GLOBAL PULSE

UNDP

Empowered lives. Resilient nations.

A GUIDE TO DATA INNOVATION FOR DEVELOPMENT

FROM IDEA TO PROOF-OF-CONCEPT

Version 1, December 2016

You don't have to be a data scientist **to innovate with data.**

Any development practitioner interested in data innovation can benefit from this guide. *You do not need expertise in data science to integrate data into your projects! If you are a forward-looking innovator excited about improving your work and the work of your institution and partners, this guide can help.*

WHAT IS DATA INNOVATION?

Data innovation is the use of new or non-traditional data sources and methods to gain a more nuanced understanding of development challenges.

Data innovation often combines non-traditional with traditional sources of data, such as household surveys, to reframe issues and shed new light on seemingly intractable problems. New, or non-traditional data sources may include digital data derived from social media, web content, transaction data, GPS devices (see pp 19 for more). Because combining data sources can provide more complete, more timely, and/or more granular information about an issue, data innovation can open opportunities for more cost- effective interventions, as well as provide entirely new insights that may have been overlooked through traditional approaches.

CASE STUDY:

Improving late-night transportation *through data innovation*

Low-income workers in Seoul needed a new transportation option for commuting late at night. There was no night bus service, and taxis were expensive and hard to find. To establish a new night bus route, city officials analysed aggregated and anonymized mobile phone usage patterns to understand the most common points of departure and destination for travelers. This insight allowed officials to efficiently create a targeted "Owl Bus" route map, which best served late-night travelers.

For more information: http://www.citiesalliance.org/node/5063

WHY DO DEVELOPMENT PRACTITIONERS NEED DATA INNOVATION?

Governments and development agencies around the world are increasingly moving away from old models of problem-solving, and searching for new, more networked models of resilience.

Data innovation is a vital element of this effort. Identifying and integrating faster, more detailed insights into development programme planning processes can lead to better-targeted responses and more efficient resource allocation. Data innovation is also part of reaching the Sustainable Development Goals (SDGs). Effective data collection, analysis, and monitoring can help policymakers to course-correct programmes and policies more quickly, leading to cost efficiencies and greater returns on investments, as well as empower communities ultimately helping to achieve the goals.

HOW CAN THIS GUIDE HELP?

This guide will provide practical guidance in designing a data innovation project, as well as tools for recruiting allies both within and outside of your organization.

It does not include specifics about data analysis; you'll collaborate with a Data Expert for those. Instead, this guide offers an understanding of the scope and scale of data innovation projects, and guidance on asking the right questions from the beginning.

This guide covers the project design phase **from the earliest hint of a data innovation idea through the creation of a proof-of-concept**.

This project design phase may take as long as the actual solution development process itself, because it requires thoughtful coordination among various stakeholders. But with this groundwork firmly in place, data innovation has a better chance of success.

ABOUT THIS PUBLICATION

Recognizing the potential of data innovation, the **United Nations Development Programme (UNDP)** innovation teams at the Regional Centre for Europe and the Commonwealth of Independent States (ECIS) and the Regional Hub for Arab States embarked on a "big data for development exploration journey," seeking to harness new sources of data to improve services or programme implementation on topics including disaster risk management, improved welfare, migration, and poverty reduction. As part of this exercise, six UNDP Offices from the two regions developed data innovation proof-of-concept projects with governmental and civil society partners, to address urgent local challenges and gain hands-on experience in data innovation.

UN Global Pulse provided technical guidance, coaching and tools to the project managers for the duration of the exercise, and the cohort shared advice and challenges at in-person workshops. This guide is based on these experiences, and on the methodology created by Global Pulse to design and implement data innovation projects across the UN system.

The joint exercise showed how alternative sources of data can and should play a role in pursuing development outcomes and, as such, hold great promise for fulfilling the **Sustainable Development Goals**—both from the perspective of pursuing the outcomes as well as enabling (close to) real-time monitoring and evaluation.

This publication is a how-to resource for UNDP staff around the world, and is designed to help further data innovation efforts within the UN and throughout the development field.

ABOUT THE UNITED NATIONS DEVELOPMENT PROGRAMME (UNDP)

UNDP is the UN's global development network, advocating for change and connecting countries to knowledge, experience, and resources to help people build a better life. UNDP is on the ground in 177 countries and territories, supporting their own solutions to development challenges and developing national and local capacities that will help them achieve human development and the Sustainable Development Goals.

Learn more online at www.undp.org

ABOUT UN GLOBAL PULSE

Global Pulse is an innovation initiative of the United Nations Secretary-General. Global Pulse is working to promote awareness of the opportunities big data presents for sustainable development and humanitarian action, forge public-private data sharing partnerships, generate high-impact analytical tools and approaches through its network of Pulse Labs, and drive broad adoption of useful innovations across the UN System.

Learn more at: www.unglobalpulse.org

HOW TO READ THIS GUIDE

This practical guide will help you lay the groundwork for integrating data innovation into your projects. If you're working on your first data innovation project, we suggest you go through the guide chronologically to make sure you remember each step and keep your project on track.

Make sure to meet the requirements of Checkpoint A before progressing to Sections Two and Three.

If you have previous experience managing data innovation projects, our suggestion is that you will refer back as you progress on your journey, dipping in and out of specific modules **as they become relevant** to your work.

Table of **Contents**

SECTION ONE
Explore the Problem & System

SECTION TWO
Assemble the Team

SECTION THREE
Create the Work Plan

MODULE 1 ..... **p. 9**
Define the problem you're trying to solve.
Problem Definition Tool ..... ***p. 10***

MODULE 2 ..... **p. 11**
Inventory and understand **the data gaps.**
Data Gaps Tool ..... ***p. 12***

MODULE 3 ..... **p. 13**
Map **all stakeholders**
Stakeholder Mapping Tool ..... ***p. 14***

MODULE 4 ..... **p. 15**
Understand who will use your data innovation results
Data Journey Tool ..... ***p. 17***

MODULE 5 ..... **p. 18**
Identify your **data wish list and define your hypothesis**
Project Concepting Tool ..... ***Annex***

CHECKPOINT A (p.22):
Official permission

MODULE 1 ..... **p. 26**
Account for **Five Key Responsibility Areas**

MODULE 2 ..... **p. 30**
Secure support from Data Holder(s)

MODULE 3 ..... **p. 32**
Find your **Data Expert**

MODULE 1 ..... **p. 36**
Develop a **concept note**

MODULE 2 ..... **p. 38**
Ensure data protection and privacy

MODULE 3 ..... **p. 41**
Plan how to **measure and share your results**

CHECKPOINT B (p. 45):
Create a Proof-Of-Concept

ANNEX ..... **p. 48**
Full library of tools for use

SECTION 1:
EXPLORE THE PROBLEM & SYSTEM

Every successful data innovation project starts with **a well-defined problem** (such as a development challenge or a policy issue) as well as a deep **understanding of the institutional system** around the problem.

With many fascinating data sources available, innovators in both the public and private sectors often get carried away collecting or diving into data sources that don't really matter to the actual challenges they face. Because they start with the data solution (instead of the problem), such projects are often unsuccessful.

This section will help you avoid this common mistake, with **five tools** for reflecting on what decisions, actions, and changes you hope to enable. It is designed to encourage rigor and spark inspiration, and organized based on common exploration processes, but you may find that specific tools are more useful individually, or at different times in your own journey. In some cases, completing one tool will help you flesh out another. Hop around the tools of this section, print them out, share them with colleagues, and make them your own.

OBJECTIVES

- Identify your data gaps
- Create a stakeholder power map
- Establish your Policy Research Question

MODULE 1:
DEFINE THE PROBLEM YOU'RE TRYING TO SOLVE.

What is the specific development challenge or policy problem you hope to address through data innovation? Articulating the problem clearly will establish important boundaries, save time and resources, and maximize project returns.

Use the **Problem Definition Tool** *(sample on the following page)* to identify the key challenge you are trying to address, and to begin thinking about what data already exists on the issue. The tool will push you to think specifically about how better measurement or more complete information can help solve this problem. What new decisions will be made? What responses can be faster?

As you complete the **Problem Definition Tool**, you should desk research previous projects in the same problem area, or around similar data sources. This can spark new ideas for potential data sources for a solution; it may also serve as one indicator of your project's chance of success.

*You can find a printable version of the **Problem Definition Tool** in the Annex.*

CASE STUDY:

Laying the groundwork for data innovation in Armenia

Tourism has great potential for growth in Armenia. In supporting the sector as a national and local economic development vehicle, the government could benefit from more precise empirical data. Commonly used data sources, such as hotel booking logs and border control records, provide only part of the picture; they can also be time-consuming and expensive to collect.

The UNDP Armenia office saw an opportunity to **help the tourism industry** based on up-to-date tourist preferences by analyzing the number of roaming foreign SIM cards in use.

The team and a local telecom operator worked together to analyse aggregated records from two tourist areas for a two-month pilot, testing whether insights about tourists' countries of origin and travel patterns within Armenia could be reached. The proof-of-concept was successful, and the team is now preparing to analyse a dataset based on a full tourist season, to share with both government decision-makers and local businesses to understand and adapt to shifting trends.

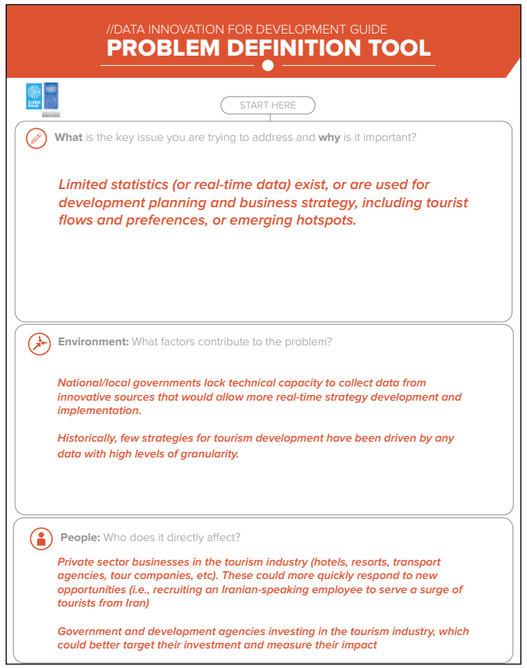
This success was possible because of careful groundwork and planning, following a process similar to that detailed in this section. Throughout this section, examples from the Armenia team's work will illustrate specific elements of the process, to help you develop a similarly thorough foundation to set your project up for success.

For more information reach out to Max Perry-Wilson (max.perry-wilson@undp.org) and Marina Mikhitaryan (marina.mkhitaryan@undp.org) from the UNDP Armenia office.

EXAMPLE: PROBLEM DEFINITION TOOL

The Problem Definition Tool completed by the Armenia team began with the fact that tourism is a key growth sector, with great potential for economic development in a country with abundant natural and cultural assets; tourism had also been identified as a priority in the national development strategy, but with very little actionable data that would allow the Government and tourism sector to design timely new services and products to capture more of a market share during the season.

The problem definition tool showed that additional information sources could help specific actors further their response to the problem. See case study on pp 8-9 for more on this project.



//DATA INNOVATION FOR DEVELOPMENT GUIDE
PROBLEM DEFINITION TOOL

START HERE

What is the key issue you are trying to address and **why** is it important?

Limited statistics (or real-time data) exist, or are used for development planning and business strategy, including tourist flows and preferences, or emerging hotspots.

Environment: What factors contribute to the problem?

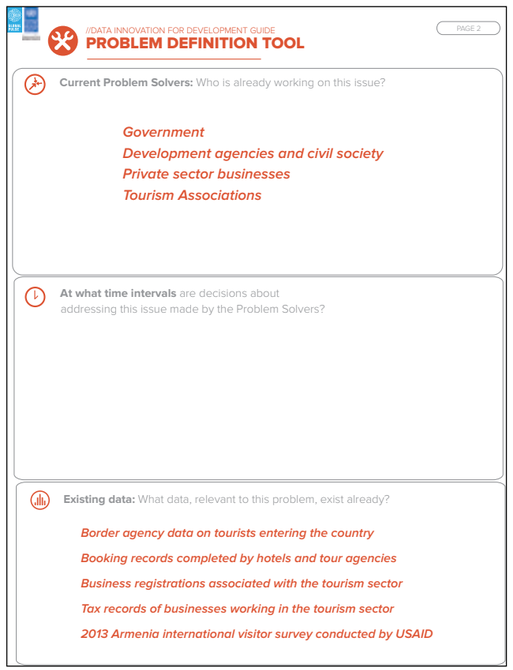
National/local governments lack technical capacity to collect data from innovative sources that would allow more real-time strategy development and implementation.

Historically, few strategies for tourism development have been driven by any data with high levels of granularity.

People: Who does it directly affect?

Private sector businesses in the tourism industry (hotels, resorts, transport agencies, tour companies, etc). These could more quickly respond to new opportunities (i.e., recruiting an Iranian-speaking employee to serve a surge of tourists from Iran)

Government and development agencies investing in the tourism industry, which could better target their investment and measure their impact



//DATA INNOVATION FOR DEVELOPMENT GUIDE
PROBLEM DEFINITION TOOL

PAGE 2

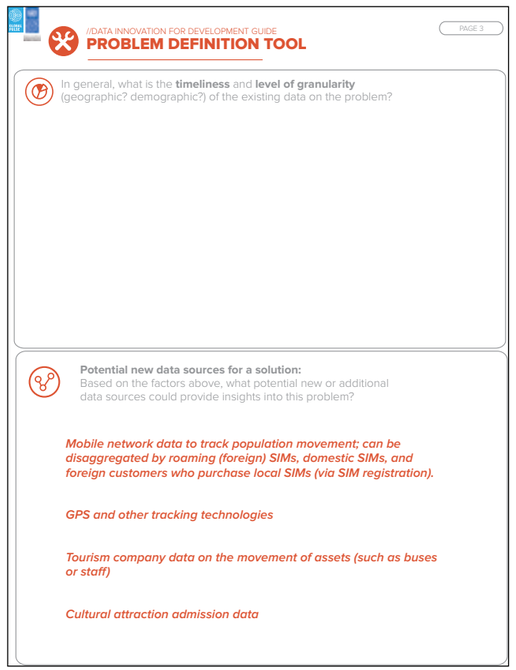
Current Problem Solvers: Who is already working on this issue?

Government
Development agencies and civil society
Private sector businesses
Tourism Associations

At what time intervals are decisions about addressing this issue made by the Problem Solvers?

Existing data: What data, relevant to this problem, exist already?

Border agency data on tourists entering the country
Booking records completed by hotels and tour agencies
Business registrations associated with the tourism sector
Tax records of businesses working in the tourism sector
2013 Armenia international visitor survey conducted by USAID



//DATA INNOVATION FOR DEVELOPMENT GUIDE
PROBLEM DEFINITION TOOL

PAGE 3

In general, what is the **timeliness** and **level of granularity** (geographic? demographic?) of the existing data on the problem?

Potential new data sources for a solution:
Based on the factors above, what potential new or additional data sources could provide insights into this problem?

Mobile network data to track population movement; can be disaggregated by roaming (foreign) SIMs, domestic SIMs, and foreign customers who purchase local SIMs (via SIM registration).

GPS and other tracking technologies

Tourism company data on the movement of assets (such as buses or staff)

Cultural attraction admission data

Using your completed **Problem Definition Tool,** establish a first iteration of one or several of your key statements: your **Policy Research Question(s).**

Your **Policy Research Question** is a short question that connects the data sources you just identified with the problem itself. Can this data help? It is phrased as a "yes or no" question:

> **Can we use [X data source(s)] for insights on [Y problem insights]?**

This **Policy Research Question** will be refined *throughout the problem exploration phase and the completion of additional tools*. It will lead to the construction of your Hypothesis: a short statement that establishes what aspect of the data source you will test for insights on specific problem indicators.

SAMPLE POLICY RESEARCH QUESTION

ARMENIA ROAMING MOBILE DATA

Policy Research Question: Can data about "roaming" mobile phone SIM cards provide insights into tourists' preferences about where they travel and plan to go while in Armenia?

MODULE 2:
INVENTORY AND UNDERSTAND THE DATA GAPS

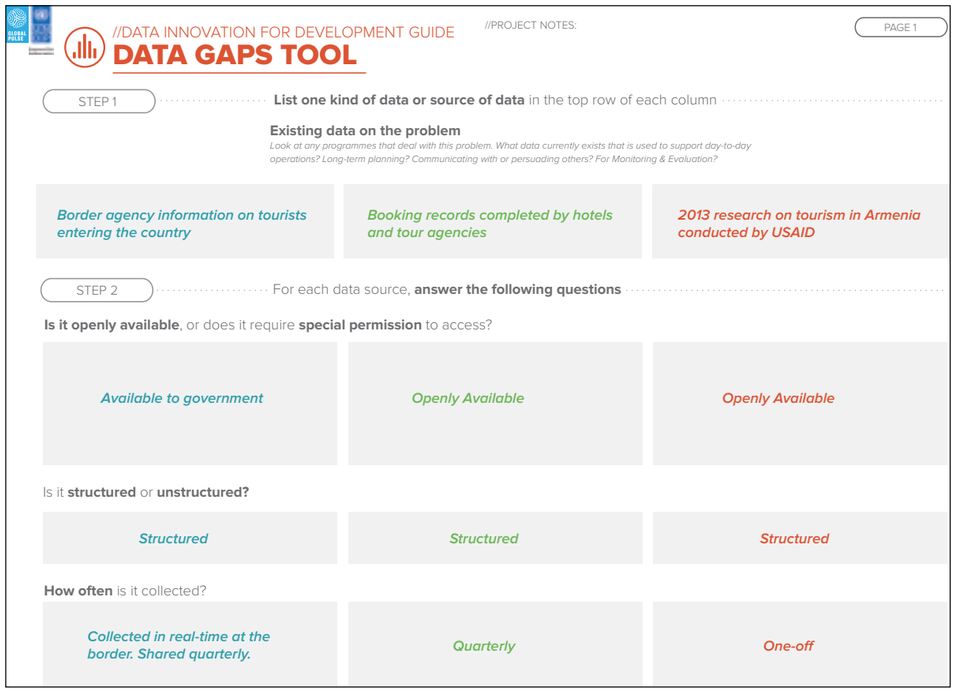
In the last module, you defined the problem you are trying to solve. **Now, it's time to think about what important data (see pp 19) you have (and lack) internally, across both short- and long-term planning cycles.**

Also think about insights you wish you had (perhaps at a different granularity, on a different dimension or more frequently), but cannot get from your current data collection methods. Bridging these "data gaps" will likely be the focus of your data innovation project.

Use the **Data Gaps Tool** to make a complete list of readily available data. The goal of this is to understand fully the data that already exists, including what you rely on for long-term planning (even if it's imperfect). Be sure to capture all data that are easily available within your organization, including what is gathered in reports and stored in databases. Consider quantitative as well as qualitative data, such as small-scale surveys, focus groups, and ethnographic methods like observation and interviews.

Note that one data source is not usually sufficient. A thoroughly completed Data Gaps Tool can highlight the fact that multiple data sources may be necessary to fill all gaps. There is a printable tool in the Annex.

EXAMPLE: DATA GAPS TOOL



//DATA INNOVATION FOR DEVELOPMENT GUIDE
DATA GAPS TOOL

//PROJECT NOTES:

PAGE 1

STEP 1 — **List one kind of data or source of data** in the top row of each column

Existing data on the problem
Look at any programmes that deal with this problem. What data currently exists that is used to support day-to-day operations? Long-term planning? Communicating with or persuading others? For Monitoring & Evaluation?

Border agency information on tourists entering the country	*Booking records completed by hotels and tour agencies*	*2013 research on tourism in Armenia conducted by USAID*

STEP 2 — For each data source, **answer the following questions**

Is it openly available, or does it require **special permission** to access?

Available to government	*Openly Available*	*Openly Available*

Is it **structured** or **unstructured?**

Structured	*Structured*	*Structured*

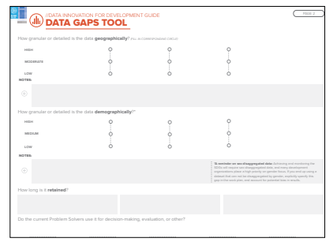
How often is it collected?

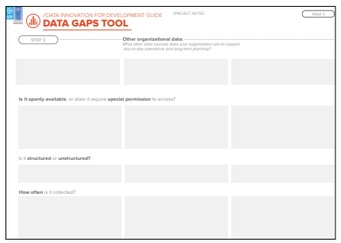
Collected in real-time at the border. Shared quarterly.	*Quarterly*	*One-off*

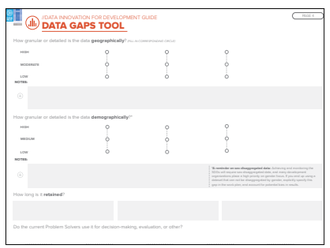
The Data Gaps Tool completed by the Armenia team shows that, while there is a wealth of data about foreign tourist activity available in hotel booking records and other surveys, there was a gap in information about internal tourist flows (particularly in relation to domestic tourists) and about what foreign tourists spend.

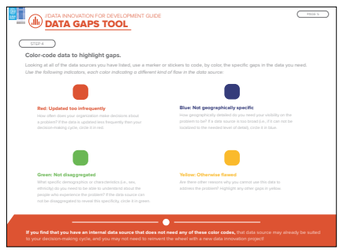
Additionally, data on international tourists were limited, as hotels and tour agencies often only submit data on those tourists paying for services through credit card, which is especially limited outside of Yerevan.

See case study on pp 8-9.









MODULE 3:
MAP ALL STAKEHOLDERS

In addition to the need to focus on the problem to be solved, another common pitfall of data innovation projects is **focusing on theoretical or technical concerns while overlooking the operational pathways necessary for real impact.**

Understanding all actors and their relative influence over the problem will help frame your plans within the existing government, civil society, academic, and other systems responsible for addressing a problem.

Your data innovation project, like all of your work, is part of a much larger system of efforts by governments, NGOs, and communities dealing with the problem. By understanding this broader network of people involved, you can better position your project to secure vital support from powerful actors and to open participation and collaboration to the people who are most affected by it.

THREE TYPES OF ACTORS TO CONSIDER:

- **COMMUNITIES:** The people who experience the problem directly, and interact with Problem Solvers.
- **PROBLEM SOLVERS:** The civil servants, NGO staff, front-line responders, and others on-the-ground.
- **POLICYMAKERS:** The people who have access to resources and control allocation.

Use the **Stakeholder Mapping Tool** to list all of the parties potentially involved in or affected by your data innovation project and organize them according to their motivations, roles, and constraints.

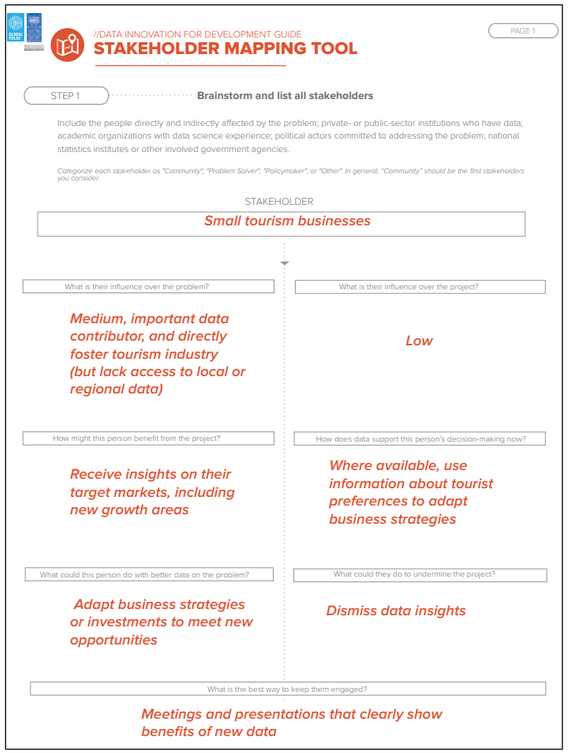
In this step, you should go beyond a default listing of partners you work with. Focus in on the individuals who have leverage and stake in the intervention you are planning, including members of formal groups and organizations as well as issue experts. Note that, while development projects often lump actors according to a simpler “supply vs. demand” categorization framework, this tool encourages you to understand actors according to their influence over the problem at hand.

There will probably be a few partners who have some capability to enact change. With this in mind, the **Stakeholder Mapping Tool** is designed to distill the particular roles of relevant partners in terms of your Policy Research Question.

EXAMPLE: STAKEHOLDER MAPPING TOOL

On the Stakeholder Map completed by the Armenia team, "community" and "Problem Solvers" are the same; tourism businesses will use the new data insights. The map also highlights tourist associations as valuable potential channels for information distribution.

See case study on pp 8-9 for more on this project.



//DATA INNOVATION FOR DEVELOPMENT GUIDE
STAKEHOLDER MAPPING TOOL
PAGE 1

STEP 1 — Brainstorm and list all stakeholders

Include the people directly and indirectly affected by the problem; private- or public-sector institutions who have data; academic organizations with data science experience; political actors committed to addressing the problem; national statistics institutes or other involved government agencies.

Categorize each stakeholder as "Community", "Problem Solver", "Policymaker", or "Other". In general, "Community" should be the first stakeholders you consider.

STAKEHOLDER: *Small tourism businesses*

What is their influence over the problem?	What is their influence over the project?
Medium, important data contributor, and directly foster tourism industry (but lack access to local or regional data)	*Low*
How might this person benefit from the project?	**How does data support this person's decision-making now?**
Receive insights on their target markets, including new growth areas	*Where available, use information about tourist preferences to adapt business strategies*
What could this person do with better data on the problem?	**What could they do to undermine the project?**
Adapt business strategies or investments to meet new opportunities	*Dismiss data insights*

What is the best way to keep them engaged?

Meetings and presentations that clearly show benefits of new data

//DATA INNOVATION FOR DEVELOPMENT GUIDE
STAKEHOLDER MAPPING TOOL
PAGE 5

STEP 2

Plot all stakeholders on the following map

Those in the center of the circle have the greatest influence, those on the perimeters have the least. Include brief notes about strategies for engagement or opportunities for support.

NOTES:

预览已结束，完整报告链接和二维码如下：
https://www.yunbaogao.cn/report/index/report?reportId=5_11960

MODULE 4:
UNDERSTAND WHO WILL USE YOUR DATA INNOVATION RESULTS

On your Stakeholder Map, there are two groups of actors who need to play a central role in the design of your data innovation solution: **The communities who experience the problem directly, and the Problem Solvers who are responsible for addressing it.**

These two groups should be the main focus of the data innovation design. If you tailor the project to their needs and constraints, you have a greater chance that they will actually use your solution effectively.

Take a moment to consider whether your Problem Solvers can be the same people and communities who experience the problem directly (i.e., the "beneficiaries" of your work). Rather than focusing on large government or NGO responders, many data innovation solutions directly empower affected communities to address problems themselves. These kinds of solutions are best practice for building resilience and sustainability.

The **Data Journey Tool** (on pp 17) explores the problem statement from a different perspective: the interactions that generate data points. You should fill out the Data Journey Tool putting yourself in the shoes of the communities you are trying to help - try to identify what data points are generated from that perspective. For example, what data points would a person living with HIV/AIDS be generating? Also think whether the people you are trying to assist and the Problem Solvers are one and the same category. If the Problem Solvers are not the same as the people experiencing the problem, you might want to fill out a second Data Jounrney Tool answering each question from the point of view of the Problem Solver - for example a UN agency or a government official that use a dashboard to implement policies.