it?

DEVELOPING PARTNERSHIPS

PARTNERSHIP AND COMMUNITY ENGAGEMENT

Proactively and responsively reaching out to and vetting potential partners.

How do we **engage with communities**?

DEMAND SIDE

Which **organizations** might have valuable data?

Who else should be part of the data collaborative?

SUPPLY SIDE

How do we develop a **data sharing agreement**?

What other **beneficiaries** need to be considered?

Who will **touch the data**? How do we vet stakeholders?

Are the partners **trustworthy**?

INTERNAL COORDINATION

INTERNAL COORDINATION AND STAFF ENGAGEMENT

Securing internal coordination and sign-off from various company actors.

Who do I need to **sign off**?

DEMAND SIDE

Do we have any **internal expertise**?

Who **else is doing what in the data space**?

SUPPLY SIDE

Who do I need to **sign off**?

Who can **lead** our partnership efforts?

Is there any **internal expertise** needed to make data re-usable?

ASSESSING FINANCIAL SUSTAINABILITY

NURTURE DATA COLLABORATIVES TO SUSTAINABILITY

Gathering the needed resources and support so as to ensure broad and long-term impact.

How do we raise **funds** to secure access and cover the costs?

DEMAND SIDE

How can we **sustain access** for the long term?

Do we have **adequate resources** to support this project?

SUPPLY SIDE

What's the **business case**?

What are the long term **benefits** of this project?

How much will this project **cost**?

MAKING AN IMPACT

DISSEMINATION AND COMMUNICATION OF FINDINGS

Raising awareness, disseminating findings and communicating outcomes from data collaboratives.

Can these **findings** inform other projects or future collaboratives?

Can these **outcomes** help or harm our business practices? (Vetting of insights)

DEMAND SIDE

SUPPLY SIDE

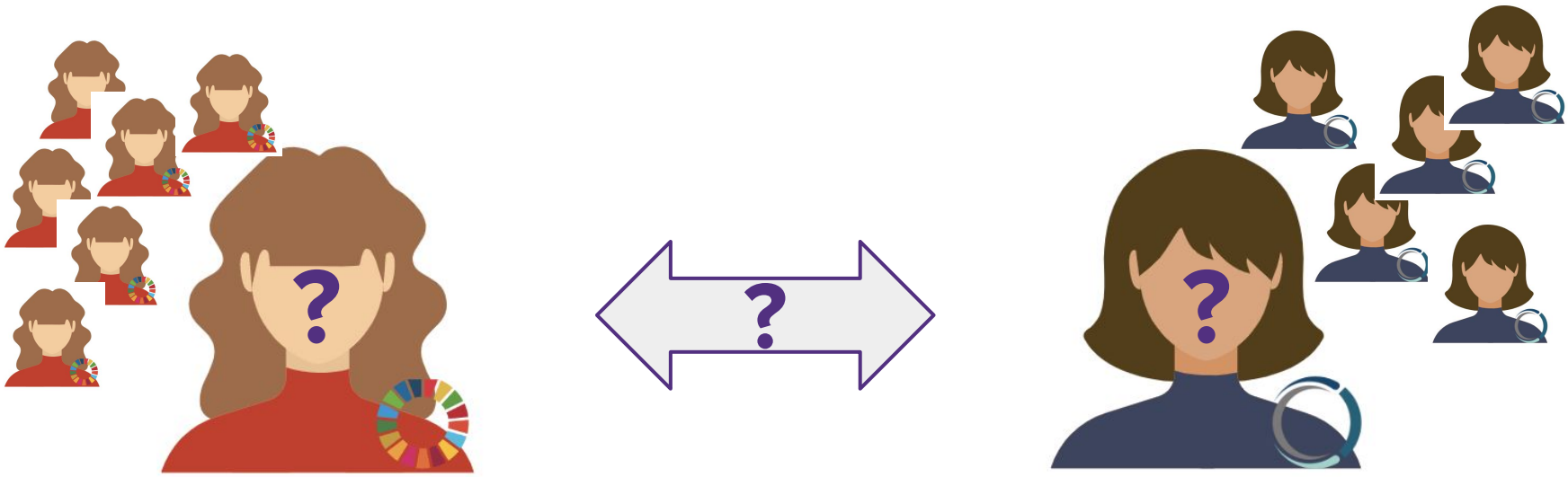
Who needs to be informed to **act** upon the insights?

How can we effectively **communicate** our findings to maximize impact?

Should we **sunset** this initiative?

How can we effectively **communicate** our findings to maximize impact?

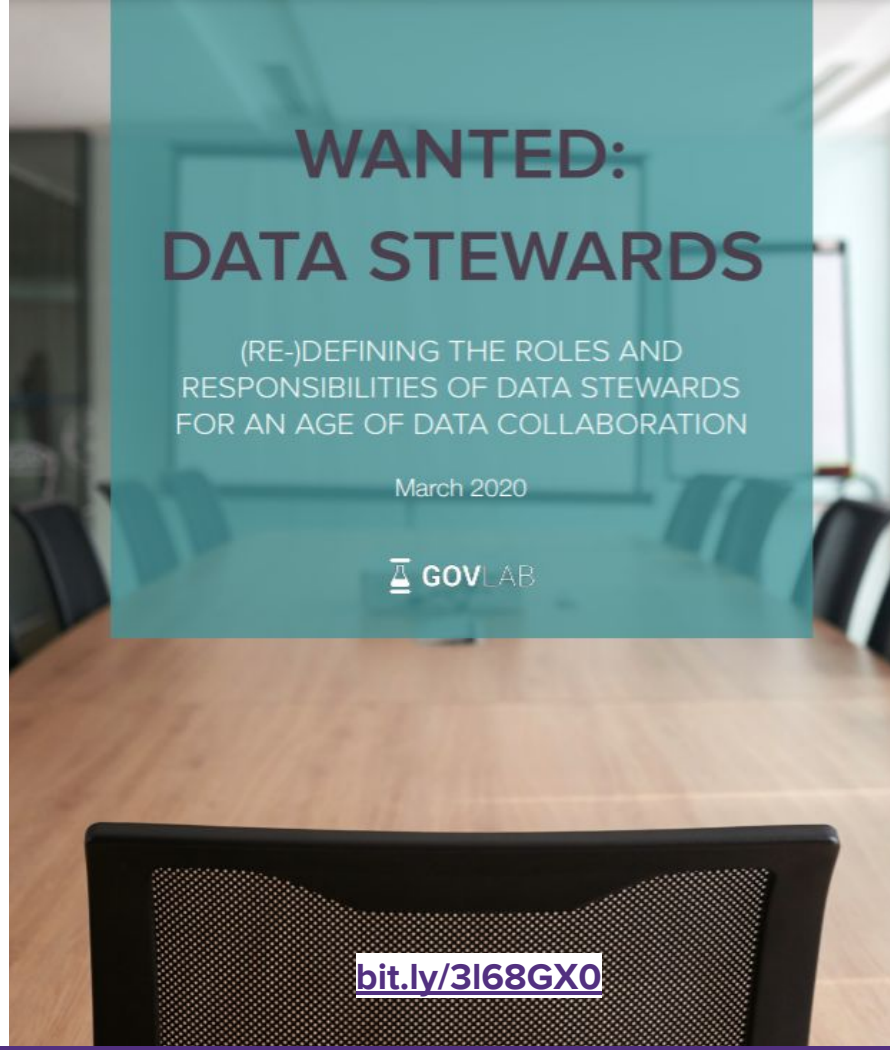
TODAY'S SITUATION



Most data collaborative efforts fall flat because of the lack of a designated data stewardship function.

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.



RE-IMAGINING DATA STEWARDSHIP

Data Stewardship within a Scientific and Library Context

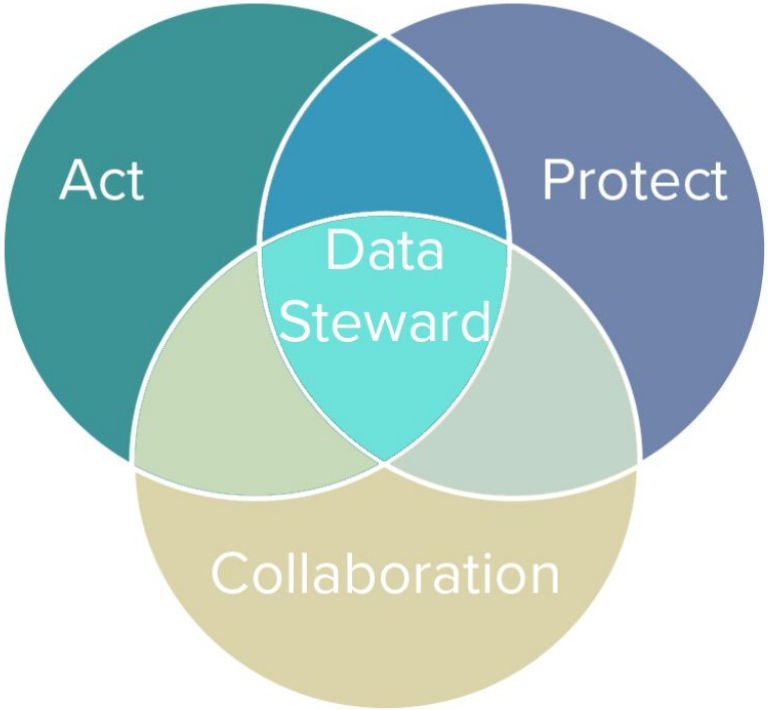
Data Stewardship within a Corporate Data Governance Context

Integrity of the Data

Security of the Data

Re-Use of the Data?

RE-IMAGINING DATA STEWARDSHIP

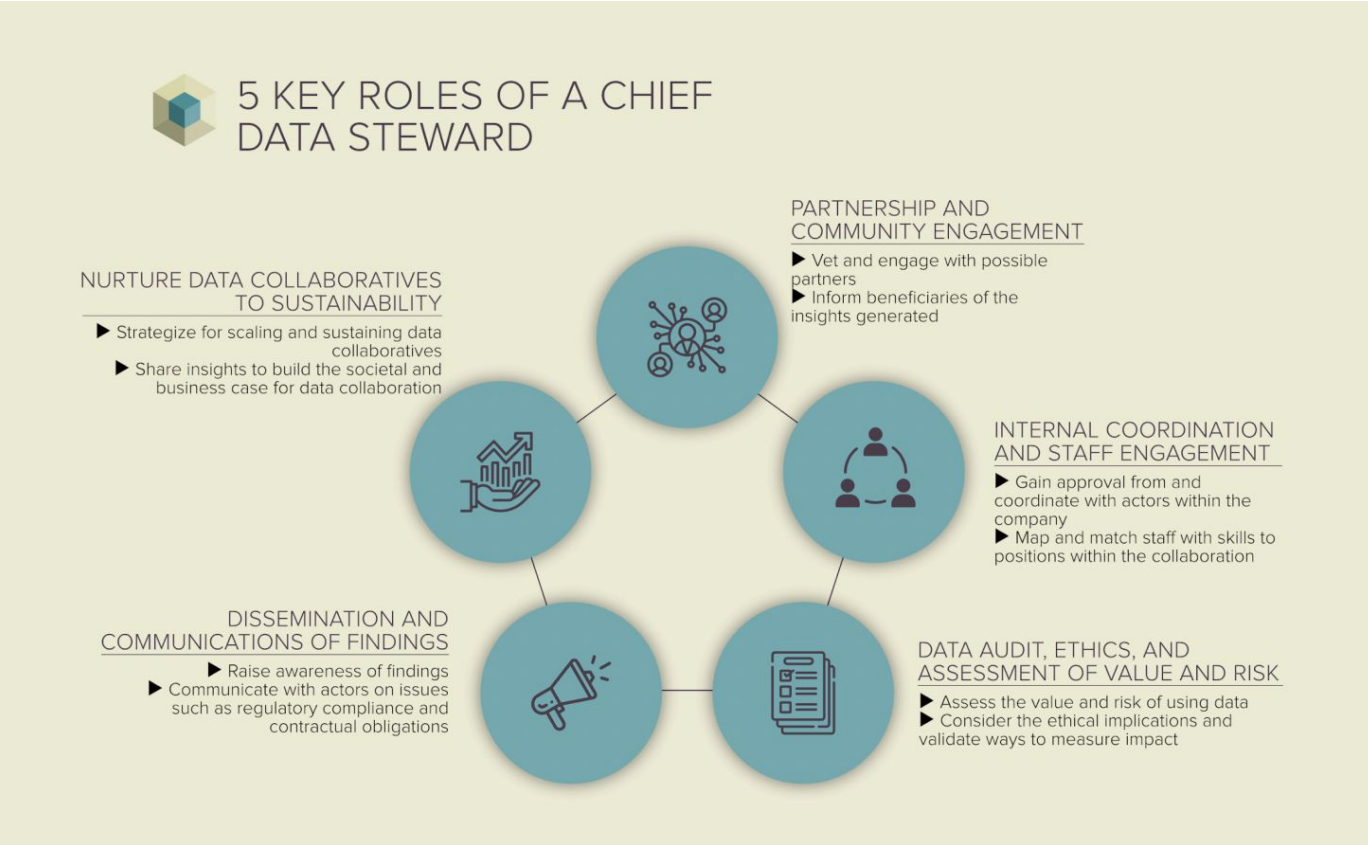


Collaborate:
Working with others to unlock the inherent value of data when it serves the public good.

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

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

THE CORE FUNCTIONS OF A DATA STEWARD



DATA AUDIT, ASSESSMENT & GOVERNANCE

Determining and assessing the value, potential, and risk of data held and needed within an organization

“Stewarding Data Assets for and in the Public Interest”

- Help Formulate and Determine **priority questions** (vis-à-vis Value Proposition or Problem Definition)
- **SCOPING and ITERATING:** Assess “minimum viable” data needed vis-a-vis the questions at hand
- Identify and document data assets
- Consider the ethical and fundamental rights implications and other **risks** of using (or not using) data
- Help establish operational, technical and governance models that are **“fit for purpose”**
- Validate ways to measure impact

PARTNERSHIP AND COMMUNITY ENGAGEMENT

Proactively and responsively reaching out to and vetting potential partners or users.

“Stewarding Relationships”

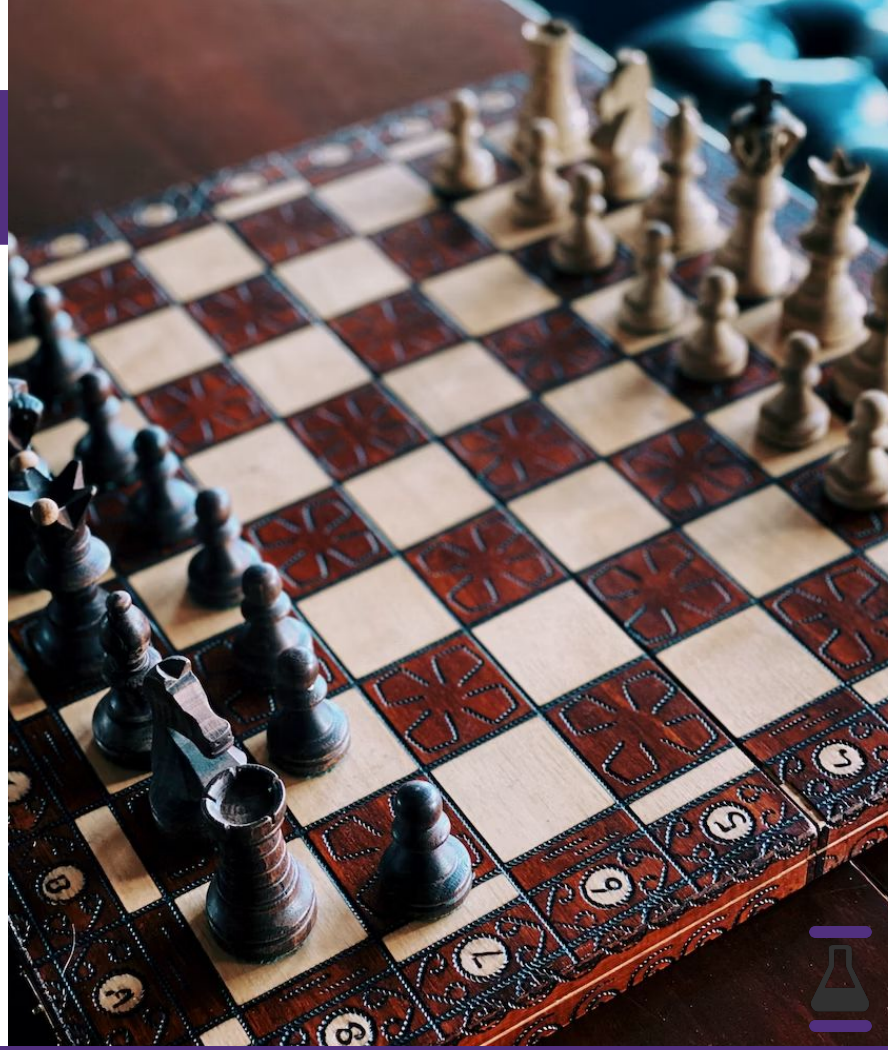
- **EXTERNAL RELATIONS:** Be the point of contact regarding re-use of data
- Identify, map, vet and engage with possible relations, partners and other stakeholders
- **USER-DRIVEN DESIGN:** engage users of data products and insights
- Help establish the **“SOCIAL LICENSE”** of re-using data through deliberation and community engagement
- Establish data agreements and other contractual relationships

INTERNAL COORDINATION AND DATA OPS

Securing internal coordination and establish data operations

“Stewarding Internal Resources, Expertise and Authorities”

- **INTERNAL RELATIONS:** Gain approval from and coordinate with actors within the organization;
- Ensure all internal stakeholders and organizational leadership are informed and aligned.
- **DATA OPS:** Map and match Internal resources, expertise and skills needed to enable data collaboration



NURTURE DATA COLLABORATIVES TO SUSTAINABILITY

Gathering the needed resources and support so as to ensure broad and long-term impact.

“Stewarding Sustainability”

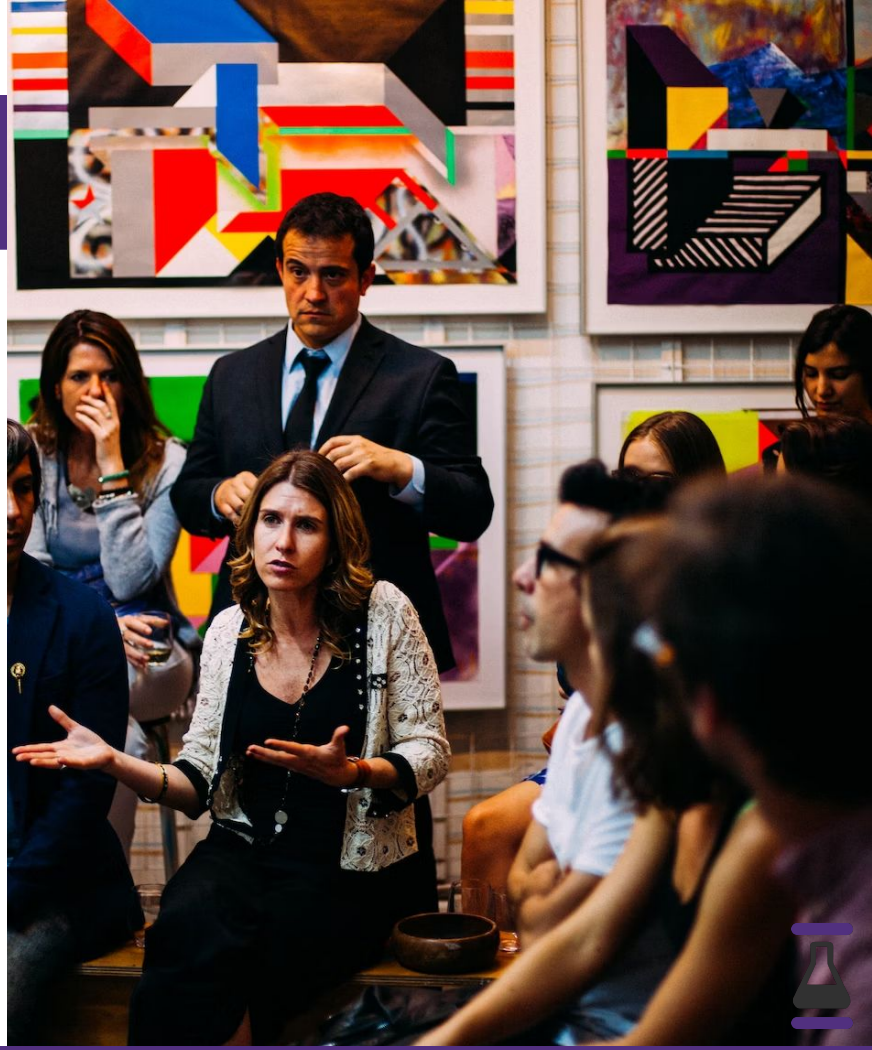
- **INSTITUTIONALIZE DATA INNOVATION:** Make re-use of data systematic (and institutional)
- **DEVELOP THE BUSINESS CASE:** Strategize for scaling and sustaining of data innovation.
- **EVALUATION:** Measure impact and share insights to build the societal and business case for data collaboration

DISSEMINATION AND COMMUNICATION OF FINDINGS

Raising awareness, disseminating findings and communicating outcomes from data collaboratives

“Stewarding Insights”

- **COMMUNICATIONS:** Raise awareness of insights with users, partners, government and other stakeholders
- Enable the translation of data intelligence into decision intelligence
- Communicate with actors on issues such as regulatory compliance and contractual obligations



The Data Stewardship Canvas

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

- What is the problem you seek to solve?
- Do you need to scope out the domain using a topic map?
- Are there certain issues you ought to prioritize based on their need, externalities and feasibility?
- What is the guiding question leading this project?

In the Toolkit:

Problem Definition Tool
R-Search Methodology
Open Data Demand Assessment and Segmentation Methodology

2. Defining the Supply of Data

- What are the expertise and capacity needs for this project?
- What is the minimal amount of data needed to make progress towards answering the question?
- What are the different data sources available for this project?

In the Toolkit:

RD4C Data Ecosystem Mapping Tool
The Periodic Table of Open Data: A User's Guide

3. Making a Value Proposition

- What is this project's value to society?
- What is the return on investment of this project?
- Do the benefits of this project outweigh the costs?

In the Toolkit:

A User's Guide to the 9Rs Framework.
Cost-Benefit Analysis: Data Collaboration

4. Assessing the Risk

- What are the risks of this project across the Data Lifecycle?
- What are the potential externalities (including environmental externalities) of this project?

In this Toolkit:

RD4C Opportunity and Risk Diagnostic
RD4C 22 Questions Audit Tool

5. Matching Demand & Supply: Operational Models

- How is the data going to flow between the project partners?
- What does a fit-for-purpose collaborative model look like?

In the Toolkit:

Data Collaboratives Canvas
RD4C Decision Provenance Mapping Tool

6. Matching Demand & Supply: Governance

- What are the 4 Ps of data governance for this project?
- Who is going to govern this project and how?

In the Toolkit:

Data Responsibility Journey
Contractual Wheel of Data Collaboration

7. Matching Demand & Supply: Tech Infrastructure

- What data standards will improve the interoperability of the data?
- How can the data be handled to balance privacy with efficiency?
- Who can access and re-use the data?

In the Toolkit:

Data Tagging Criteria and Exercise

8. Using Data Responsibly

- What are the ethical implications of this project?
- Do you need to establish a social license for this project?
- How can you assess and mitigate the environmental impact of your project?

In the Toolkit:

Data Responsibility Journey

9. Measuring Impact

- How will you capture the impact and success of this project?
- How will you know when to end this project?

In the Toolkit:

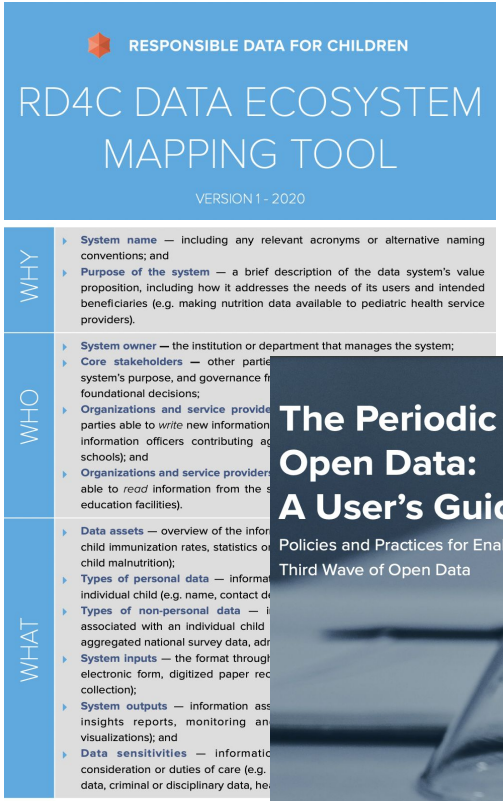
Building a Logic Model to Assess Impact

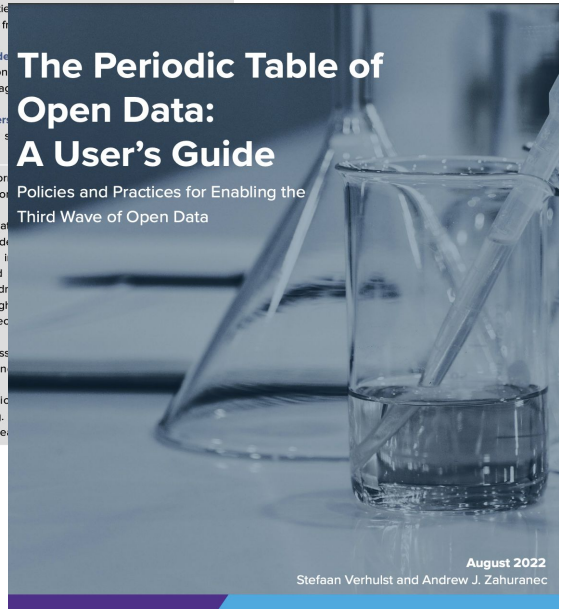
DEFINING THE DEMAND FOR DATA

- What is the problem you seek to solve?
- Do you need to scope out the domain using a topic map?
- Are there certain issues you ought to prioritize based on their need, externalities and feasibility?
- What is the guiding question leading this project?

DEFINING THE SUPPLY FOR DATA

- What are the expertise and capacity needs for this project?
- What is the minimal amount of data needed to make progress towards answering the question?
- What are the different data sources available for this project?





MAKING A VALUE PROPOSITION

- What is this project's value to society?
- What is the return on investment (ROI) of this project?
- Do the benefits of this project outweigh the costs?

Oper

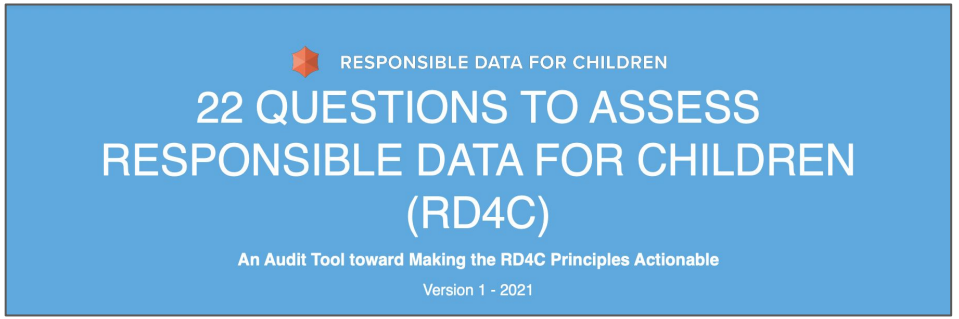
GOVLAB

Cost-Benefit Analysis: Data Collaboration

Stakeholder	Direct Benefits		Direct Costs	
	Recurrent	One-off	Recurrent	One-off
Data Holders (private and public sector)	• VALUE Improved value of own data sets and access to research, analysis methods and models previously not available, adhering to the principle of embracing expertise and efficiency • RISKS Removal of legal and reputational risks (access model by public entities agreed and validated) • REPUTATION and TRUST Reputational benefits and increased trust by publicly demonstrating efforts to create public value by enabling responsible data reuse		• PROCESSING Costs of normalization and making datasets available for reuse (including in absence of compensation depending on the use-case). • PARTNERSHIP Cost associated with formalising partnership – including contractual arrangements (resulting from more such arrangements)	• INVENTORY Costs of cataloguing and identifying data that can be valuable for public interest purposes • SETUP GOVERNANCE new Data Steward position, developing internal data governance approach to comply with relevant laws and regulations

ASSESSING THE RISK

- What are the risks of this project across the Data Lifecycle?
- What are the potential externalities (including environmental externalities) of this project?

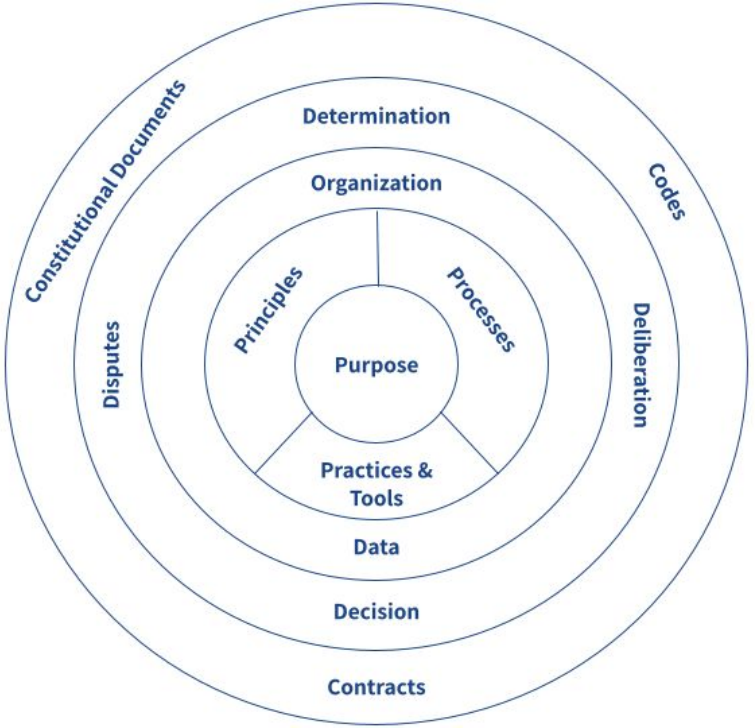


MATCHING DEMAND & SUPPLY: OPERATIONAL MODELS

- How is the data going to flow between the project partners?
- What does a fit-for-purpose collaborative model look like for this project?

MATCHING DEMAND & SUPPLY: GOVERNANCE

- What are the 4 Ps of data governance for this project?
- Who is going to govern this project and how?



The 4 Ps of Data Governance by Stefaan G. Verhulst and Andrew Young

MATCHING DEMAND & SUPPLY: TECH INFRASTRUCTURE

- What data standards will improve the interoperability of the data?
- How can the data be handled to balance privacy with efficiency?
- Who can access and (re-)use the data?

Draft 1.0 - Data Tagging Criteria and Exercise, The GovLab September 4, 2020

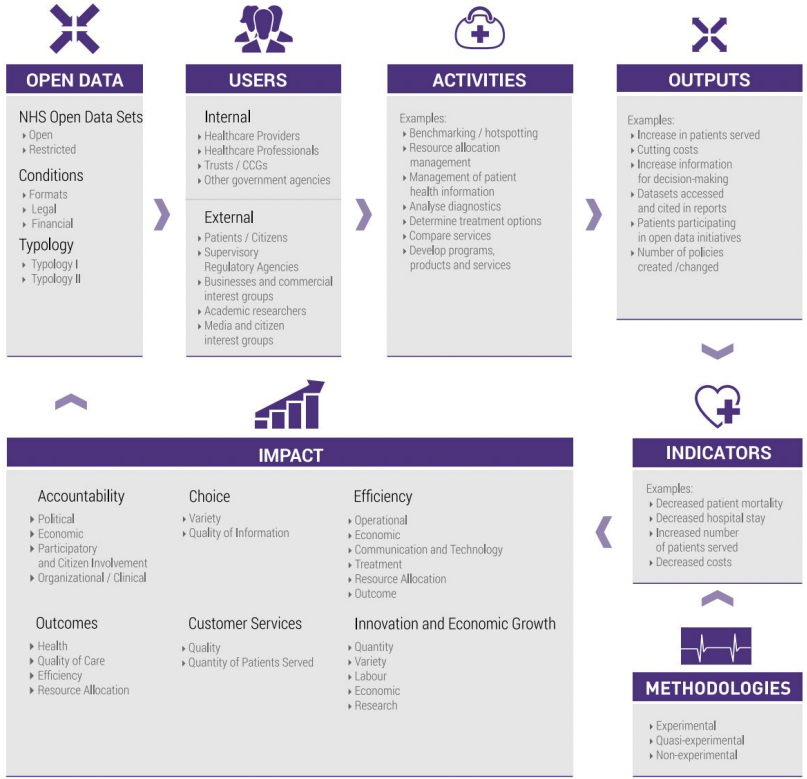
Data Life Cycle Questions	Release Risk Factors	High/Low-Risk Tagging	Open vs Closed? (Spectrum of Conditionality)
How was the data acquired or collected?	Was Consent Obtained (for Reuse)? Data Lineage? Obtained Ethical Clearance through Review Board?		
What data rights are associated with the data?	Licensing Regime? Ownership Expectations? Chain of Trust?		
Are there laws and regulations that need to be complied with?	Regulatory compliance? Cross-jurisdictional considerations?		
How is the data currently stored and processed?	Security? Ease of Access? Auditability?		
Are there any personal, demographic and/or enterprise sensitivities within the data?	Fidelity and Sensitivity? Personal and demographic identifiability (Direct/indirect)? De-identification?		

USING DATA RESPONSIBLY

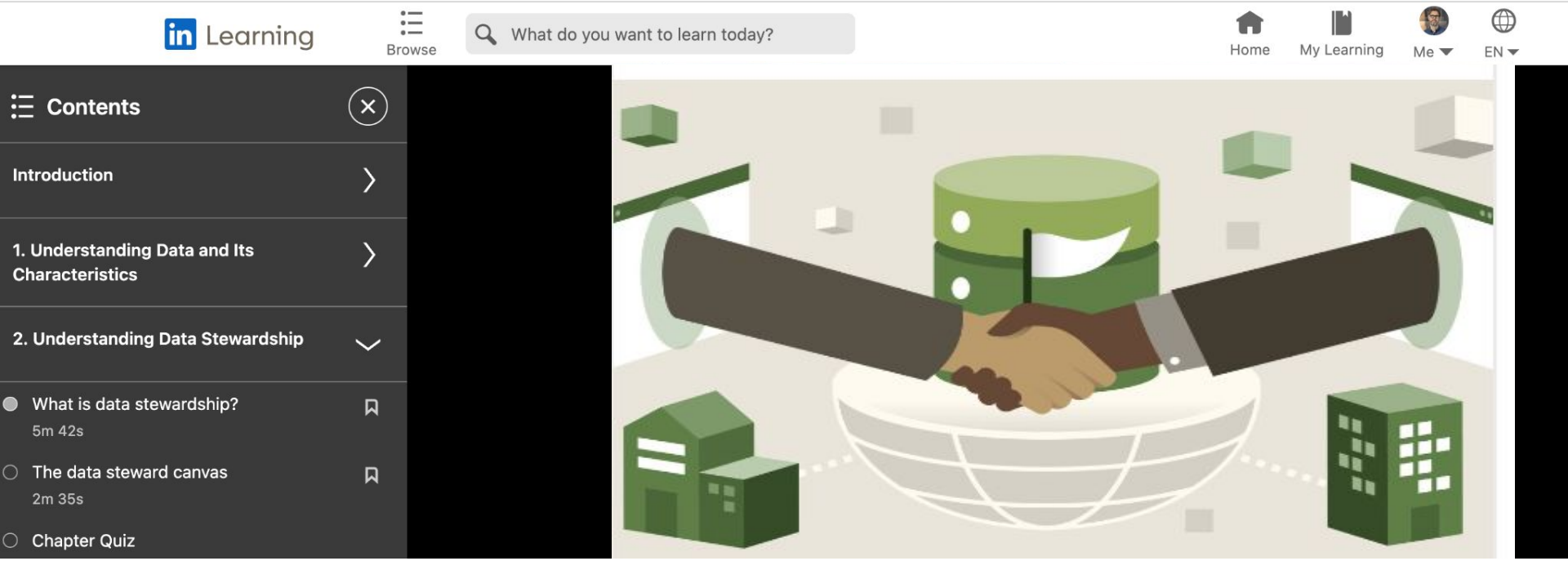
- What are the ethical implications of this project?
- Do you need to establish a social license for this project?
- How can you assess and mitigate the environmental impact of your project?

MEASURING IMPACT

- How will you capture the impact and success of this project?
- How will you know when to sunset this initiative?

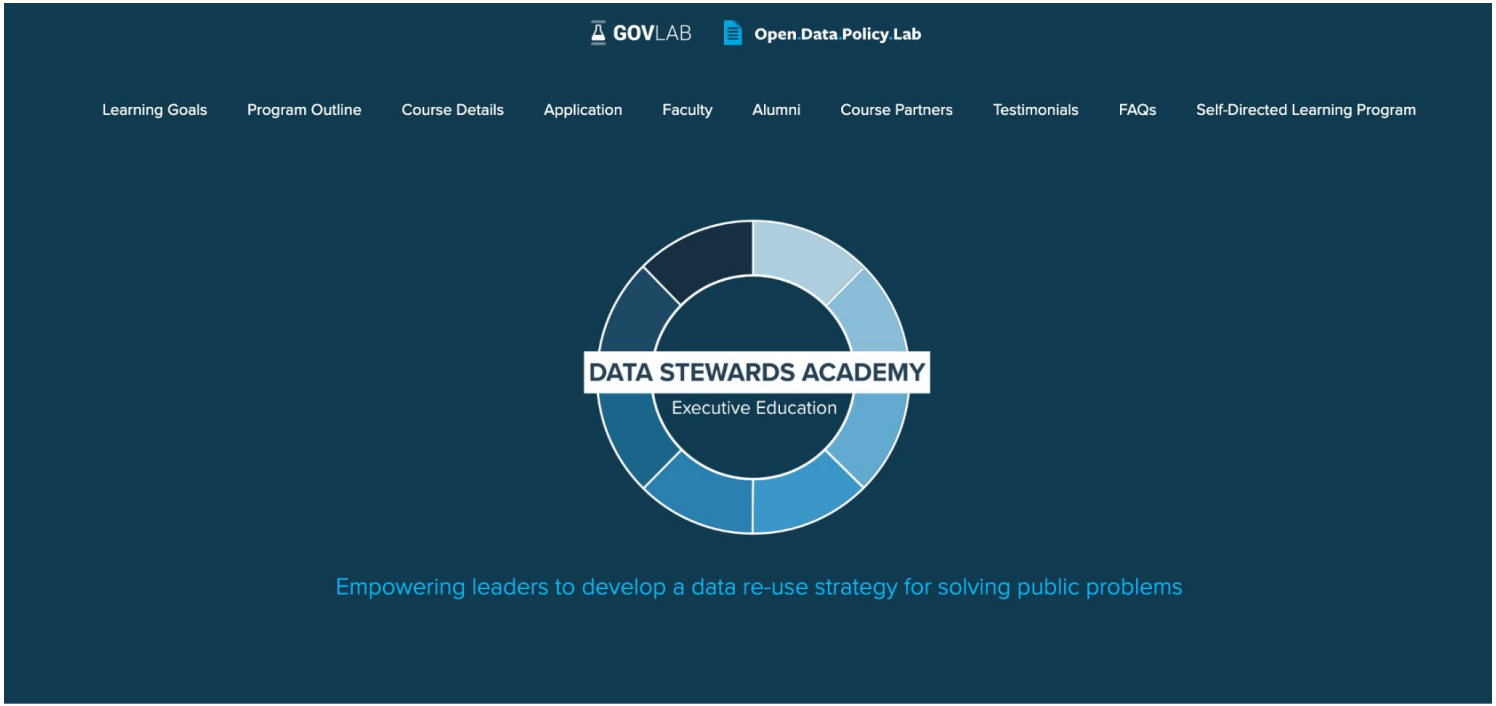


Global Data Stewardship



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

Data Stewards Academy



https://course.opendatapolicylab.org/


THEGOVLAB