****Note to reviewers****

We welcome comments, revisions and suggestions on our draft of this playbook and are keen to iterate and treat this document as a dynamic community artefact.

As we build out the playbook over the next couple weeks, we will also be adding the following elements or components:

- *Stakeholder curated checklists at the end of each 'play'*
- *A select set of case studies or 'methodology' practices as illustrated on p.24*
- *A detailed section outlining our methodology and learnings for this playbook that will inform the focus and approach of our forthcoming work*
- *A set of open questions that this playbook may be limited in addressing, but may be explored in the upcoming focus areas (Health & Urban Governance)*

For any questions or clarifications, feel free to reach out to our team at contact@aapti.in or suha@aapti.in

Thank you!

Fostering Participatory Data Stewardship:

A Playbook **[WIP title]**

aapti institute

September 2022

Table of Contents

EXECUTIVE SUMMARY **6**

HOW TO READ THIS DOCUMENT **8**

INTRODUCTION **10**

Participation – a foundational pillar of data stewardship 11

SECTORAL GUIDE | ENVIRONMENT & SUSTAINABILITY **14**

Challenges with accessing environmental data & consequent impact 15

Existing pathways to environmental justice that can inform participatory and community-oriented governance 16

Foundational Play : Paradigms for community participation in data stewardship **18**

Challenge 1: Participation is structured as a 'one-off' engagement and does not persist throughout the lifecycle of data usage & governance **20**

Strategy 1.1: Identify incentives of data generators (individuals and communities) through consultation to better identify approaches to embed participation ***21***

Strategy 1.2: Remodel the prevailing regulatory landscape for data governance to embed mechanisms for community participation throughout the lifecycle of data usage ***23***

Challenge 2: Data gathering and sharing efforts risk being one-sided and exploitative in the absence of clear incentives that would deliver broad-based public benefit for communities that share data **25**

Strategy 2.1: Tap into existing networks or organisations that possess trust-based and collaborative relationships with communities and identify how value can be jointly delivered or enhanced ***26***

Strategy 2.2: Creating supportive legal and technical infrastructures that prioritise 'participation by design' for meaningful community participation in data stewardship ***28***

Challenge 3: Data gathering efforts although defined as participatory continue to be surface-level and non-diverse as structural barriers (class, gender, ethnicity/race, age, citizenship status), vulnerabilities or capacities (technical, financial and data literacy) are not accounted for **30**

Strategy 3.1: Forge partnerships with existing community-based organisations to solve for issues of representation within data stewardship initiatives ***30***

Strategy 3.2: Locate and empower engaged community members to build bottom-up data-oriented communities and facilitate more diverse onboarding for stewarding organisations ***31***

Play II: Scaling community governance through data stewardship - moving beyond pilots **33**

Challenge 1: Ensuring financial sustainability beyond the pilot phase **34**

Strategy 1.1: Assess value delivery and product accessibility in order to identify suitable business models ***35***

Strategy 1.2: Funder flexibility to understand long-term stewardship goals ***36***

Challenge 3: Limited technical capacity 40

Play II | Sectoral Guide | Environment and Sustainability 42

Challenge 1: Limited accountability and transparency around Environment, Social and Corporate Governance (ESG) reporting and Environmental data sharing by private companies 42

Strategy 1.1: Enhance ESG reporting and identify alternative sources to fill in data gaps 42

According to the Carbon Disclosure Project, a non-profit group that runs an environmental disclosure system, private companies trail public counterparts when it comes to reporting greenhouse gas emissions and climate risks. 42

Challenge 2: Funding challenges for citizen science projects 44

Strategy 2.1: Identify and amplify networks to facilitate funding for citizen science efforts ***44***

Play III: Technical pathways for operationalising community participation in data stewardship 45

Challenge 1: Lack of community participation in the innovation/design process that is hindered due to technical barriers has knock-on effects for community participation in the life cycle of a stewarding initiatives 46

Strategy 1.1: Using technical tools to enable co-designing with the purpose of furthering effective participation ***46***

Challenge 2: The chosen technology and its implicit design bias could exacerbate digital divides and inhibit participation based on lack of capacity and access to such technology 48

Strategy 2.1: Employing tools that make technology and data more accessible ***48***

Challenge 3: Lack of quality data and validation mechanisms can result in misuse and distrust rendering such data unsuitable for research or policymaking 49

Strategy 3.1: Employing data verification mechanisms and maintaining extensive metadata ***49***

Play III | Sectoral Guide | Environment and Sustainability 50

Challenge 1: Furthering of environmental injustice owing to technical barriers that prevent communities from being involved in the design process of citizen science projects lead 51

Strategy 1.1: Co-designing the project for full and effective participation of communities ***51***

Strategy 1.2: Incorporating indicators/tools that ensure responsible inclusion of communities ***52***

Challenge 2: Implicit design bias can create digital divides and barriers to access technology 53

Strategy 2.1: Incorporating best practices and tools that enable responsible inclusion of communities ***53***

Challenge 3: Lack of quality data and validation mechanisms can result in its misuse and distrust that would render such data unsuitable for research or policymaking 55

Strategy 3.1: Maintaining data quality, employing verification mechanisms and contextualising data 55

Strategy 3.2: Involving participants in data collection methods 56

Play IV: Ecosystem enablement – role of the public sector 58

Challenge 1: Funding efforts tend not to focus on community-oriented and community-led projects – funding programs typically incentivize institutional data collection efforts which don't look at the community as an active participant 60

Strategy 1.1: Instituting funding programs with a specific focus on data stewardship efforts 60

Strategy 1.2: Encouraging private entities to fund data stewardship efforts 61

Challenge 2: There is a disconnect between community actions and policy measures – engagement of the public sector at a grass-roots level is often missing, and when present, the information provided by communities is not often reflected in policy measures 62

Strategy 2.1: Recognising citizens as a key stakeholder in data collection that influences policy making 62

Strategy 2.2: Incorporating community generated data and learnings in policy measures 63

Strategy 2.3: Implementing robust and accessible data quality practices and standards 63

Challenge 3: Access to information and data collected by the public sector is very limited – in many cases the data itself is not made available and in cases where it is, access is complicated by the modalities of access 65

Strategy 3.1: Creating an infrastructure that leverages access to public sector data and enables open data exchange between public sector and citizens 65

Challenge 4: Private sector involvement in data governance for good is lacking - Information collected by the private sector is siloed, with either no access, or high access costs, and compliance of private sector with regulations is problematic 66

Strategy 4.1: Incentivise the private sector to share data with the public sector and data stewards 66

Strategy 4.2: Leverage data stewards to ensure accountability in data sharing 67

Play IV | Sectoral Guide | Environment and Sustainability 69

How can the public sector support citizen science efforts? 69

Challenge 1: The public sector's role in funding citizen science efforts 70

Strategy 1.1: Instituting citizen science funding programs 70

Strategy 1.2: Encouraging private entities to fund citizen science efforts 70

Challenge 2: Partnerships and engagement with communities 72

Strategy 2.1: Building tools for community participation 72

Strategy 2.2: Incorporating community generated data and learnings in policy measures 73

Strategy 2.3: Collaborating and co-designing at the planning stage of the citizen science project 73

Strategy 2.4: Contextualising data and maintaining quality data standards 74

Challenge 3: Support citizen science efforts through easier availability of public sector data and transparency 75

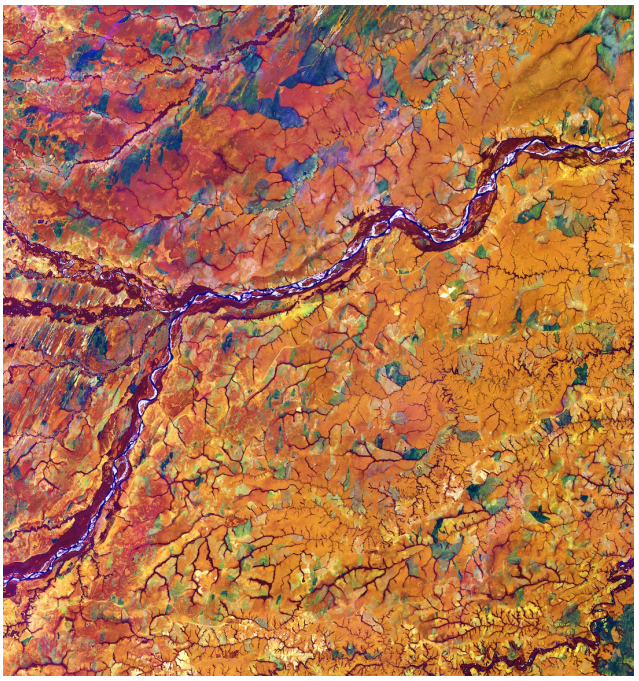
Strategy 3.1: Creating a digital ecosystem-based infrastructure that leverages access to public sector data and enables open data exchange between public sector and citizens 75

Strategy 3.2: Integrating citizen generated data into official datasets 75

Challenge 4: Questionable veracity of data provided by private entities 77

Strategy 4.1: Leveraging data stewards to improve accountability 77

EXECUTIVE SUMMARY



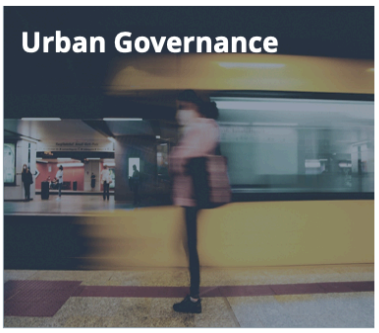
In 2020, Aapti released a report that set out a taxonomy of data stewardship, a paradigm of data governance that seeks to centre individuals and communities in the collection, usage and sharing of data. The discourse of data stewardship we've been building on since then has put forward frameworks and best practices for how intermediary entities can support individuals and communities in deriving value from data – to empower, advocate for and serve as a tool for negotiating and accessing rights – both online and offline.

While stewardship has since gained significant traction (particularly models like data trusts and data cooperatives) from researchers, policymakers and practitioners alike, it is still characterised as an emerging field.

This calls for greater focus on unpacking how stewardship may be best translated into practice. Our Stewardship Navigator tool attempted to address this in part – and sought to provide stewards or steward-like initiatives with a set of questions and possible design choices to consider.

After the release of this tool, however, we learned that many of these choices are influenced and predicated on sector-specific nuances, requirements and realities. We understand that the aspiration of building participatory data governance systems is critical to pursue, but the contours of what participatory stewardship is and is not for different stakeholders, and how participation can be actualised needs to be understood further.

Therefore, to deepen our insights and recommendations, we instead chose to identify key sectors where stewardship may prove to be most actionable and valuable:



These sectors were identified by considering the following factors:

- **Demands from the ecosystem –** *Growing consciousness around data rights, possibilities for collaboration and a realisation that better frameworks are required for data consolidation and governance*
- **Potential for societal value to be derived from data sharing –** *Use-cases exist where the value of data can be instrumental in driving collective or societal good*
- **Traction around data stewardship –** *Collaborative or participatory efforts and steward-like initiatives are developing around this domain and may benefit from greater knowledge, resourcing and visibility.*

Our first release of the playbook focuses on ***Environmental Conservation & Sustainability.***

HOW TO READ THIS DOCUMENT

The Introduction section provides a general overview of the concept of data stewardship, the notion of participation in data governance, and the aim behind drawing up this playbook.

The Environment and Sustainability sectoral guide within the introduction section elaborates on the role of data stewardship in this space, delineating the value proposition of participatory mechanisms for environmental data governance.

The table below sets out what strategies are relevant to which category of stakeholder.

Audience	Guide to plays

Stewards / steward-like initiatives	• **Play I**
	  ○ Strategy 1.1
	  ○ Strategy 2.1
	  ○ Strategy 3.1
	  ○ Strategy 3.2
	• **Play II**
	  ○ Strategy 1.2
	  ○ Strategy 1.3
	• All of Play III
	• **Play IV**
	  ○ Strategy 4.1
	  ○ Strategy 4.2

Policymakers and public sector actors	• **Play I**
	  ○ Strategy 1.2
	  ○ Strategy 2.2
	• **Play III**
	  ○ Strategy 3.1
	• All of Play IV

Private sector	• **Play III**
	  ○ Strategy 1.1
	  ○ Strategy 2.1
	• **Play IV**
	  ○ Strategy 1.2

Technologists	• All of Play III
Funders	• **Play II** ○ Strategy 1.1

	• **Play IV** ○ Strategy 1.1

Civil society organisations (CSOs) play a crucial role in enabling the data stewardship ecosystem. Through our research, we have witnessed CSOs engagement at almost all possible stages of the stewarding ecosystem. Therefore, we are of the understanding that the playbook as a whole is of relevance to CSOs, with specific relevance depending on the focus area of individual CSOs. As such, we have not attempted to classify any particular strategies as being geared solely towards CSOs.

Finally, while each strategy is geared towards specific stakeholders, we do believe that there is value for all stakeholders in looking through strategies not specifically directed towards them as it can help build a better understanding of the data stewardship ecosystem.

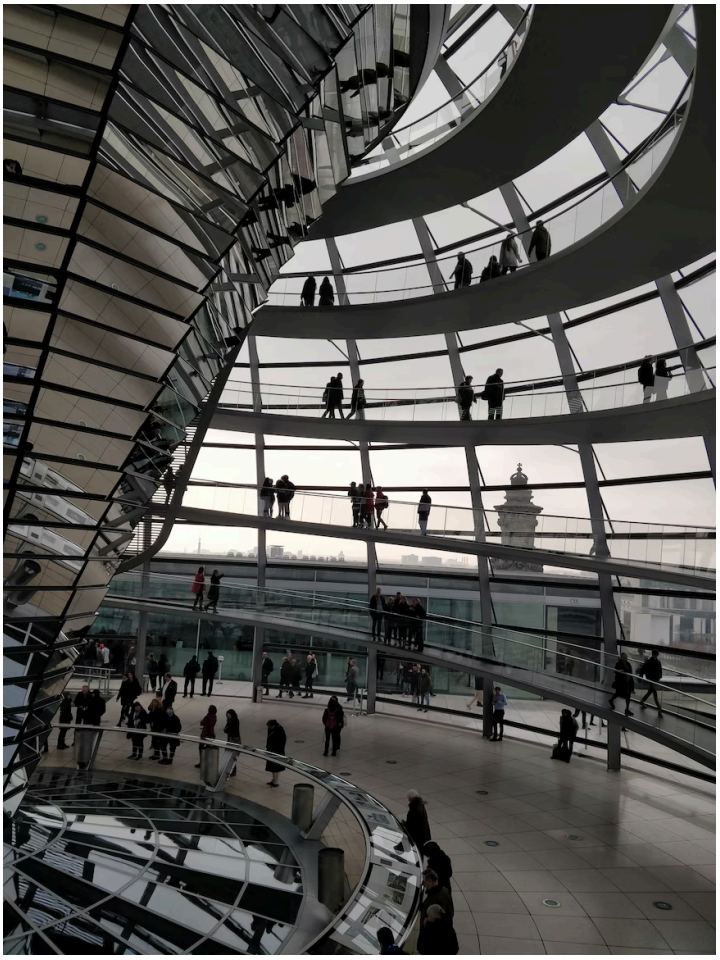
If you are interested in the challenges and strategies for the Environment and Sustainability sector, please refer to the sectoral guides:

- For Play II
- For Play III
- For Play IV

INTRODUCTION

In the current data economy, the ability for individuals and communities to participate seems to be both narrowly defined and limited in application. Varying expressions of digital rights across jurisdictions and conceptualizations of where individuals fit into the vast patchwork of data fiduciaries, intermediaries and regulators further complicates this situation.

At an individual level, most of us are familiar with exercising 'choice' in the data economy through the provision of consent – a process that has been criticized as broken and not extensive enough. Moreover, the role of individuals and communities – the primary producers of data – are largely rendered invisible, affording limited control over the use of their data by public agencies and private entities alike.



However, it is important to acknowledge that there is collective societal value that can be unlocked with data and central to this unlocking is the meaningful participation of individuals and communities in data governance, in a way that goes beyond notice and consent mechanisms. Yet, several barriers impede how participation can be defined, incentivized and scaled in the broader stewardship ecosystem.

Solving for these issues requires both an acknowledgement that current systems effectively stymie substantive participation and a deeper investigation of what practices can be drawn from historical and contemporary modes of community mobilisation to define new conceptions, modalities and pathways for participation.

What can participation look like in the data economy and what existing literature and use-cases demonstrate a promising path forward?

Participation – a foundational pillar of data stewardship

This course of inquiry finds resonance in the evolving discourse around data stewardship which pushes for a more agential and equitable approach to participating in the data economy. It imagines a more just ecosystem that enables individuals and communities and communities to exercise their data rights in accordance with their respective data priorities, while ensuring necessary safeguards are in place for the protection of their privacy as well as rights over their data.

Data stewardship's participatory ethics finds its genesis in Elinor Ostrom's design principles for community-led governance of commons – which presents a third avenue that exists beyond the state and private sector – to effectively manage the utilisation of resources at the level of communities. This framework has also been historically reflected in the governance of common resources including the management of ecological services and pooled goods like the oceans and atmosphere. A significant artifice of her work is the relational nature of common pool resources, highlighting the interconnectedness of the commons, communities and livelihoods that determine how resources are governed and how members of the community participate in such governance processes.

Participatory data stewardship similarly furnishes a spectrum of participation for communities to be involved in the use of their data. It can help ensure that communities are involved in decision-making processes throughout the lifecycle of data – from its collection through to its processing, storage and sharing to eventual deletion. Sherry Arnstein's 'ladder of citizen participation', can be used to diagnose the current data governance practices as low levels of participation, and point towards the many possibilities data stewardship provides to enhance participation. This framework is adapted by the Ada Lovelace Institute for data stewardship as well, illustrated in the figure below, surfaces a useful starting point for community participation in data stewardship.

Source: Participatory data stewardship, Ada Lovelace Institute

In application, participatory data governance systems are emerging through various models of data stewardship that include data cooperatives, data collaboratives and data trusts – many of which furnish reliable mechanisms for unlocking the social value of data for research and innovation by enabling individual and collective consent-driven data sharing. These models of data stewardship also represent diverse approaches to enabling participation at varying degrees. For instance, data cooperatives offer a high level of participation if structured similarly to their offline counterparts - in this model, all members of a cooperative are provided with a set of voting rights that can be exercised to take granular or collective decisions around data. Data collaboratives on the contrary are more conducive to participation of different types of stakeholders, and coalesce these entities around a common purpose. Governance in data collaborative models are often co-defined by these stakeholders. With Data Trusts, while their form has been imagined differently by practitioners, this model is predicated on a representative, advocacy based system of participation, possibly rooted in a board of 'trustees', with a clear tether to the beneficiary community. These models of stewardship outline lucid terms of use of community data and prescribe standards for accountability that would ensure community data is used according to their interests and imperatives.

However, based on our research over the last three years, even considering plurality across steward model types, shaping responsible data governance requires careful designing and strategizing on multiple levels – technical, community, policy and partnerships. Further, defining pathways, degrees and mechanisms to enable participation also vary, with certain domain-specific examples providing blueprints for civil society and the public sector to uphold participatory ethics.

As data privacy and governance frameworks evolve globally, there is a need to situate the role of a steward and understand what needs to be put in place to create responsive regulation. There is a need to diagnose how various stakeholders intersect and what their roles may be in the ecosystem.

Considering this, the intent of this playbook is to:

- **Diagnose key challenges** that inhibit individuals/communities from accessing data's value or impede participation in data governance
- **Coalesce strategies** that can function as 'plays' for respective stakeholders in the stewardship ecosystem
- **Draw insights from existing stewards and case studies** to highlight possible ways forward
- **Spark conversation across stakeholder groups** on what's needed to scale existing initiatives and what levers can help shape a thriving stewardship ecosystem

SECTORAL GUIDE | ENVIRONMENT & SUSTAINABILITY

Climate change has profoundly impacted the health of the planet, threatening the extinction of more than a million species, the loss of biodiversity and irreversible harm to the air and oceans - all common resources. The crisis has been further exacerbated by limited and poor understanding of data which has in turn obscured the understanding of human impact on the climate and the vulnerabilities it produces for marginalised communities. For instance, the Koli community, a traditional small-scale fishing community in Mumbai, have pointed towards a lack of substantive traditional ecological knowledge in guiding climate mitigation policies. This community is not only disproportionately impacted by climate risks resulting in shrinking fish catches, but their livelihoods are also threatened by development/infrastructure projects.

In order to understand the impact of climate change on marginalised communities, it is necessary to have environmental data which is inclusive of substantive traditional knowledge. Environmental data has become central to bottom-up movements which prioritise the participation of local communities in development initiatives and enables them to set their own goals. Advancement in data analysis and data gathering technologies have also expanded the potential of collection and sharing of environmental data. Tools developed by organisations like Digital Democracy like Mapeo have been useful in collection and sharing of environmental data, building biodiversity science that is informed by bottom-up community knowledge and involvement.

Though promising, tools like these are exceptions within a data economy where ultimate control over data is largely concentrated by state and private actors often with minimal accountability - increasing the scope for misuse. Weaponization or manipulation of

environmental data has systematically undermined conservation efforts as well as the rights of indigenous communities. The actions of the WHO in deleting India's air pollution data from its portal is one such instance where data misuse has stymied transparency and climate mitigation efforts.

Challenges with accessing environmental data & consequent impact

It is often the least well-resourced actors who stand to suffer most from climate change (with actors in the Global South being particularly vulnerable). This is further exacerbated by digital transformation-related solutions which may deepen inequities. Increasingly, it is becoming evident that the loss of agency around data and information that relates to common pool resources (such as water bodies and forests) can have a knock-on effect around conservation efforts and is also beginning to impact the livelihood of marginalized communities.

At the core of this exists a fundamental challenge in the management of public resources. The current state of overexploitation - resulting in a pillage of the commons - stands opposed to the long-standing traditions of natural preservation and a delicate ecological equilibrium which indigenous communities have successfully maintained. Prioritising a return to community-oriented modes for governance of natural resources will be necessary, given that ongoing digitisation processes threaten to amplify existing environmental injustices.

Mainstream environmental policies and interventions have also failed to acknowledge the unique role indigenous communities can play in stewarding the earth's natural resources. Studies have shown that community led protection of diversity results in better conservation of the environment. The Namibian government recognised community-based natural resource management of 82 conservancies covering about 20% of the country's surface. These initiatives resulted in improved living conditions of local communities while also restoring animal populations. However, it's also important to note that it is now indigenous and marginalised communities who face the greatest degree of impact of climate risks - and bear the brunt of an ever increasing burden of pollution. The added burden that may be imposed by stewarding these resources today must be matched with adequate support from the entities responsible for pollution.

Relatedly, technocratic decisions, the enforcement of which is entrusted unto experts and bureaucrats, are often at odds with the decision-making processes within local communities. This reflects quite clearly in the impact of digitisation of land records under the Bhoomi program in Bangalore, where as a consequence of this program, corruption increased and relatedly so did costs for farmers. Instead, the e-governance scheme enabled efficient corporate capture of land for developers and prioritised the interests of IT companies.

Existing pathways to environmental justice that can inform participatory and community-oriented governance

Elinor Ostrom's work demonstrates how communities can forge a framework to govern their common resources in a mutually beneficial and sustainable manner. Ostrom draws from a rich history of community-initiatives and organisations which have driven development efforts in aid of natural resource management. One such initiative is the incorporation of the concept of '*buen vivir*' undertaken by some Latin American countries. *Buen vivir* is based on the belief that true well-being is only possible as part of a community. *Buen vivir* began to grow in popularity in Latin America in response to depletion of natural resources, climate change and the necessity of developing an alternative to a model of economic development informed primarily by the experience of the Global North.

Similarly, the Maine Lobster Fishery is credited for its steady yield and sustainable growth for many decades due to the lobster fishers' community who monitor their livelihoods through an informal process of control, mutually negotiated rules for resource extraction and territorial governance. Elsewhere, the Gond Tribe in Mendha in the district of Maharashtra has been recognized for its movement towards self-rule and forest conservation. One of the most significant actions that the villagers took was declaring the land in the village as village commons under India's Forest Rights Act, 2006. Considered an empirical experiment, Mendha Lekha has since demonstrated a positive ecological impact on its village ecosystem and economy that is guided by a participatory resource governance system, ethics of consensus and inclusion.

Building on Ostrom's work, Frischmann developed the Knowledge Commons framework specifically tailored to the properties that distinguish knowledge from natural resources. The framework refers to knowledge as a broad set of intellectual and cultural resources which includes group information, science, knowledge, creative works and data. Frischmann's framework puts special emphasis on managing rival resources within a knowledge common since shared resources may not be fully independent to other resources. This flexibility within the knowledge commons approach has the advantage of encompassing a wider range of interests than a private right.

Knowledge commons have existed and continue to in various countries. Indigenous communities within Australia, Micronesia, Melanesia and Polynesia have developed a database including information like recording animal and plant behaviour to historical observations on biological or physical factors in climate events. This database follows strict guidelines on ability to be operated smoothly in environments with limited technical expertise, ease of use, free and open source software, respecting cultural sensitivities, portability and sustainability. This approach on commons has many lessons

for data, especially as we begin to think of it as a collective/community resource and not just an individual one.

The role of stewardship in institutionalising community-governance over resources and related data/information

Discourse on environmental stewardship and biocultural rights are also closely tied to environmental justice and collective governance. These can be similarly drawn from while building blueprints to translate stewardship into action in the environmental domain.

This requires recognizing the degree to which vulnerable and marginalised communities are disproportionately affected by climate risks. The tenets of the environmental justice movement embed ethics of procedural and distributive justice alongside a recognition of communities' rights to self-determination. These principles find resonance within contemporary discourse on data justice. For instance, Taylor's framework for data justice is based on three pillars: visibility, digital (dis)engagement and countering data-driven discrimination. The first pillar deals with both privacy and representation. The second pillar deals with freedom to control the terms of one's engagement with data markets focussed on potential benefits to low-income communities. The third pillar is power to identify and challenge bias in data use and freedom to not be discriminated against.

Building on data justice work and proposing solutions on social asymmetries and injustice, data stewardship has been proposed as a solution to monitoring data, inclusion of marginalised communities, contextualising tech-based solutions and tackling climate change. Data stewardship will be relevant to climate-related goals and for sharing technical knowledge and information with decision makers and stakeholders locally, regionally and globally.

Stewardship can be imagined and implemented in a variety of forms - one thread of this development is focused on indigenous data sovereignty. Indigenous data sovereignty (IDS) has been defined as "indigenous people's rights to control data from and about their communities and lands, articulating both individual and collective rights to data access and privacy." Some projects aiming to improve environmental stewardship have embraced IDS principles. For example, in Canada, the Arctic Eider Society is developing a platform to study ice-monitoring which is conceived as an instrument to empower indigenous self-determination. The platform's privacy features include an option to assign indigenous stewardship to user content, giving granular data access to specific communities, regional, and other affiliated local organisations.

These various bodies of research and practice around community and collective governance of resources are necessary to continue unravelling. They provide significant grounding for designing pathways for greater control and participation in data decision-making for indigenous communities. Similarly, it may help situate the role, responsibilities and related requirements data stewards may have for empowering indigenous people in their digital lives.

Considering existing discourse and our learnings on stewardship, we believe that the environment sector is ripe for collective action on data through models of stewardship. Therefore, this playbook provides a set of emerging strategies for stakeholders to help address challenges that emerge from enacting stewardship in the environment sector.

Play I: Paradigms for community participation in data stewardship

Our research highlights that while participation seems to be an area of interest, many data stewards are still reconciling how this can be factored in and prioritised – particularly during the early stages of business development and at the point of scaling these initiatives. This is corroborated in part by the emergence of efforts across the ecosystem, like Mozilla Foundation's Data Futures Lab which seeks to support builders and 'supportive entities' capacities in the ecosystem to move beyond extractive, status quo data governance models and instead prototype more just, inclusive and participatory efforts. Despite this progress, our interviews with experts and practitioners suggest substantial ground is yet to be covered – and in the absence of blueprints around participation, this remains an area that demands greater attention.

This play considers the barriers that data stewards/intermediaries and supportive ecosystem enabler organisations face when looking to define, design and pilot more participatory mechanisms and subsequently, outlines emerging findings and recommendations. These insights build on an existing domain of literature by like-minded organisations in this ecosystem and aim to kickstart greater dialogue and action around making participation more tangible.

Diagnosis of challenges – What inhibits participation of communities in the stewardship ecosystem?

Challenge 1: *Participation is structured as a 'one-off' engagement and does not persist throughout the lifecycle of data usage & governance*

Most contemporary efforts for data collection and use focus on participation of individuals and communities at the early stages of the data lifecycle – when the project is being mapped out and when data is being collected.

Avenues for participation even within citizen-science oriented efforts assume that participation is sufficient at the stage of seeking consent. However, emerging research suggests that traditional paper or oral based consent mechanisms fail to be inclusive of diverse capabilities and awareness of data values and related data rights and protections.

This challenge is likely due to a few variables that may include the capacity/resources of a steward, limited individual/community collectivization around the value of data, insufficient regulatory frameworks for participatory data governance and protections. It may also relate to the limited understanding of data and related perception around the value and possibilities of participation for individuals and communities.

These limit opportunities for citizens to be in control of how data may be used and distributed, and subsequently offer little option to be a part of the broader data value chain.

Few engagements account for participation of communities throughout the lifecycle of data – from collection to processing through to storage, sharing with third parties and the different purposes for which data is used. The consequent marginalisation of communities could be attributed to a variety of factors ranging from resource constraints of stewards (*See Play II for more on this challenge*) who may not have the financial or physical means to enable granular consent provisioning for the use of their data. Elsewhere, the problem of lack of engagement could also be attributed to the absence of meaningful blueprints to inform and involve communities in data decisions. The emergence of consent dashboards and comic-based consent narratives are emerging pathways to address this long standing problem but are few and far between to constitute a robust blueprint for enabling participatory data governance.

The above limitations are only bolstered by the prevailing regulatory landscape around data and information markets which is overwhelmingly concerned with privacy protection as the overarching goal of regulatory action. As a result, individual and community participation in data sharing and governance is limited to consent provisioning through cookie notices and prolix of terms of use as displayed across a variety of digital platforms. Not only are cookie notices inaccessible in themselves, but they also lead to 'consent fatigue' among users, actively undermining their ability to make informed decisions about their data. For instance, a study about cookie consent

mechanisms examining 80,000 unique users on German websites has demonstrated that they are manipulative inasmuch they do not offer a reject button on the notice. Furthermore, users are "nudged" towards assenting to such notices through deceptive design and where privacy-preserving options exist, they lay buried or inaccessible within cookie banners.

Given the manifest limitation of the 'notice and choice' approach furnished by cookie notices, there is a need to move away from this "privacy model" of data regulation in favour of an "accountability model" for data governance. The "accountability model" places dual emphasis on one's ability to control the downstream use of data and hold data users liable for their actions. As a result, meaningful consent and responsible data use become the guiding imperatives for regulation of data.

Strategy 1.1: Identify incentives of data generators (individuals and communities) through consultation to better identify approaches to embed participation

In order to enhance the possibilities for participation, it is necessary to first be aware of the varying value, incentives, bandwidth, interest and capabilities of individual communities in participating in data collection and governance--related decision-making. Worth noting is that there is likely to be significant variation in incentives at an individual vs community level, this may present challenges where individual values may at times be at odds with collective community goals. Assessing these factors will be necessary in better understanding how stewarding entities can structure participation and for those supporting these efforts, what must be done to facilitate greater engagement.

Tools such as the Data Maturity Assessment help a variety of organisations take stock of their data goals by examining three axes: purpose, practice and people – to produce a data lifecycle evaluation. Purpose relates to an organisation's strategy, analysis and applications of data while practice encompasses the infrastructure, quality, ethics and security protocols in place to handle data. The 'people' axis looks at the decision-makers within the organisation and their approach to data use. These assessments, however, are likely to have the greatest utility for established organisations - and carrying out these assessments also requires significant commitment and organisational buy-in - which may be viewed as burdensome for newer 'stewards' or data intermediary entities.

Similarly, other tools that may be used to assess data priorities include the Stewardship Navigator - this guides potential data stewards (or interested parties) and the communities they aim to serve through essential considerations about structure, sector and internal data governance standards of an entity, providing pathways towards responsible data use and sharing for public benefit.

The results of the above assessments function as precursors to crucial conversations with a community of data producers who can determine their subjective interests and incentives for participation in data governance. On the one hand, certain communities may (choose) not to be involved throughout the lifecycle of data. To this end, mechanisms of delegated representation may prove to be useful to ensure trusted intermediation and sharing of their data. A valuable example of such a mechanism for delegated representation can be found in the MindKind Study, supported by the Wellcome Trust. The goal of the study is to establish a Global Mental Health Databank – a kind of data collaborative – through volunteer contribution of mental health data by youth in participating institutions across India, South Africa, United Kingdom and United States. Participants in the study – in effect, the research community – are involved in co-designing the questions that the study should aim to address as well as design a mobile platform to collect mental health data from other youth. Therefore, the involvement of the community in the MindKind Study extends so far as project design, data collection and recruitment of youth data contributors to the platform, while the data bank and the purposes for which it can be used are not in the scope of the community's control. The role of the community falls within the 'inform' and 'consult' components of the Arnstein spectrum of participation.

Elsewhere, communities may wish to be involved at every stage of the data lifecycle – right from its collection to processing and sharing with third parties for certain pre-defined purposes with certain approved and vetted data users. Such participation is also anchored in perceptions of value generated through data use and sharing, and whether community members can exercise meaningful control over their data – falling in the 'empower' category of Arnstein's spectrum.

For instance, a group of patients with multiple sclerosis might be more invested in sharing their data for medical research on the disease as opposed to monetizing their data for drug development by pharmaceutical companies. The patient community here chooses to prioritise creation of public value through knowledge generation and research on multiple sclerosis, as opposed to monetizing data for drug development that generates proprietary value for a narrow subset of pharmaceutical companies. MiDATA, a health data cooperative founded in Zurich, Switzerland, attempts to precisely create this sort of public value through its platform by allowing users to share their data for specific medical research projects, among which multiple sclerosis features prominently. The cooperative structure herein allows its members to contribute data for causes and projects that resonate with community values by sharing data through democratic voting mechanisms within a general assembly of cooperative members.

Strategy 1.2: Remodel the prevailing regulatory landscape for data governance to embed mechanisms for community participation throughout the lifecycle of data usage

The prevailing landscape for regulation of information markets is predominantly preoccupied with consent and data protection as means to authorise data sharing in the context of personal data. The problem with such an approach is that it violates the contextual integrity of privacy in ways that hinder individuals and communities from engaging effectively with downstream uses of their data, to direct its use by actors and for purposes that ultimately benefit communities that produce this data. The European Union's General Data Protection Regulation, 2016, India's Digital Personal Data Protection Bill, 2022 and Ghana's Data Protection Act 2012 are some of the data protection legislations that follow this "privacy model" of regulation. Such regulations suffer due to their inability to comprehend the 'social value' of data, failing to account for the many positive externalities (ex: research through combining different datasets) and negative externalities (ex: potential privacy loss through data de-anonymization) that are inherent to data-driven innovation. Communities neither benefit from the positive externalities nor have any avenues to mitigate risks produced by negative externalities due to their fundamental lack of autonomy over data decisions. The paradigm of creation, collection and use is even worse in the context of non-personal data, where there is yet no significant legislation that recognises the interests of communities over non-personal data. Aside from fledgling efforts in Europe and India there are no significant policy endeavours worldwide to govern non-personal data sharing. Data such as energy use information, crop yields, air quality data are created through shared endeavour and have important insights about communities, and have immense public value. Communities however do not derive any benefit from such data.

Therefore, it is incumbent upon public institutions to recognize the invisibilisation of communities and remedy the same through the introduction of participatory mechanisms for data governance such as data stewardship. Precursors of such an impulse to further participation of communities find expression in the EU's Data Governance Act which recognises data cooperatives. However, the language of the Act frames cooperatives in a very restrictive way, with significant debate over whether individuals can even delegate their rights under the GDPR to cooperatives.

Elsewhere public authorities are contemplating standard-setting for data stewardship, opening doors for institutionalisation of participatory mechanisms for data governance within policymaking. Canada's CIO Strategy Council has proposed operational models for data stewardship like data trusts, data collaboratives and data cooperatives as a part of its National Standards for Responsible Data Sharing.

Strategies 2.2 and 2.3 of Play IV (Ecosystem enablement – Role of public sector) surface valuable frameworks to contemplate how public institutions can support and further community participation in data stewardship efforts.

Challenge 2: *Data gathering and sharing efforts risk being one-sided and exploitative in the absence of clear incentives that would deliver broad-based public benefit for communities that share data*

Despite the many positive externalities afforded by data, emerging technologies such as machine learning have led to new approaches to the collection, use and sharing of data that are often extractive, inequitable and disenfranchise communities from participating in the governance of their own data. For instance, users signing up on digital health platforms have little control over how their data is used, just as gig workers are excluded from the audit of algorithms that govern their working conditions.

This practice is also prevalent in mainstream scientific and academic data gathering exercises which often maintain and reproduce the 'researcher-subject' dynamic, where the researcher extracts data and knowledge often with little in return to the 'subject' of enquiry - often treated as passive agents in these one-off, transactional engagements.

Persistent concerns around the misuse of data, combined with a lack of bottom-up engagement with communities that are affected by data use have systematically eroded public trust in the process of data sharing.

This trust deficit manifests itself as 'data hoarding' and 'data fearing' scenarios – two inter-related phenomena where communities and organisations either restrict access to data and prevent it from being leveraged for public benefit or communities withdrawing consent for use of data due to fear of privacy loss that altogether prevent data from being collected in the first place. This is further exacerbated as communities are not able to visualise the benefits (tangible or otherwise) due to the complex routes of data usability. This is

For marginalized communities (indigenous peoples, gender or sexual minorities, refugees, disenfranchised groups, etc.) that have historically been subject to these extractive relationships between the state or academic communities, the consequences of this dynamic impacts more than just the willingness to engage, participate or trust in data collection or sharing exercises. More insidiously, many of these dominant frameworks of knowledge, data gathering, and usage can often subvert, limit or contribute to the erasure of indigenous ways of knowing and traditional forms of knowledge and practices.

Strategy 2.1: Tap into existing networks or organisations that possess trust-based and collaborative relationships with communities and identify how value can be jointly delivered or enhanced

Many organisations that actively hold data assets now presume a greater responsibility or custodianship over the data of their beneficiaries (either end user organisations like community-based organisations, individuals or communities) and are keen to understand how their rights can similarly be protected. For instance, trade or credit unions are increasingly concerned about the harm poorly stewarded data may inflict on their members and in parallel are alive to the possibilities data presents in furthering collective negotiation efforts. Building the capacities of these entities to enhance their role as data stewards can be a meaningful way of facilitating greater participation and decision-making around data. Prospect's Lighthouse 'purpose-made digital governance maturity test for trade unions' offers a useful case-study of how this guidance can be designed to be domain or organisation specific.

However, some of these stewards or steward-like entities may not have an active tether to the community. This may be intentional on the part of some organisations who choose to focus their resources and vision on building out the technology or tools to enable self governance and collective decision-making.

*****PescaData Video to be inserted here*****

For organisations with a similar vision to Pesca Data, there is a need to build the capacities of these end-user organisations to leverage their tool efficiently and self-sufficiently. In this process, they invest significant resources to support the onboarding process as well.

However, the relationship between these end user organisations can also be a feedback loop of sorts – where the needs and priorities of their beneficiaries can be better surfaced. In these cases, stewards would do well to first identify organisations, associations or collectives that work closely with communities – ideally those that have pre-existing relationships of trust.

From there, a steward's role more specifically would be to understand how these organisations deliver value to their communities – for example, many of these organisations provide legal advocacy services or are an intermediary organisation that creates offline architectures to facilitate greater access to state services and associated rights. Many of these efforts can be better aligned with or strengthened through participatory and secure data collection and governance processes. For instance, through an organisation called Rainforest Connection, GIS and bioacoustics data has been leveraged by indigenous communities to alert instances of poaching and can be

used as a tool to highlight encroachment of land by illegitimate entities in legal cases. In this instance, Rainforest Connection provides the technical resources and guidance around the collection of data – yet the end goal of its usage is determined by the communities who can choose how and where (or for which use-case) that data can be put to its greatest value.

The value and associated "use-cases" around data act as both a precursor and incentive for communities to participate or see tangible benefit(s) in engaging with a steward or steward-like entity.

A steward attempting to define and deliver on the "value" of data in isolation, without the consultation of communities and end-users is often a complex task. Mainstream discourse valorizes the monetization of data in the current data economy, even if that's not always what individuals or communities actually desire. This sole focus on monetizing data as the only expression of its value can also create perverse incentives for communities to forgo their data along with the rights it bears as well as stewards who may privilege monetary returns over community imperatives. Monetization of data is likely to deepen existing inequalities, and make privacy a privilege for the rich. Therefore, it is important for stewards or steward-like initiatives to broaden their understanding of the value of data and how this can surface differently depending on use-case and communities' subjective needs and requirements.

For newly established organisations (technical with limited resources/personnel), other supportive entities in the ecosystem may provide the missing key of what this value can look like and therefore they should explore how to collaborate to combine and supplement this value addition.

Digital Democracy, a nonprofit that has co-developed an open-source mapping tool for indigenous communities (Mapeo), identified a few qualities in a partner that are important to account for. For instance, those with technical personnel/capacity are better placed to provide communities with greater on-ground onboarding support and meaning making around data. These partners are also better enmeshed in the societal conditions, realities and needs of the communities to support them in identifying new use cases for data.

One such methodology that can be leveraged by partners (whether end-users or the stewards themselves) is to carry out data feedback sessions, described in the case study box below:

Methodology Case Study #1 –

Data Feedback Sessions [Abalobi & Blue Ventures]

In partnership with Blue Ventures, Abalobi released a report with a toolkit on 'Community Engagement with Data' . This document outlines their experience and recommendations for practitioners building out community based marine management systems. Key to their methodology is facilitating 'data feedback sessions', a participatory strategy which brings together communities in a forum to engage on questions, reflect on experiences and take active decisions around both the insights from data and its possible usage or sharing.

These feedback sessions are described to have a several benefits in that they support communities in; recognizing data rights, empowering communities (through building credibility around knowledge contribution and enable decision-making input), building trust and resolving conflicts, data validation and informing decision-making,

adaptation, and change.

Their toolkit highlights a few different considerations and principles around the content, approach and tools that can be leveraged during these sessions, highlighted in Figure 1.

This methodology is particularly relevant for those in the environmental and sustainability sector as it is oriented around an 'adaptive management approach', a framework adopted by conservationists that foregrounds substantive community involvement in the monitoring, planning, management of resources.

Strategy 2.2: Creating supportive legal and technical infrastructures that prioritise 'participation by design' for meaningful community participation in data stewardship

An essential corollary to "privacy by design", the concept of "participation by design" borrows from the former and refers to a range of technical instruments available to public institutions to embed community participation in the process of data governance and sharing. Such a move to embed participation within the technical architecture of platforms can redress long-standing asymmetries in the digital economy where individuals and communities have little visibility into how their data is used by private corporations and public agencies that collect their data. Supportive regulation that carves out a role for communities in data governance is necessary and Strategy 1.2 of this play delves into the specifics of what such legal infrastructure might look like.

More importantly, it is imperative that public institutions also invest in the creation of digital public infrastructure - 'digital solutions that enable basic functions essential for public and private service delivery, i.e., collaboration, commerce, and governance" - that would allow communities to participate meaningfully in data decisions. Examples of

such DPIs include the X-tee data exchange layer built and managed by Estonia's Information Systems Authority that provides a secure information exchange which is confidential and interoperable. Estonian citizens can access a variety of services such as health insurance, digital signatures, banking and voting through their digital identifiers which is linked to the X-tee framework, retaining control over who has access to this information and how it is shared.

In a similar vein, India's banking sector regulator, the Reserve Bank of India, has rolled out the Account Aggregator framework – a data intermediary that facilitates consent-driven data exchange between financial information providers (FIPs) (ex: an individual's bank account) and financial information user (FIU) (ex: credit lending agencies). When FIUs request data from FIPs, the AA will request the data principal (owner of data) for consent to share the data. The promise of the AA framework lies in creating an efficient and connected financial information ecosystem that is powered by user consent and recognition of one's agency over their data.

Strategy 3.1 of Play IV (Ecosystem enablement – role of public sector) and Strategy 1.1 of Play III (Technical pathways for operationalizing community participation in data stewardship) expand on the above by proposing pathways for community-driven data governance and sharing.

Challenge 3: *Data gathering efforts although defined as participatory continue to be surface-level and non-diverse as structural barriers (class, gender, ethnicity/race, age, citizenship status), vulnerabilities or capacities (technical, financial and data literacy) are not accounted for*

Several data gathering projects that rely on citizen generation or "donation" of data often presuppose a few different factors that contribute to a 'citizen's' ability to substantively participate. Varying incentives, access restrictions and abilities are often not accounted for when designing these data-related projects or infrastructures.

For data intermediaries that are cognizant of these barriers and are intent on building more diverse and inclusive data pipelines and spaces for effective participation, this lack of representation is more of an unintended outcome – largely the result of narrow sampling methods or limited pathways to "customer acquisition". For example, a recent mobility data cooperative Posmo, based in Switzerland, captures data from a limited set of citizens that are characterised as able-bodied, city dwelling and perhaps from a particular age range. This sample size is reflective of the early outreach PosMo carried out to its existing community of supporters, colleagues and other value-aligned individuals.

This challenge is particularly acute among emerging stewards who are in the process of defining the modes and mechanisms to best acquire customers or build out their membership. This lack of diversity in the representation of data generators must be addressed as without adequate inclusion both in the data collected and the ability for a range of individuals being able to participate - this may lend itself to reinforcing existing data biases.

Strategy 3.1: Forge partnerships with existing community-based organisations to solve for issues of representation within data stewardship initiatives

Reimagining data futures is critical to ensure that data governance and the communities that are helming these efforts are sensitive to problems of lack of diversity that are pervasive within data stewardship initiatives. Forging partnerships with community mobilisation and advocacy organisations present a tangible pathway to resolve challenges of inclusion faced by stewardship entities. Research by Aapti, undertaken as a part of its efforts to build the Stewardship Navigator tool, documenting stewardship initiatives across the world indicates that 56.6% of all such initiatives originate in the Global North – countries based in Europe, North America and Oceania. The Open Data Institute's Data Institutions Register contains a log of 204 organisations working as stewards, of which 89.70% entities operate in the Global North. Consequently, low- and

middle-income countries in the developing world, as well as marginalised communities within the Global North, find little mention within such databases. While it is likely that there are overall fewer stewardship or steward-like initiatives in the Global South, the difference in numbers is unlikely to be at the level of causing an overwhelming majority of institutions in such databases to be from the Global North. Nonetheless, there is cause for greater emphasis to be put by actors across the stewardship ecosystem to push for more stewardship initiatives in the Global South and in marginalised communities world over.

As a result, there is a pressing need to move away from Euro-centric visions and practices of data governance to ensure that the stewardship community is alive to the experiences of discrimination and exclusion faced by disenfranchised groups. Stewarding organisations can stand to benefit from partnering with initiatives such as Data for Black Lives – "a movement of activists, organisers and scientists committed to the mission of using data to make concrete and measurable change in the lives of black people". In turn, Data for Black Lives partners with organisations working for racial justice to counter bias inherent to data and algorithmic systems. Forming partnerships with such initiatives can help stewards solve for twin issues of lack of diversity and representation as well as scalability by leveraging existing networks within communities that Data for Black Lives enjoys.

Other valuable partners include the Environment Data and Governance Initiative – a research and advocacy network working with organisations and communities concerned with climate change, science policy, good governance, and environmental and data justice. The EDGI hopes to focalize stewardship of public knowledge about environmental issues by enhancing the use of existing environmental data, through tools like Jupyter Notebooks, which can support greater awareness and data-driven decision-making. Ethics such as intersectionality and a commitment towards anti-oppression are further affirmed through partnerships with grassroots communities that lead climate action. Lastly, EDGI also offers much needed technical support to communities to gather, process, make sense and act using their data.

Strategy 3.2: Locate and empower engaged community members to build bottom-up data-oriented communities and facilitate more diverse onboarding for stewarding organisations

In the absence of well-defined communities or relevant infrastructures (e.g unions, collectives, community-based organisations and self-help groups), stewards or enabling organisations should invest in identifying specific individuals (champions and early adopters) from the community interested and aligned with the goals of data stewardship.

Stewards may be best placed to also identify these members of the community that are typically not represented in other broader community groups. For instance, while fisher women or women in the fishery management supply chain contribute significant labour, in many communities this remains invisibilized and unrewarded. This asymmetry of power was recognized by Abalobi, a social enterprise that empowers fishers through co-created ICT technologies and data analytics products based in South Africa.

Upon reaching out to some of these members, Abalobi found that women were also more likely to both demonstrate interest and, in some instances, possessed greater capacity and bandwidth to leverage their technologies. They also showcased a greater engagement and incentive to participate in the co-creation, development and governance of Abalobi technologies. Abalobi chose to centre the fisherwomen as key pioneers in building their own collective data cultures and associations. At a structural level this meant creating a layer of foundational organisational governance composed of representatives from these communities. Responsibilities involved also seeking and onboarding new members through the articulation of benefits.

This intentional sampling and overarching self-governance model Abalobi have put into action is closely tied in with their theory of change – which is to empower and build capacities for more agential and transparent collection and usage of data.

Play II: Scaling community governance through data stewardship - moving beyond pilots

While human centric data governance models have been on the rise, to unlock the value of responsible data sharing and bottom-up governance - there have been myriad roadblocks along the way. Yet, the landscape of data stewardship has matured considerably in recent years; both in the volume of stewardship efforts, as well as an increased interest from other players within the ecosystem to magnify initiatives that foreground community value. One of the primary challenges of these stewardship efforts has been an inability to scale - to move beyond pilots or otherwise temporally bound structures. Scalability, particularly for data stewards, is a multi faceted and multi stakeholder challenge. While this play speaks to stewards themselves, it is not without the support of other pillars in the ecosystem that sustainable scale can be realised. There is a dearth of literature and evidenced research for stewarding entities to fall back on when considering revenue models, partnership avenues and more. Avenues for funding remain limited or unknown to many stewarding initiatives, despite a marked increase in investors' interests in data sharing. Technical capacity, both in the communities being served as well as within the steward itself also poses a major challenge.

Collective governance is strengthened most significantly by an increase in the volume of the collective, the capacity of the collective, and support for the collective. Thus, if we envision data stewardship as a means to unlock wider societal value from data, while also preserving and meaningfully amplifying data rights – scalability becomes crucial. The question of scale operates not just at the level of an individual steward, but is a concern even at an ecosystem level. Scaling of data stewardship initiatives, or rather, an increase in the number of stewarding initiatives across communities, is also a matter of concern. Our research has found a disconnect between various stewardship efforts despite the inherent possibilities of sharing learnings, challenges and opportunities. Without this, the ecosystem of data stewardship and consequently, its benefits, will remain scattered.

Challenge 1: *Ensuring financial sustainability beyond the pilot phase*

The field of data stewardship has been dominated by pilot projects and experiments that fail to reach scale or are abandoned when funding dries up. Relying on time-limited funding, such as grants can be challenging to predict future revenues. To maximise the value and impact of their work, data stewards are faced with the task of managing institutional change towards cementing the culture of data stewardship. While data stewards might generate revenue from varied sources (including selling data and membership fees) the most common source of income / funding for stewards are grants / funds received from institutions - be it private philanthropies or public sector actors. While a sustainable business model for a steward will rely on a healthy mix of funding sources, stewards should look to prioritise specific sources based on their lifecycle. Critical to carrying this out is to draw a distinction between earned revenue (revenue derived from supply of data and allied services) and non-earned revenue (donations, grants and other funding sources). Typically, a steward will have to rely more on non-earned revenue sources during the initial stages of their lifecycle, and slowly shift the dependency towards earned revenue as they grow. A healthy diversity in revenue sources has been associated with sustainability of an organisation.

This is easy enough on paper, and generally a model followed by most businesses. However, with data stewards this is particularly challenging for a few reasons: (a) the altruistic goals, of data protection and seeking public value for data, are ones that people are ordinarily not inclined to pay for; (b) data collected by stewards is of value only when it is of a certain critical mass, making it representative of the population it relates to; and (c) sharing data under conditions of privacy with strict purpose limitations is not how the digital economy has worked so far. Identifying a clear business model is therefore critical for the financial sustainability of a steward, but also a challenging one.

The public sector can also play a crucial role in ensuring the growth and sustainability of data stewardship initiatives. However, funding from the public sector is more often than not limited (for more, please see Play IV). Similarly, tensions may arise if funding agencies and data stewards have goals and priorities that are not in sync. Funding for data management is often irregular and of limited time and scope. This, in turn, also affects the potential to increase technical capacity of data stewardship efforts through involvement of developers or technologists, keeping most pilots suspended in small-scale, volunteer efforts. Philanthropic funding is often fixed over multi-year contracts. There is a danger that funding will not keep pace with growing data volumes impacting the scalability plans of data stewards.

Strategy 1.1: Assess value delivery and product accessibility in order to identify suitable business models

Earned revenue is generated from two main sources: (a) from the community members (either through membership fees or through services provided to members); and (b) from external sources (through the provision of data and allied services).

At the outset, identifying the governance structure of the organisation and the value that the steward seeks to deliver to the community are key in delineating strategies for membership or subscription pricing. Research shows that membership rather than subscription fees might offer additional benefits for some data institutions. Membership encourages active participation and is used in a way that ensures people contributing data are the one governing access to it. Membership models convey a sense of belonging, trust and community based on shared values and interests. Members may be expected to contribute in both monetary and non-monetary ways, for example with their energy, expertise and time. On the other hand, subscriptions are a simpler, transactional exchange of services for a fee. Crossref, a membership organisation that assigns and maintains identifiers for research outputs, supports membership because it emphasises the notion that data stewardship is a collective endeavour.

However, not all members will always have the time and energy to participate in the stewarding organisation in the same way. Based on their driving aims, a steward could have a decision making system that involves each and every one of its members, or can opt for a model where the governance structure of the stewardship can delegate the management burden and entrust the steering responsibility to properly chosen people or organisations. Stewards can opt for a hybrid model, and accordingly stagger membership fees, thus maximising on revenue from community members.

Equally important are the incentives for members to participate in the activities of the steward. Most participatory stewarding organisations are structured as community-owned, and it is important for the steward to determine how returns from the steward's activities are distributed amongst its members in a way that incentivises participation, but not at the cost of overall membership figures. Research into sustainable revenue-allocation schemes for data cooperatives shows that a "Robin-hood" model works best, where the right amount of additional incentive is provided for privacy-sensitivity. The research cautions that such incentive cannot be too lopsided as it can be taxing on one set of members, possibly fracturing the membership group.

In terms of external revenue sources, data stewards may seek to change how they deliver value by restructuring their business models to make their products and services more accessible. E.g. Idaho Health Data Exchange partnership with Amadeus platform

proved to be successful in enhancing value for patients. The partnership with Amadeus platform provided dedicated resources in technical onboarding of new data providers and their interfaces. This increased the overall value of Idaho Health Data Exchange participants while creating system-wide improvements in the value of patient care.

This example also relates to another manner in which stewards can generate revenue outside of their core business. While stewards can sell the data they collect, stewards can also look into partnerships providing in-kind support, such as technical infrastructure, administrative support, or access to a particular set of actors.

The task of assessing value delivery and product accessibility is not one-time. Based on the stage of the steward, the results of assessment are liable to change and evolve. It is therefore important for stewards to periodically undertake this exercise. Our research and conversations with CSOs has identified key questions and considerations for data stewards based on where in the lifecycle of a steward they are.

Strategy 1.2: Funder flexibility to understand long-term stewardship goals

For funders, the priority must be to provide sustainable funding to support infrastructures required for long-term stewardship of data. Supporting bottom-up approaches that are already in place with grants is a recommended approach instead of developing new data governance programmes. On a practical level, funders should include multi-year flexible funding, streamlined applications and reporting based on commitment to build relationships based on feedback, transparency and mutual learning. For example, Co-impact, a philanthropic collaborative, recommends trust-based philanthropy in everyday practice to address social issues.

Grant seekers are often encouraged to customise their proposals to fit funder priorities, which may have been developed based on inadequate consultation with the target audience. This can create tensions on the functioning of a data steward and its intended goal. Giving grant seekers the space to step back and proactively articulate their own strategy and vision can lead to greater sustained change and success over time. Pooled funding models encourage collaboration among funders. This can help to reduce the transaction costs associated with multiple processes for managing, verifying and sourcing. It reduces the risk of duplicating efforts. Network-building efforts from funders like collaborative philanthropies can be useful in ecosystem strengthening as well - e.g. arranging convenings of funders (and/or grantees) to expand their knowledge on ongoing work.

Challenge 2: *Identifying incentives and community-building*

A factor that is important for a steward to scale up, is the ability of a steward to continually expand their user base. However, this has proven to be a major stumbling block for many stewards. Our research and conversations with ecosystem stakeholders indicates that this might have to do with incentives for being part of a data stewarding initiative.

The primary goal of most data stewarding initiatives, if not all, is to provide community members with more control and agency over their data, including in matters of who it is shared with and what uses it is put to. However, this is unlikely to be a sufficient incentive for most people. While there are groups of people concerned about their privacy and the impact to their rights from misuse or widespread sharing of their data, this remains a relatively restricted group of people. This is a problem that is more pervasive depending on the context and the community in question.

The problem of scaling of data stewardship, particularly for non-steward stakeholders in the ecosystem, is not restricted to scaling of one particular initiative alone, but scaling of the concept of stewardship as a whole. Despite the growth of data stewardship in discourse on participatory data governance, stewardship remains a rather esoteric concept. Even within the data governance community, many people and organisations are entirely unaware of the concept. And in some cases, organisations that would be categorised as data stewards were unaware of the literature around this concept, or were unsure of the taxonomy to be used in finding adequate resources.

Strategy 2.1: Identifying incentives beyond agency over data

As mentioned above, agency over one's data is often not incentive enough for people to join a stewarding initiative. Marginalised or poorer communities from the Global South, for example, typically either do not know or care about their data rights. For example, a research effort undertaken by Aapti Institute and the Open Data Institute for the Global Partnership on AI involved co-designing data trusts for climate action. This included the design of three data trusts: one for cyclists in London, one for small shareholder farmers in India, and the last for climate migrants in Peru. While there was clear indication that there was a demand for a bottom up data trust from cyclists in London, this demand was missing in the case of India. Farmers and CSOs working with farmers we spoke all articulated that agency and control over data was quite low on a list of priorities for farmers, and unless we could show incentives in terms of better financial access or access to wider markets, farmers would be unlikely to sign up for a data trust.

This issue of identifying incentives is especially critical at the stage when stewards are looking to expand their user base. Undoubtedly, there are likely to be data minded

community members who have not joined a stewarding initiative simply because they were unaware of the initiative. In such cases, the usual strategies of widening communication channels, and advertising the initiative to the target audience will be successful. However, in order to reach a wider audience and gain the increase in membership base that is required to scale, it is imperative that stewards identify incentives beyond simply agency and control over data. These incentives will vary based on the community the steward is looking to serve and the nature of data they collect. However, given that the data the steward collects can be put to various uses, once stewards have built a strong foundational dataset, they can use the services and partnerships this dataset unlocks to provide varied incentives to attract new members. Aapti's recent work as part of the 17 rooms project involved assessing the value of adding a data layer to an existing agricultural cooperative. As part of this, we identified the specific benefits that would accrue to the cooperative members from adding the data layer. These included better access to credit by building a digital identity for farmers, generating data on group funds to help secure more credit, and possible improvements in yield data-driven advice on better farming techniques. Women farmers Aapti spoke to noted that these were much more attractive incentives for them.

It should also be noted that the incentives need not be linked solely to the data collection effort alone. Abalobi, a steward for small scale fisheries in South Africa, for example has helped visibilise the labour of women in the fishing value chain and has been able to realise actual value to them. This is only one in a range of incentives that Abalobi can speak to that are a result of its data stewarding efforts - including strengthening the community, capacity building, and sustainable fishing.

Strategy 2.2: Identifying and communicating challenges, enablers and strategies

Scaling of data stewardship, at the ecosystem level, relies on proper communication of learnings and principles from existing initiatives and reporting what the challenges the initiatives have faced and enablers they have had based on their socio-economic and political contexts.

In our engagement with data stewardship over the past three years, we have come across numerous steward or steward-like initiatives from across the globe, as well as various other organisations in the stewardship ecosystem that would benefit from the work / learnings of other organisations in the ecosystem. However, in many cases, there was a lack of awareness of the other organisations, or even a dearth of literature around these efforts. As an anecdotal example, in addition to Abalobi referenced above, we spoke to 2 other stewards / steward-like initiatives operating in the small fisheries space, and came across three more such initiatives, all from different parts of the world. Even in the case where these initiatives were aware of the work the others were doing,

which was not always the case, there was a dearth of resources that an initiative could access to inform them of the lessons learned by the others.

The burden of creating these resources however must not be placed entirely on stewards themselves. Civil society organisations can play a major role in identifying the key learnings from these stewards and distilling them into resources that can be made accessible widely. An example of this is the paper that Aapti Institute wrote on the lessons that could be learned from the various stewards for small scale fisheries, and some of these learnings were applicable not just to fisheries, but to stewards in other sectors as well. Indeed, even this Playbook serves as a resource to provide stakeholders in the stewarding ecosystem with knowledge about other initiatives that they could possibly benefit from.

The community focus a steward has might put a limit to the scale of members the steward can reach. In many cases, attempting to expand beyond that specific community might not be the most feasible option from a business perspective. However, the model itself can be replicated in another community, or another context, and resources that speak to the challenges faced by the steward as well as strategies that worked for them can be critical in helping new initiatives thrive. Additionally, such resources also serve as important informants for governments and funders in their process of setting their funding priorities, and identifying possible avenues for funding.

Challenge 3: *Limited technical capacity*

Collecting data, implementing stringent systems and protocols to ensure data security, being able to analyse the datasets for insights and building applications to collect data are all complex tasks that require a high degree of technical capacity. The 2022 MongoDB Report on data and innovation revealed that 73% of respondents agreed that working with data is the hardest part of building and developing applications. Addressing these challenges is critical to ensure a data stewards ability to scale successfully - not only is it a core component of how the steward functions, it can also have significant impact on internal efficiencies. The challenges faced in technical pathways for operationalising community participation in data stewardship are dealt with in Play 3. This Play will focus instead on strategies stewards can take to address the challenge of limited technical capacity.

Strategy 3.1: Collaboration between data stewards and tech developers

The ability to manage and anticipate risk in design and development, particularly those systems that are complex because of sensitive data can prevent 'techlash'. Stewards can look to partner with tech developers who are looking to avoid the pitfalls of the current digital economy and instead are focused on creating new alternatives designed to foster a more participatory future. The UK Behavioral Insights team for example, has designed two platforms: Your Priorities, a citizen engagement platform connecting citizens with the government, and Applied, a recruitment platform, with a focus on fostering diversity.

Such organisations can anticipate user needs more accurately given their wealth of experience. This will help build applications that are easy and intuitive to use, thereby improving user retention and helping acquire new users. However, stewards don't always work with communities that are digitally literate. It is therefore key that such partnership ensures that developers should focus not only on safe and secure design, but also on how the interface of that technology remains accessible, developing applications that work equally well for groups that lack digital skills in order to allow for equal access. Play 3 looks at how co-designing strategies can address such challenges.

Strategy 3.2: Easier data discovery, privacy enhancing technologies and trusted research environments as drivers of scalability

Enabling easier data discovery and its potential application can scale a data steward. To address the challenge of stale data, stewards can start with a core discovery platform. This can empower data stewards to uncover context for data usage, reduce time to understand impact analysis and derive meaningful insights from data. Further, including Privacy Enhancing Technologies (PETs) can offer the steward the ability to accelerate

secure data collaboration/exchange, build trust and maximize data value without compromising privacy. Given the privacy is on the forefront of a data steward's activities, including PETs can enable a host of data collaborations that make data more valuable to internal teams as well as external partners while preserving privacy. However, as the Global Partnership on AI notes, there are challenges to data usability commonly faced when working with PETs, and it is incumbent on other stakeholders in the ecosystem to address these in order to provide stewards who are typically smaller with the ability to compete with larger organisations.

Play II | Sectoral Guide | Environment and Sustainability

Challenge 1: *Limited accountability and transparency around Environment, Social and Corporate Governance (ESG) reporting and Environmental data sharing by private companies*

Studies indicate that the growth of ESG investment is inextricably linked to data. Companies that pay to collect and analyse environmental data consider those datasets to be business assets. As a result, they are generally not eager to share them. Companies often treat environmental data as confidential business information. Data may offer the company that collected them some competitive advantage over their competitors. The preservation of such data can come in conflict with the imperative to share research data. The majority of valuable data is siloed with private companies, inaccessible for the wider public. Moreover, in countries with weak climate information structures, this lack of data can also affect larger businesses and limit their investments. Companies are using data-led innovation to unlock sustainability gains in line with stated commitments to net-zero carbon emissions, in line with the goals of the Paris Agreement. Turning that data into actionable insights can help power positive change while improving profitability at every level.

Strategy 1.1: Enhance ESG reporting and identify alternative sources to fill in data gaps

According to the Carbon Disclosure Project, a non-profit group that runs an environmental disclosure system, private companies trail public counterparts when it comes to reporting greenhouse gas emissions and climate risks.

Companies are also facing increased pressure from investors, lenders and regulators to disclose their environmental and social activities. Assuring sustainability disclosures that are accurate should enable the government, CSOs, funders and the private sector to better understand the impacts of economic activities on the environment. Trustworthiness of sustainable disclosures should enable companies and investors to differentiate themselves from less sustainable counterparts and attract investment.

Within the ESG framework, current ratings used to measure sustainability performance are often limited and flawed. According to an MIT study, the divergence in ESG ratings is based on a “fundamental disagreement about the underlying data.” The EU in its Non-Financial Reporting Directive has articulated the scope of wide potential audience in assessing a company’s environmental footprint encompassing investors, consumers, civil society, employees impacted by company’s activities. Many existing data institutions steward data that is or may become relevant for ESG ratings. E.g. iNaturalist, a citizen

science project empowers people to document information about flora and fauna. The Energy Data Co-op allows people to pool and share data to improve the efficiency of their homes by changing how energy is used. These initiatives could play a role in stewarding environmental data that is relevant for rating agencies and regulators which can help investors and fund managers to make informed decisions. This can also help third parties to evaluate and compare information across different companies and products. It could be possible to fill some of the data gaps from new and unique sources which can help investors to help evaluate a company or investment.

Challenge 2: *Funding challenges for citizen science projects*

Access to funding remains one of the most common challenges that citizen science projects encounter. Overcoming funding challenges is critical to enhance the capacity of citizen science and its impact on social and environmental impacts. Being largely volunteer-led initiatives, funding is crucial to the ability of citizen science initiatives to upskill, become technologically more adept, restructure for scale and thus, sustain themselves. Beyond funding, the support and guidance of investors can play a big role in ensuring the success of revenue models and dynamic adaptation to changing economic landscapes.

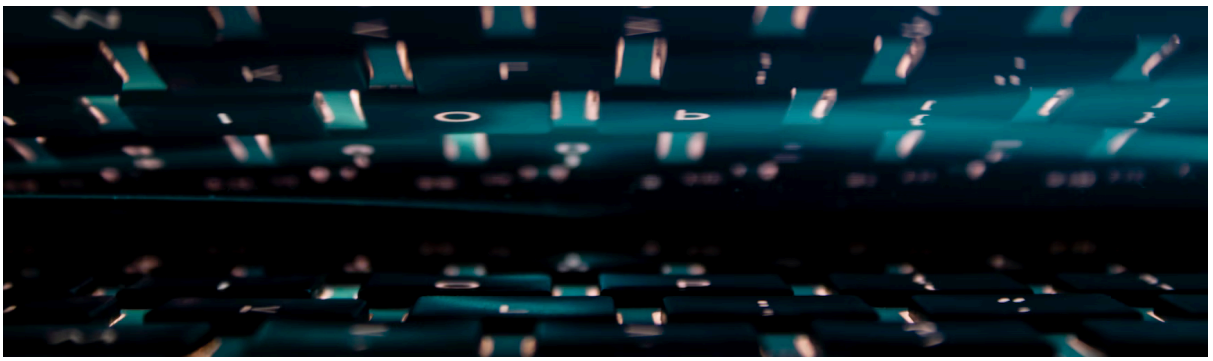
Strategy 2.1: Identify and amplify networks to facilitate funding for citizen science efforts

Persuading funding agencies and consequently, securing funding can be difficult - particularly if schemes for funding are not designed to cater to citizen science initiatives and structures. It is also crucial that citizen science organisers understand local communities and involve them in co-design initiatives, bolstering their participatory frameworks while approaching philanthropic or other funders. Considerations such as data types, representation of communities, feedback loop, insight sharing, community needs, social relevance, and strategic priorities are just as relevant to taking such initiatives out of microcosms or pilot-frozen stages.

In order to adequately cover these pre-funding bases, and to gain support in funding proposals or applications, there are currently various organisations who help citizen science organisers in getting funding. For example, Stickydot, an organisation based in Brussels helps citizens science organisers in the application process for seeking funding. Similarly, Citizen Science Network - Austria provides information on existing and completed citizen science projects, mentoring, networking and suitable partners to citizens science organisers.

With a growing number of promising stewards in the environmental space, such networks that both connect stewards and aid them in gaining access to funding are a crucial part of the ecosystem. Existing stewards and citizen science initiatives that successfully secure funding cycles can act as nodes in these networks to aid other initiatives, while funders must grow increasingly cognisant of the channels, challenges and focus areas of citizen science funding.

Play III: Technical pathways for operationalising community participation in data stewardship



Evidence demonstrates that research significantly depends on the degree and quality of public participation in project design. This calls for inclusive processes to be used in conjunction with the adoption of robust methodologies that allow concerns of citizens and local communities to be expressed. This requires integrating these concerns into the research cycle at its onset and then facilitating the participation of groups of citizens in all phases of the research process.

Play I highlights how most contemporary efforts in the data economy lack substantive and effective participation and are often exploitative. Challenge 3 in Play I looks specifically at how structural barriers and capacities play a role in hindering meaningful participation. The strategies therein suggest pathways to solve for issues of representation and diversity for stewarding initiatives.

This Play looks specifically at a key capacity in data stewardship – the technology. Given the intrinsic nature of technology when talking about data collection, sharing and management as well as the near constant use of technology in most data stewardship initiatives, it is of critical importance to understand how existing technological capacities affect participation of communities in stewardship initiatives.

Existing research has shown how various technical barriers and (lack of) capacity has impacted community engagement in digital initiatives. This points to the constant need to consider and adapt new tools and research design channels that allow collective decision-making and shared access to outputs. While selecting the technology, different criteria should be considered, such as ease of use, digital literacy, acceptance levels of tools and the extent to which the intended aim of accessibility and openness is served by the tools. This play lays down action strategies to be implemented in order to make inclusive participation of communities and individuals technologically feasible.

Diagnosis of challenges – how does lack of technological feasibility hinder data stewardship efforts?

Challenge 1: *Lack of community participation in the innovation/design process that is hindered due to technical barriers has knock-on effects for community participation in the life cycle of a stewarding initiatives*

The large diversity of technology and the lack of a model for describing the possible interactions between people, data, systems and services, is a major barrier to citizen engagement and bottom-up design. Communities are unlikely to engage, identify and then design solutions for civic problems if they do not understand smart technology and how large, complex datasets can be used by stewards for problem-solving, for large citizen science initiatives, and for environmental governance in general. It therefore becomes imperative to ensure that community involvement is made feasible. This could be done by adopting co-designing approaches. However, one must also be wary that in many cases this is also simply an issue of capacity - participatory organisations working to advocate for better environmental governance might not have the time, or the resources, to help citizens engage with data.

Strategy 1.1: Using technical tools to enable co-designing with the purpose of furthering effective participation

Employing tools to support users to design and implement robust processes can enable effective involvement and help align their objectives and interests that they have in data stewards and related activities. For example, city inhabitants were involved in the co-design of a smart city service prototype using a living lab approach. Living lab is a concept to support user driven ICT systems and its processes. A user centric model was used which contained an iterative process consisting of 3 main cycles that aimed to engage users from the preliminary stages of planning to understand their requirements and to maintain the correlation between usability and features of the application.

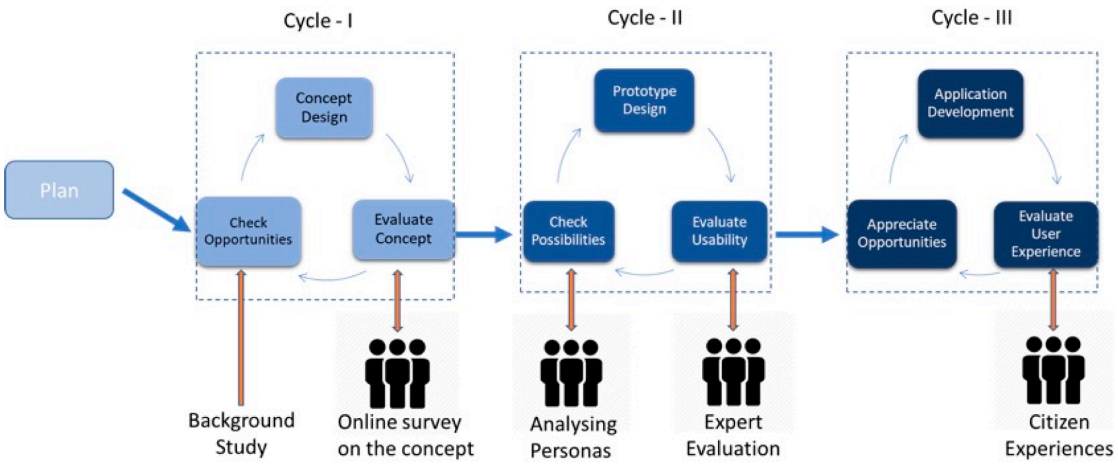


Figure: The iterative process of engaging users from planning design to developing the application

A City Commons Framework is another example of a model built on participatory action research that was developed and tested for Citizen Sensing, and designed to orchestrate large scale citizen engagement around urban issues. Voices for Change was a project that also included lay people in the data analysis stage of the research process and their strategy focused on making all aspects of the research process as participatory as possible by involving them as co-researchers.

Challenge 2: *The chosen technology and its implicit design bias could exacerbate digital divides and inhibit participation based on lack of capacity and access to such technology*

To implement these co-designing approaches to their fullest, inclusive processes must be used in conjunction with the development and adaptation of technology that will ensure that the concerns of citizens and local communities be specified and expressed. ICT skills depend on the socio-economic status, educational level, and lack of host society language skills and access to ICTs.

While designing, there must be a collaboration with developers and human-computer interaction (HCI) specialists so that design issues around data quality, sustained participation, and adoption of tools and technologies are better understood. GenderMag for instance, is an effective method that finds and fixes gender inclusivity bugs in software interfaces and workflows to eliminate bias and make the design more inclusive.

Further, web developers need to incorporate WCAG 2.0 Level AA principles - that lay down guidelines to ensure interface components are perceivable, operable, understandable and predictable. These guidelines seek to make web content more accessible and usable to a wider range of people. These principles state that information and user interface components need to be presented in a way that users can perceive, understand and can easily navigate. The overall interface design should also take into account cultural and environmental characteristics. There is a need to ensure early engagement with end users by placing the user at the centre of the design and development lifecycle..

Strategy 2.1: Employing tools that make technology and data more accessible

There is a need for the technology infrastructure to support the inclusion of community. This could be done by employing tools that enable data automation and visualisation to enhance the ability of citizens to easily navigate and perceive data that is being collected and analysed. Tools that enable training boost learner opportunities and enable learners to collaborate. Further communicating to the data subjects on data access requests will increase transparency and help build trust within the community. Tools that help in data standardisation, automation for continuous monitoring of data and comprehensive data quality assessments also need to be used. Simple tools like Google Charts (an interactive web service that creates graphical charts from user-supplied information) also provide a quick means of visualising data online as configurable charts and graphs. Tools that enable standardisation and automation increase efficiency by improving the quality of data and by increasing ease of access and use.

Challenge 3: *Lack of quality data and validation mechanisms can result in misuse and distrust rendering such data unsuitable for research or policymaking*

Lack of quality data and validation mechanisms leave room for a number of risks. These risks include not just outright misuses of personal information for private gain or to cause harm, but also the use of inaccurate or incomplete information or its transmission at the wrong time, to the wrong destination, or in the wrong transmission mode. Additionally, also of concern in the environmental space is the need to find data that is right for the particular objectives sought to be studied, i.e., data that is fit for purpose. The absence of consensus on a good strategy for de-identifying and aggregating data is a major issue for quality improvement, as it is for much public policy and research.

Strategy 3.1: Employing data verification mechanisms and maintaining extensive metadata

Data quality assessments must be undertaken which include cleaning, comparison with authoritative data, linked data analysis, semantic harmonisation and model-based validation. This could help improve the quality and veracity of the data through various phases of the research cycle. Further, extensive metadata helps create context and is helpful to communicate the 'known quality' of the data. Communicating the context in which a particular high-volume data set has been created enables data re-use. Such contextualisation and creation of metadata becomes extremely relevant when a dataset is applied for another purpose or combined with other datasets and re-used for research or policymaking.

For instance, smart city data relies on large-scale deployment of devices from multiple vendors and networks – in such a dynamic environment, the quality of data collection samples will inevitably be compromised. This could be solved with the use of lightweight dynamic semantic models that help provide interoperable descriptions of data, their quality and attributes. Thus, quality can be updated by data processing software and APIs. This is particularly useful for aggregate data that gets integrated from various sources, wherein the provenance parameters could help trace the quality of information for each source and the quality aspects of the processing algorithms applied to data.

Play III | Sectoral Guide | Environment and Sustainability

For citizen science projects to be rooted in encouraging bottom-up forms of participation and empowering communities, we need to make sure that their involvement is made technologically feasible. Implicit biases in design decisions limits an individual's or group's ability to access or use technology reinforcing existing gaps of peoples' environmental advocacy abilities and exacerbating environmental injustice. Levels of participation in citizen science range from 'citizens as sensors' (crowdsourcing) and 'citizens as interpreters' (distributed intelligence) to levels where participants are more involved in problem definition and collection protocols (participatory science) or are even part of the entire development of the scientific process (extreme citizen science).

We must consider the necessary training and scale of facilitation roles for citizen science to combine practical and theory-based knowledge for specific types of research facilitators. This means improving facilitation by learning first-hand about specific how-to guides such as design thinking, collaborative project management in combination with existing methodologies and practices from the social sciences (such as PAR). We must involve citizens and communities in a way that enables them to design and implement, jointly with scientists or in an autonomous way, valid and robust research processes.

Accessible visualisation tools can provide a better understanding of participatory research processes and outputs in terms of scientific data, as well as open online platforms for public deliberation on the interpretation of scientific results. Data visualisation is helpful for feeding analysed data back to participants and for presenting results to policymakers. Reaching consensus regarding derived actions can also benefit policies and policymakers. Play IV elaborates more on the role of policymakers.

This play lays down action strategies to be implemented in order to make inclusive participation of communities and individuals technologically feasible.The following challenges and associated play and sub-strategies offer a set of solutions and possible best practices that emerge from existing case studies.

Diagnosis of challenges – How does the lack of technological feasibility inhibit participation in the stewardship ecosystem?

Challenge 1: *Furthering of environmental injustice owing to technical barriers that prevent communities from being involved in the design process of citizen science projects lead*

Majority of climate studies (87%) practise an extractive model in which outside researchers use Indigenous knowledge systems with minimal participation or decision-making authority from communities who hold them. The lack of information and inability to use technology due to insufficient access and proficiency in the language, cultural barriers and lack of trust can inhibit interaction and consequently lead to social exclusion. Moreover, this can lead to lower quality data when there is a lack of adequate representation. Therefore, we need to involve citizens and communities in a way that enables them to design and implement valid and robust research processes, jointly with scientists or in an autonomous way. For instance, the easyRights project developed a Co-Design Approach Service design that introduces human-centred approaches to designing or redesigning services. This contains service design tools and templates to support users to take an active part in the redesign of such services. These templates use visual representations to facilitate co-design participants to immerse themselves in the user-centric approach of service design and learn from it.

Strategy 1.1: Co-designing the project for full and effective participation of communities

Co-designing becomes a prerequisite to include Indigenous knowledge for 'full and effective participation of community members - from design – which includes proposal development, defining research questions and objectives, implementation – which includes fieldwork and data collection to its analysis. Collaborative road mapping sessions with software companies can be conducted to ensure that all members are aligned with the code written, preventing time and resources in modifying the software to suit all needs. Further building interventions that rely on open-service architecture, enables the community members to easily take over the role of designers. Co-designing can also be done in a way where the citizen is prioritised by reactive scripting for design by placing them as the 'designer in the loop' thus, relying on their ability to exchange information and knowledge.

Example

One way to engage communities in a way that there is collective decision making is by co-defining problems at the outset through generation, review, and discussion of specific research questions amongst all participants. For instance, STEM4youth uses a participatory design to enable co-defining. Further conducting iterative validation mechanisms for the results of each phase such as voting on proposals. These iterative

mechanisms help encapsulate community desires and local knowledge which could be recorded to assist decision-makers in applying these community preferences.

Open systems developed tools for knowledge generation, each associated with a reflexive research stage. The tools were tested and refined during six co-creative processes using a series of activities based on a research design principle - alternate phases of divergence and convergence.

This method allows ideas to be generated in a participatory way in the first sequence and in the second sequence, the participants jointly select options; through pooling and decision-making mechanisms such as dot-voting, or dotmocracy, and thermometers of concepts.

Strategy 1.2: Incorporating indicators/tools that ensure responsible inclusion of communities

Diagrams, canvases, and gamification techniques can be used to channel citizens' social concerns and needs into the research process. These facilitation mechanisms, such as user feedback for different types of volunteers in co-created research designs provide opportunities for people lacking a voice to reveal otherwise hidden or contentious societal problems. Providing online forums to develop research topics and define priorities along with conducting on-site community workshops could prove useful to local communities, policymakers and citizen science practitioners alike.

Example

Open Humans is a community-based platform that enables personal data collections across data streams, giving individuals more personal data access and control of sharing authorizations, and enabling academic research as well as patient-led projects.

Open humans has 3 themes across its authorisation flow - members, projects and data objects. A Member can join Projects and approve them to read/write/grant access to their data. These Projects can be created by any member and a community review process was developed for Project approvals going forward; new projects are shared with the larger community for public comments, inviting feedback from all members. Members of Open Humans can also participate in the approval of new projects that want to be shared on the platform via a community review process.

The Balloon Mapping project by Public lab provides low-cost technology for communities to create high-resolution landscape imagery with various applications, such as to evaluate the effectiveness of bioswales in absorbing pollutants. The Balloon Mapping tool served as the technology infrastructure that supports community activism without the need for extensive ongoing expert assistance following its deployment.

Challenge 2: *Implicit design bias can create digital divides and barriers to access technology*

The chosen technology and its design can inhibit participation based on the skills needed to use it. The more skills necessary to participate, the less likely it is that people are to be able to participate in a meaningful way. On the other hand, choosing a process that relies on sustained and intensive community engagement inhibits people from joining later on, as these people have missed part of the process. Therefore, this urgently calls for adoption of tools and practices that support the most valuable part of any system — the people within it.

New technologies need to be adaptive to lead community actions in a successful manner - such pathways include the creation of citizen science apps, the development of guided field trips, gamification, role-playing, virtual fairs, data visualisations, or demonstrating processes that are otherwise hard to observe. For instance, VR (Virtual reality) enables critical learning opportunities in an entertaining and engaging process. A study discovered that an immersive and accessible VR environment allows learners to collaborate and provides exciting opportunities in science education and public outreach (EPO) practices

Strategy 2.1: Incorporating best practices and tools that enable responsible inclusion of communities

Citizen science projects must ensure visibility such that data collected by a particular individual should at the very least remain visible to that individual. For instance, TestMyBrain provides visual approaches for data analysis - by allowing participants to sift through images to perform tasks such as matching/classifying images. Virtual citizen science projects like EyeWire uses real time communication systems or forums to provide better support to the community.

For citizen science initiatives to function, greater usability of apps and the support provided (such as personal training) to participants are influential factors. The data infrastructure needs to be flexible enough to accommodate the diversity of data types such as text/tabular/geo-spatial and various media types -iNaturalist is a good example of one such biodiversity aggregating platform. Platforms such as iNaturalist and Natusfera allow citizen science volunteers to find biodiversity monitoring projects and directly upload biodiversity data. Citsci.org allows participants to create their own citizen science projects to initiate data collection and analysis via websites and/or mobile applications. Other platforms such as Zooniverse, provide cyberinfrastructure supporting data analysis via tasks such as classification, annotation and tagging.

Example

Web Accessibility Initiative provides web accessible tools training that web developers should aspire to meet. Tools that contribute to the design and development of components of FAIR infrastructure and platforms can be used for responsible data management planning. The conceptual model from the COST Action CA15212 was established to improve data standardisation and interoperability in citizen science activities. It utilises past models and contributes to current standardisation efforts, such as the Public Participation in Scientific Research (PPSR) Common Conceptual Model and the Open Geospatial Consortium (OGC) standards. Open Humans encourages project administrators to be clear about both data management and security in a thorough community guide including best practice details on how to communicate to participants which data access is being requested and why in clear and plain language.

The Project BudBurst website, for example, trains participants to collect and publish data and provide education materials. The project also supports a mobile application mainly designed to facilitate data collection.Citizen developers allows for people with little or no coding experience to write their own applications through the use of low code or no-code technology. The increasing availability of development frameworks help project owners create websites and other tools to support citizen science projects without the need to write complex software from scratch. At a fundamental level, WordPress, Django, Wix and Weebly are examples of frameworks that provide means for interacting with participants and also support responsive design to deliver content appropriate for display on mobiles, desktops and tablets. For more advanced users, frameworks such as PhoneGap and Ionic help developers write websites in HTML and JavaScript. Responsive design enables websites to be viewed according to the device being used to access them, by adapting layouts, media items and other content to different resolutions and screen-sizes. For projects that seek to host a website and a mobile site or app, styling tools such as Bootstrap and Boilerplate help simplify this process.

Challenge 3: *Lack of quality data and validation mechanisms can result in its misuse and distrust that would render such data unsuitable for research or policymaking*

Data from citizen science is immensely valuable as it provides evidence that professional science cannot gather. However, aspersions cast on the collection method and its validity by citizens is the main reason that it faces distrust and scepticism from scientists and policymakers. The misuse of data is likely to occur when there is inaccurate data documentation, this could lead to overcorrection which can cause more errors and ultimately exacerbating distrust and suspicion of all citizen science data. This has caused many to discard it as unsuitable for research purposes. From a more general research design perspective, the validity and the reliability of data are most important. Datasets should be consistent and must have adequate distribution of the target population area. Especially in citizen science projects operating large datasets; reliability and quality ensure trust and aligns with policy requirements and stakeholders' interests.

Strategy 3.1: Maintaining data quality, employing verification mechanisms and contextualising data

To ensure a minimum standard of data quality, a plan or protocol of data collection must be set out at the start of a project. This is to manage different levels of expectations of data quality from different stakeholders. All stakeholders should be invited to co-develop standards for data quality and explicitly state the data quality levels they expect in order to form an agreed approach to data quality (citizens, scientists, researchers, funders and policymakers) through online forums.

Contextualisation i.e., communicating the context in which a particular high-volume data set has been created enables data re-use. Metadata, attribution, and curation are the most prominent examples of data contextualisation. Appropriate documentation and metadata are the most effective deterrents against using unsuitable data. Metadata helps create contextual information – can include title, number of participants, contact details, data lineage and geographic extent of data. This helps in assessing the quality of data as data points such as identity of observer or location accuracy can be determined. Further, contextualisation and creation of metadata becomes extremely relevant when this dataset is applied for another purpose. When these datasets are combined with others to be re-used for a different purpose (like policymaking), the context needs to be carefully examined

Preemptively restricting data inputs can help identify unforeseen sources of errors or other problems that can be fixed before the project starts. This can be done through

methods such as profiling which helps the data collectors to understand the quality challenges and pre-testing by gathering sample data before a citizen science project begins using both expert and beginner contributors. When new data is submitted, effective data validation mechanisms need to be implemented such as averaging data records, flagging errors and providing feedback to the users. These mechanisms can improve the quality of data and user trust in the data.

Automatic quality assessment involves the use of software-based systems to carry out a quality assessment of the data generated or collected by a citizen science project. There is a wide range of approaches, such as data mining algorithms, which filter and search for problematic data, statistical analysis (plausibility of data), and qualifying systems

Model-based quality assessment goes beyond automatic filtering technique and tackles residual errors using an explicit model of how the phenomenon of interest is expected to vary in space or time. Examples include de-biasing procedures and generation of contributor ratings, based on identified sources of systematic errors in the archive of observations.

Example

The PDI laboratory information management system (LIMS) tracks samples, receives instrument output, stores metadata, and produces chain-of-custody certification, thus standardising and automating processes. For instance, in Puerto Rico a data collection app (for animal identification) was developed by PDI (Partnerships for Data Innovation, which provides scientists with state-of-the-art technology and tools for robust data management) wherein data uploaded could be accessed instantaneously by scientists and was extremely effective in automatically calculating proper dosage and other relevant information - reducing post processing time and the margin of error.

Strategy 3.2: Involving participants in data collection methods

Comprehensive data collection protocol that covers all stakeholder expectations must be implemented. Poorly designed or overly complex protocols can also create skill inequality if some protocols assume a specific level of scientific training before they can be used carrying the risk of excluding whole communities.

A possible solution is to introduce a permanent channel or forum that participants can use to contact creators and provide input. Tasks of data collectors can be made more straightforward by pre-filling files with often used values and providing examples for how observations are recorded. These data collection tools need to have a simple user interface design and must be compatible with a variety of skills and objectives of potential citizen scientists. Integrating simple and affordable visual ways by use of

images and gamification features to communicate concepts which are often complex or specific to the scientific world.

Programs with very simple data collection methods report much higher rates of participation, but data is often biased and noisy. Therefore, using simple techniques such as data collection forms or semi-structured surveys reduces the need for rigorous training, at the same time ensuring that data is collected in a prescribed format. For instance, Ushahidi and Open Data Kit (ODK) provide a way to easily develop customised surveys and set up websites and mobile applications that can be distributed to crowdsource information. These frameworks also allow project owners to aggregate, visualise and analyse the data collected.

There are a range of tools available to support project co-ordinators and volunteers in processing, analysing and visualising data - Earthwatch's Freshwater Links and UCL's Extreme Citizen Science: Analysis and Visualisation (ECSAnVis) are examples where users can visualise data coming from a variety of remote databases. The ECSAnVis project focuses on the creation of software called Sapelli that runs on mobile phones and allows non-literate people to collect information about their area with an app that participated in the design that is suitable for such users. This software was tested on the community and was built considering those who have little or no understanding of tech, but could still fully take part in the process of building research questions and collecting data.

Example

The main tool for data collection in D-NOSES to collaboratively build odour maps is the citizen science open app OdourCollect. Communities affected by odour issues can map them collaboratively so they can be viewed by all. The project enhances digital inclusiveness by creating tools for ensuring inclusiveness in specific social environments : for example, for the elderly or people with difficulties in accessing digital technologies. The D-NOSES methodology seeks to empower citizens and key stakeholders to generate, access, and use data related to odour pollution. The collected data is then used to inform and co-design possible solutions to manage odour problems

Data collection tools by SPOTTERON provides for data infrastructure including handling and storage – data analysis toolkit.

Play IV: Ecosystem enablement – role of the public sector

Unlike the previous plays, this play focuses on the role of a specific stakeholder – the public sector. This focus on the public sector is a result of the critical role the public sector has in ensuring the growth and sustainability of data stewardship initiatives. To start with, data stewardship needs a robust policy ecosystem to aid its functioning. Recognised rights over data, established means of grievance redressal, and protocols for collection and sharing of data are all key tools that data stewards leverage in their work. These tools are best put in place through policy action by the public sector. The Subject Access and Data Portability rights in the EU GDPR and UK's Data Rights Act, 2018 (which mirrors rights from the EU GDPR) provide individuals with the right to access data about them held by a company and have an explanation as to why they collect this data, who and who they share it with. These rights have had far reaching ramifications for Uber drivers in the UK and Netherlands who were able to access their data held by Uber to prove that they were entitled to benefits as regular employees / workers and were not self-employed contractors.

Policy action can also serve to inhibit the working of stewards. For example, the EU Data Governance Act's characterisation of rights under the GDPR in the context of data cooperatives (recital 31) has led to debate around the ability of data cooperatives to function to their full potential. It notes that these rights are personal rights of the data subject that cannot be waived. A possible interpretation of this is that data principal's cannot delegate their rights to a data cooperative. This evident ban on data mandatability has had serious repercussions for the functioning of data stewards - as they will now be unable to access data rights on behalf of their members. In order to function, stewards will need to either collect data from members directly through an application - something that is expensive to do - or have members individually request data from companies - something that dissuades members from joining a steward. It must be noted that this interpretation would be in sharp contrast to the broader data

protection goals of the EU. Nonetheless, what is clear is that the public sector - through its role as a policymaker - has immense sway in the sustenance of data stewards.

Policy making is not the only sphere of influence for the public sector. Active support (financial and otherwise) from the public sector can be a major boon to data stewardship efforts that either don't yet have a self-sustaining business model or where access to private funding is not easily available.

While some of the previous plays have addressed aspects of the public sector's role, this play looks at the challenges faced with the engagement of the public sector with stewardship and community-oriented initiatives and suggest possible pathways for redress.

Diagnosis of challenges – What are the challenges around public sector engagement with the stewardship ecosystem?

Challenge 1: *Funding efforts tend not to focus on community-oriented and community-led projects – funding programs typically incentivize institutional data collection efforts which don't look at the community as an active participant*

Our research – including desk research and conversations with data stewards and other community-oriented organisations - highlighted that funding for community-oriented efforts from the public sector was a point of serious concern, and this was echoed in the conversations we had with organisations in this space. Issues highlighted by organisations include limited access to funding from public sector sources as well as majority of funding going towards efforts that were directed at more institutional collection of data that does not involve citizens in the collection and contribution process. These efforts tend to miss out on data points that realistically cannot be obtained through other means.

Strategy 1.1: Instituting funding programs with a specific focus on data stewardship efforts

Public sector organisations can look to set up national or regional level funding programs aimed at funding data stewardship efforts specifically. In doing so, the programs can have separate funding packages for different focus areas – thus helping to address multiple areas that suffer from a lack of data collection, as was the case with the UK government's data trusts program, that funded stewardship initiatives aiming to tackle illegal wildlife trade and reduce food waste. These funding programs can be set up by national or state governments as well as public sector organisations, or multilateral bodies. For example, the Global Partnership on AI – a multilateral organisation comprising various national governments – has funded research projects to look at the viability of data trusts to tackle climate change related issues. While the division of funds set out by national or state governments can be for broader thematic areas / practices, funds from public sector organisations can be divided to specific sub-groups within their practice domain, thus helping address data gaps at a minute level. In order to ensure that such funding efforts are stable, it is also important that they are a sustained policy effort and not one-off projects. The European Commission, with its Horizon 2020, launched a program that, inter alia, funded community centric research and innovation efforts. This was succeeded by Horizon Europe, a program with a similar objective. These programs have funded initiatives like the Community Observation Measurement & Participation in AIR Science (COMPAIR) that focuses on increasing citizens' capacity to monitor, understand and change their environmental impact in relation to air quality. The availability of long-term funding for these efforts

can go a long way in ensuring their sustainability. Community Observation Measurement & Participation in AIR Science (COMPAIR) that focuses on increasing citizens' capacity to monitor, understand and change their environmental impact in relation to air quality. The availability of long term funding for these efforts can go a long way in ensuring their sustainability.

Strategy 1.2: Encouraging private entities to fund data stewardship efforts

While the public sector is a key source of funding for data stewardship efforts, it is not the only one. As outlined in other plays, the private sector also has a major role to play in funding and sustaining community-oriented models. While certain philanthropic funds provide funding for data stewardship efforts, the public sector can also enact policy measures to increase the funding of private entities in data stewardship efforts. A key way this can be done is to include data stewardship efforts as a specific activity within Corporate Social Responsibility (CSR) legislations, thereby providing an incentive for corporates to fund such efforts. Further, public sector institutions can also provide certification or validation to data stewards that have a proven track record, assuring corporations of the quality of work of such organisations.

Challenge 2: *There is a disconnect between community actions and policy measures – engagement of the public sector at a grass-roots level is often missing, and when present, the information provided by communities is not often reflected in policy measures*

Citizen generated data has enormous potential to become valuable sources for policymaking. To fully utilise its value, governments need to assess whether the data generated is fit-for-purpose to make informed decisions. To this end, governments need to consider the quality, interoperability and formats of the data and collaborate in a way that such citizen generated data contributes to policy making. The onus is on the policy makers to meaningfully engage such that these efforts are aligned to policy objectives – this can happen through deep collaboration and by building trust with the community.

Engaging with a community provides a public sector actor with a more grounded and realistic understanding of the issues facing the community, thereby allowing them to get a better grasp of the sort of data that needs to be collected.

Partnering with local organisations / partners in the community can also address issues of capacity that a public sector actor might face, and in many cases provide a more accurate landscape of the issues – citizens might not feel entirely comfortable in providing public sector actors with the truth, or in some cases might exaggerate the situation in the hopes of seeing expedited action.

Strategy 2.1: Recognising citizens as a key stakeholder in data collection that influences policy making

Stakeholder engagement / public consultations are a major component in designing policy that is reflective of the needs, concerns and capacities of the relevant subject area. However, in many cases, these consultation processes are not truly bottom-up, and lead to the exclusion of key stakeholders. For example, when the Indian Ministry of Agriculture and Farmers' Welfare sought to implement a digitisation initiative in the sector in India, they released a foundational document outlining their vision and invited comments from the public on this. However, not only was this document released without any consultation with farmers – the one stakeholder most likely to be affected by the policy – this document was released only in English, in a country where most farmers are not literate in English. This put the onus on the civil society to either translate documents into languages that farmers could access, or to explain the policy measures outlined in the document. . Policy design processes can be exclusionary in numerous other ways – but a focus on identifying the key communities and stakeholders likely to be affected by policy measures, and then engaging actively with

these communities at various stages of the policy making process right from inception, can address this issue. Hearing, firsthand, the needs and challenges of the community in building data stewarding efforts can go a long way in making the consultation process more inclusive and result in a much more effective policy. Finally, efforts such as the EUROCITIES Citizen Data Principles are also very effective in displaying the commitment of public sector actors to engaging with citizens and instituting inclusive engagement and consultation practices. This has been taken a step further with the DECODE Project in the cities of Amsterdam and Barcelona, where actual tools were created to allow citizens to allow their personal data to be used for public good, on their own terms.

Strategy 2.2: Incorporating community generated data and learnings in policy measures

Collecting data that is then not put to use, or does not result in change in policy, can often have a demoralising effect on citizens – disillusioning them and reducing participation in future efforts. Policies that are tuned to address the needs and challenges expressed by communities – and represented in the data and information they collect – can go a long way in encouraging further participation from communities. A great way to demonstrate that the policy has been informed by community actions, is to provide explanatory notes to policies that provide the rationale for a particular policy measure and include data and information shared by the community as supporting evidence for the nature of policy action being proposed. By showing the community that the public sector is willing to believe in the data collected by the community, and act on it, responsive policies also help increase trust in public sector actors. For example, the European Union released a research report that explored various citizen science efforts with the aim of providing the European Commission with an evidence base of citizen science activities that can support environmental policies in the European Union. The report ended with recommendations on how to leverage the contribution of citizen science to environmental policy. Such policy directives, taken in tandem with funding efforts such as Horizon Europe (mentioned above) can only serve to spur new data stewarding efforts.

Strategy 2.3: Implementing robust and accessible data quality practices and standards

A major issue with community generated data often highlighted by public sector actors is the quality and authenticity of data. This is a fair concern, as citizen generated data is often through low-cost and accessible digital technologies that may not necessarily be of the required standard for policy implementation. Policymakers need to make sure

that validation and quality assurance methods are employed by data stewarding initiatives. To ensure a minimum standard of data quality, a plan or protocol can be set out by policymaking bodies or sectoral regulators that lay down standard operating procedures and quality assurance methodology that can be followed. In doing so, the relevant actors can also provide toolkits that translate technical standards and processes into language that is easier to understand for everyday citizens, thus improving accessibility. The United States' Environmental Protection Agency provides a great example in this regard with their Citizen Science Central Toolkit and quality assurance methodology.

Challenge 3: *Access to information and data collected by the public sector is very limited – in many cases the data itself is not made available and in cases where it is, access is complicated by the modalities of access*

Easier availability of public sector data opens doors for countless opportunities to enhance targeted and informed research efforts by data stewards. This is steadily being recognized by policymakers world over as they look to make more public sector data available to encourage research and innovation. However, data held by the public sector is often retained in closed silos for numerous reasons. In cases where data is made available, the problem of access is exacerbated by the modalities of access which are often exclusionary or involve bureaucratic processes – disincentivizing community members from attempting to access such data.

Strategy 3.1: Creating an infrastructure that leverages access to public sector data and enables open data exchange between public sector and citizens

Policymakers need to create an ecosystem to enhance greater collaboration between government and stakeholders. Policymakers must release information at required quality levels to improve the widespread reporting efforts. The use of open digital infrastructure and tools can promote access to environment data and facilitate collaboration with various stakeholders. Using shared standards and APIs can boost the integration of citizen generated data with official datasets. Using open-source software in citizen generated data would also allow full control of procedures and workflows, which enhances reliability and encourages open transparent and fully documented practice.

Challenge 4: *Private sector involvement in data governance for good is lacking - Information collected by the private sector is siloed, with either no access, or high access costs, and compliance of private sector with regulations is problematic*

An oft-highlighted problem with the digital economy as it stands today is the fact that significant data collection efforts are being carried out by private companies that wall off these datasets in private silos. The data that companies collect can have immense value to the public – for example, mobility data collected by companies like Uber, Ola, and Lyft can provide governments and city planners with valuable insight to design better public infrastructure and tweak policy for improved sustainability. While Uber has taken a small step towards this by sharing some of the data it collects through the Uber Movement platform, this initiative is nowhere near enough. It does not provide information on where people start and end most of their trips, which is key to understanding commute patterns. Additionally, the Movement platform shares data only for a select handful of cities, despite Uber's large global presence. Such siloing of data in private hands is playing out across various sectors including agriculture, pharmaceuticals, and energy. This trend of siloed data is no surprise however, given the distinct lack of incentive on private entities to share this data.

Complicating the issue of private sector involvement in data governance for public good is their compliance with regulations designed to provide citizens with greater control over their data. For example, even though the EU GDPR recognises rights of access, correction and deletion, the modalities of this can be very difficult, designed in a way to dissuade from effectively exercising these rights. A report by Worker Info Exchange highlights the problematic behaviour of companies in responding to data access requests, with companies often providing data in non-machine-readable formats or not even providing certain data that was requested. Additionally, while such rights are critical to empowering individuals over their data, they presuppose a degree of literacy, interest and ability to assert data rights, which are often missing, especially in Global South contexts. While data stewards can play a key role in enabling citizens to assert their rights, they need to be empowered to do so.

Strategy 4.1: Incentivise the private sector to share data with the public sector and data stewards

The public sector can play a crucial role in improving the availability of privately held data by incentivising the private sector to share data with the public sector. The simplest method that we see this taking place in is through regulation that mandates the private sector to share data. For example, in India, the 'Karnataka On-demand Transportation

Technology Aggregators' Rules require cab aggregators to provide records of passenger details, trip origin and destination, and fare collected on demand from authorities. Similarly, the Convention on Biodiversity, an international treaty, has set up a system that allows for fair and equitable sharing of genetic resources. This, along with the Bermuda Principles, have been key in the rapid development of a COVID vaccine through the open sharing of data. The Committee of Experts on Non-Personal Data Governance Framework in India, in their report, highlighted the need for a mandatory regime of NPD sharing in public interest. While such a mandatory regime might be problematic and does not adequately account for business interests and intellectual property, our research has shown that an ecosystem-based voluntary approach to data sharing which focuses on enabling infrastructure, incentivising sharing, and adopts a voluntary structure at its core can be extremely beneficial in getting the private sector to share data for public good.

Strategy 4.2: Leverage data stewards to ensure accountability in data sharing

While incentivizing data sharing through regulation and improved infrastructure are crucial steps to improving availability of privately held data, they are not always enough. In the case of the Karnataka cab aggregator rules mentioned in the previous strategy, a lack of enforcement has meant that there is no easy public access to the data collected by cab aggregators. Similarly, as highlighted above, individuals are facing it hard to assert their rights under the GDPR, with companies engaging in bad faith practices that disincentivise individuals from further engaging with their rights. This stems from a combination of a lack of well-defined regulations as well as a lack of capacity from regulators.

Rather than overburdening regulators, governments can instead empower verified, trusted, independent data stewards to perform the role of an intermediary that ensures accountability. Regulations that provide data stewards with the ability to request information from private companies, under strict guidelines, can go a long way in ensuring accountability in data sharing. For example, Worker Info Exchange facilitates data requests on behalf of gig workers from companies such as Uber, Ola and Deliveroo. In doing so, they take away the burden of data requests from individual drivers and are also in a better position to ensure compliance with such requests from companies. Optery is another example that facilitates data deletion requests for consumers.

Regulation can also be designed in a manner to support stewards to do this. While the GDPR is largely silent on the ability of an individual to delegate their rights, and the Data Governance Act is contentiously worded on data principals' ability to delegate their rights from the GDPR, academics have assessed legal bases for how delegation of data

rights can be carried out under the GDPR. Legislative clarity on this will empower data stewards to better assert and enforce rights of individuals with their consent, in their interest.

Play IV | Sectoral Guide | Environment and Sustainability

How can the public sector support citizen science efforts?

Organisations that engage in citizen science endeavours face numerous challenges. Access to adequate funding in this ecosystem is difficult, and even with access to funds, organisations tend to spend a lot of it on public engagement and promoting their efforts in a bid to involve a greater number of people in their activities. And when they manage to collect a sufficient amount of data, many citizen science organisations also face the problem of the data either being discredited or disregarded for use in designing accurate and responsive public policy. This is not entirely without reason, as concerns have been expressed within the scientific community numerous times regarding the accuracy of citizen science data. However, these issues are not beyond remedy and, as some public bodies have observed, citizen science can play a vital role in filling crucial data gaps in the public sector.

The public sector, with its ability to address a number of these issues, has a crucial role to play in ensuring the promotion and sustainability of citizen science efforts. This play outlines certain areas through which the public sector can get involved with and support the growth and sustenance of citizen science efforts.

Challenge 1: *The public sector's role in funding citizen science efforts*

Similar to the general strategy, specific public sector funding for citizen science efforts can address numerous issues. Set out below are a couple of strategies that public sector actors can follow to improve funding towards citizen science efforts.

Strategy 1.1: Instituting citizen science funding programs

Funding citizen science efforts can have multifaceted benefits to the public sector beyond just filling data gaps. This of course serves as a primary benefit as funding can aid existing efforts and encourage fresh citizen science efforts that can fill critical gaps faced by the public sector, as noted by the National Statistical Offices of Kenya and Ghana. Funding can also address other issues that are faced within the realm of citizen science efforts. Guaranteed funding for specific focus areas can ensure that citizen science organisations do not pivot to a different area. As a condition precedent to providing funding, actors should require grantees to sign contracts that stipulate certain conditions that will ensure basic levels of data quality, citizen protection as well as settle matters relating to intellectual property of the data collected. While public sector actors should look to fund models that have the potential for scaling or implementation in different contexts, this must not serve as an overriding concern, as data collection in certain areas can require specific types of models.

Example

Multiple national governments have funds set up with the specific objective of funding citizen science efforts. Prominent examples include CitizenScience.gov in the United States (which provides a catalogue of federal funding opportunities and ongoing citizen science initiatives) and Citizen Science Grants in Australia. Multiple public sector actors in the United States have also set up sectoral funding for citizen science efforts, such as the CitSci Fund by the US Department of Agriculture's Forest Service, designed to involve citizens in better management of forests and conservation of resources. NASA's Citizen Science for Earth Systems Program (CSESP) funds citizen science efforts to enhance NASA's Earth observing endeavours. This is a specific fund set up by NASA within its broader Citizen Science Projects program. These websites are also key resources that provide a glimpse into the way the contracts with grantees can be structured.

Strategy 1.2: Encouraging private entities to fund citizen science efforts

While the private sector does have a history of funding community-oriented programs in the environment space, these efforts can sometimes skew away from meaningful

citizen participation. Citizen science efforts can address these issues, and encouraging private entities to fund such efforts can boost their growth and sustainability.

Example

Corporates are already funding citizen science efforts in various parts of the world. e-Mammal India is an effort by the organisation Sahyadri Nisarga Mitra that seeks to promote scientific understanding among children of age group between 11-14 predominantly from rural and tribal schools of Maharashtra, India while also obtaining crucial information regarding elephants. By training students to install and use camera traps, the project documents the population size, activation patterns and habitat use of elephants in the area. The project is supported by ICICI Bank Ltd. While the Indian Companies Act, 2013 provides that funding by corporates for "ensuring environmental sustainability, ecological balance, protection of flora and fauna, animal welfare, agroforestry, conservation of natural resources and maintaining quality of soil, air and water" as a valid activity for CSR funding, a specific call out to citizen science efforts that aid the aforementioned aims would be a welcome addition that encourages corporates to fund citizen science activities.

Challenge 2: *Partnerships and engagement with communities*

Strong engagement by the public sector with communities for citizen science efforts can serve to aid both public sector actors and communities gain a better understanding of environmental issues. However, as was pointed out by experts we spoke to from organisations such as the NRDC and WaterCAN, public sector actors can be very hesitant or averse to engaging with communities or partnering with community-level organisations.

Involving the community in data collection efforts can help provide the community with a stronger understanding of the issue they are facing, along with possible causes and / or impacts arising from it. Data collection efforts carried out by community members also help spread awareness within the community and in cases can also see communities themselves alter their decision making and planning in order to tackle some of these issues. Strong engagement of the community by the public sector in data collection efforts that then translate to responsive policy outcomes serve to strengthen trust between the community and public sector actors. For the public sector, citizen science data can be a valuable source of information for policy making and can contribute to policy formulation, implementation, monitoring and evaluation. However, data quality is critical in the context of policy making - and actors need to ensure the quality of data and assess whether they are fit for purpose in order to make informed decisions.

Strategy 2.1: Building tools for community participation

While citizen science efforts are gaining traction, a barrier to broader engagement is the lack of accessible literature on how citizens can help and participate and tools that allow them to contribute. A large portion of literature tends to be oriented towards scientists with training and are thus too technical for average citizens. Similarly, many tools used for data collection or monitoring tend to require a degree of knowledge. Even a repository such as github, which is considered quite basic by those in the science community, can be difficult to navigate for average citizens. Designing guides, playbooks and toolkits that provide easy to understand information on the methods of data collection, how to use the relevant tools, and the ways the data will be put to use can vastly improve community participation by lowering the knowledge barrier to entry.

Example

The United States Environmental Protection Agency has a repository of information directed towards empowering citizens to involve themselves in collection of environment data. This includes a Citizen Science Central Toolkit, guide on hazardous

air pollutants, quality assurance methodology, and a guide with a general description of several different types of air monitoring and sampling equipment.

Strategy 2.2: Incorporating community generated data and learnings in policy measures

Collecting data that is then not put to use, or does not result in change in policy can often be a demoralising effect on citizens – disillusioning them and reducing participation in future citizen science efforts. While this is outlined further in the next area, proper utilisation of reliable data generated by citizen science efforts to inform policy can serve to encourage citizen participation in citizen science efforts. This demonstrates to them that the public sector actors are willing to accept the data collected by the citizens and take the required action on this basis, thus empowering them to further participate in actively making their community better. It can also encourage communities in other areas to undertake similar efforts while serving as proof of concept to other public sector actors to engage with more communities.

Example

Owing to a paucity of observed data due to the limited number of weather stations, the Ghana Meteorological Agency (GMA) faced issues in providing accurate weather forecasts and verification. To combat this, the GMA launched Whatsapp based platforms called "Let's Talk Weather in Ghana" where citizens can provide information on weather forecasts and feedback on weather events in Ghana. This feedback is then used by GMA to evaluate and verify forecasting data, and the information from citizens has helped the GMA fill in gaps. Crucially, citizen scientists were happy to see that their photos and observations contributed to the GMA weather forecast capability, with increased interest in this effort leading to a growth in the number of Whatsapp groups. It also led to an increase in the trust between the GMA and local citizens.

Strategy 2.3: Collaborating and co-designing at the planning stage of the citizen science project

Policy makers need to make an assessment on how a citizen science approach can contribute to the policymaking process. Based on this assessment, the decision to involve and co-design can be taken and management mechanisms can be formulated. This assessment must ideally involve a cost-benefit analysis taking into consideration factors such as impact of the project, time, resources and expertise required. Most importantly, policymakers must factor in sustainability and the long-term impact for society. Therefore, it becomes imperative for policy makers to think through different

pathways of change during the planning stage as part of results-based project design. This deep collaboration and co-designing can ensure meaningful engagement between citizens and the public sector thereby aligning objectives of those participating in the project as well as that of the policymakers. Independent bodies like NGOs can facilitate this building of trust between policymakers and the public. This puts a clear responsibility on citizen science practitioners and policymakers to manage participants' expectations and not overpromise the impact an initiative will have.

Examples

COBWEB is an example of a citizen observatories platform for data collection and sharing that uses standards and interoperability principles. The Welsh government has been involved since the start of COBWEB and contributes the views of government and decision makers to the development of the project. D-Noses, a citizen science project that built odor maps was a result of collaboration between local governments and other stakeholders The Scottish Environment Protection Agency (SEPA) has developed best-practice guidance on appropriate design of citizen science projects to support public authorities and access to citizen science design tools. It also identified where citizen science efforts could contribute to regulatory efforts of the public sector.

Strategy 2.4: Contextualising data and maintaining quality data standards

Policy makers need to evaluate the value of citizen generated data by taking into consideration quality, interoperability, format and scalability. To ensure a minimum standard of data quality, a plan or protocol must be set out. Policymakers need to make sure that validation and quality assurance methods are employed. They need to be aware of and take into account potential biases while making decisions. Further, for the data to be relevant for policy, it has to be highly contextualised and adapted to the level of intervention. The use of metadata and established spatial data infrastructure (SDI) initiatives can increase accessibility, openness and support the uptake of data across various scenarios and add immense value to decision makers. More extensive metadata is helpful to communicate the 'known quality' of the data. The policymakers need to record this metadata using a published standardised approach that allows assessment of the project.

Example

In the UK, the Environmental Observation Framework (UKEOF) has formed a citizen science working group to share good practice and improve environmental data quality. It has a step-by-step guide on a tool to assess the costs and benefits of citizen science.

Challenge 3: *Support citizen science efforts through easier availability of public sector data and transparency*

A lack of easy access, and in many instances any access, to public sector scientific data is a hurdle to citizen science efforts. Availability of public sector data to the public would also encourage transparency – thereby improving trust – as well as improve the reliability of data. However, very few public sector actors have made their data sets public, and in cases where they have, there are gatekeeping criteria that often make access problematic and a bureaucratic process.

Strategy 3.1: Creating a digital ecosystem-based infrastructure that leverages access to public sector data and enables open data exchange between public sector and citizens

Data held by the public sector must be released and made accessible to the community to streamline and improve widespread reporting efforts in citizen science. This would encourage transparency and enhance reliability of data- enabling policymakers to fully integrate citizen generated data into their official datasets to inform their decisions. A collection of these resources, when made available to the community, not only builds trust but also empowers them to lobby for better action by the government.

Examples

The EPA includes actions on how to contribute to streamline environmental reporting as well as how to share information on best practice and lessons learned among its members. The US Government's environment justice tool is one of its kind open-source tools and provides public forums to have discussions to discuss what data should be included and help troubleshoot issues. Government agencies solicited input from the public on finding reliable data sources to measure metrics such as pollutant levels. Green Paper Citizen science strategy for 2020 in Germany presents the understanding, requirements and potentials for citizen science. Germany developed a central citizen science platform which helped in building the citizen science community in Germany, in collaboration with the capacity-building program for citizen science, GEWISS. It includes various support tools, such as guidance and development of quality procedures for citizen science projects.

Strategy 3.2: Integrating citizen generated data into official datasets

Botellón no me deja dormir, is a platform that mentions the benefits of having a collection of resources that people can trust - necessitating the need for policymakers to

have API platforms and to enable sharing, calibration methods and cleaning the data to make it more robust. While assessing the value of data produced, especially in low-income communities - the government must look at it as a means of empowering and engaging communities. Therefore, rather than necessitating adherence to rigorous protocol and sustaining them for long periods of time - citizen science projects can behave as tools to lobby for better services and data collection by government agencies as these projects would be best positioned to identify gaps in the public sector data collected. The need of the hour is to meet the citizen science projects halfway to effectively leverage instead of bombarding these initiatives with additional monitoring process requirements.

Examples

The European Biodiversity Observation Network (EUBON) showcases how the public sector sought to fully integrate biodiversity data into decision making. The Scottish Environment Protection Agency (SEPA) developed best-practice guidance on appropriate design of citizen science projects to support public authorities and access to citizen science design tools. It also identified where citizen science efforts could contribute to regulatory efforts of the public sector.

Challenge 4: *Questionable veracity of data provided by private entities*

Data provided by private entities in the environment space has been repeatedly called into question. Environmental impact assessments (EIAs) which are meant to protect and preserve the environment are failing in this promise. In many cases, approvals are granted despite independent studies contradicting the data provided in the EIAs – leading to increased scrutiny over the credibility and accuracy of EIAs. This in turn calls into question the competence and capacity of regulators to effectively monitor and enforce regulations. Additionally, it also places serious doubts over their independence. This is furthered by attempts, such as in India, to create EIA systems that lack transparency. This is also replicated in mechanisms such as Carbon Offsets where project documents approved by verifying agencies in the US containing misleading information. The situation in this particular case is exacerbated by the fact that verifying agencies are not officially regulated by the US Government.

Strategy 4.1: Leveraging data stewards to improve accountability

Be it reduced trust in regulators, or a lack of capacity within government agencies to enforce regulations, the public sector can tap into the potential that data stewards and citizen science efforts provide in ensuring accountability. Trusted citizen science efforts can be empowered to independently verify information provided by private entities in mechanisms such as EIAs or Carbon Offset approval documents – with this providing multifold benefits of ensuring veracity of information provided and building trust in government agencies.

Action	**Yes**	**No**	**N.A.**
Do you fund citizen science efforts?	☐	☐	☐
If yes, is this part of a general fund?	☐	☐	☐
Is there a fund dedicated specifically to funding citizen science efforts?	☐	☐	☐

Does the fund specify focus areas within citizen science for which funding is available? (as opposed to being a fund for any type of citizen science effort)	☐	☐	☐
Are there strict monitoring and reporting requirements as part of the terms and conditions of funding?	☐	☐	☐
Do the terms and conditions of funding recognize IP rights over the data collected by the citizen science efforts?	☐	☐	☐
Do the CSR regulations in your jurisdiction allow for funding of citizen science efforts to be considered as CSR?	☐	☐	☐
Is there a specific call out of citizen science activities as one of the eligible fields for CSR funding?	☐	☐	☐
Are there any other forms of incentives for private sector actors to fund citizen science efforts?	☐	☐	☐
Do you have tools that make citizen science efforts more accessible to the general public?	☐	☐	☐
If yes, have there been consultations / engagements with citizens to gain feedback regarding these tools?	☐	☐	☐
Do you use citizen science data to inform policy measures?	☐	☐	☐
Has citizen science data resulted in any changes in policy measures?	☐	☐	☐

Do you conduct consultations / engagements with citizen scientists to understand their views on policy measures that can be affected by their efforts?	☐	☐	☐
Has there been any identification of data gaps that can be addressed by citizen scientists?	☐	☐	☐
Have you partnered with community organisations to carry out citizen science efforts to address such data gaps?	☐	☐	☐
Do you make an assessment of citizen science project on how it can contribute to policy making?	☐	☐	☐
Have you conducted a Cost Benefit analysis to determine extent of collaboration (factor in sustainability and whether it empowers communities)?	☐	☐	☐
Have you made the decision to co-design and collaborate with the project, based on the above assessment?	☐	☐	☐
Have you built trust and methods to ensure meaningful engagement with citizens to align with their objectives for participation?	☐	☐	☐
Have you developed best practice guidelines and standards on appropriate design of citizen science project?	☐	☐	☐

Have you managed the expectations of citizens on impact of project?	☐	☐	☐
Have you ensured minimum standards of data quality are employed?	☐	☐	☐
How to increase accessibility and openness of metadata for contextualization?	☐	☐	☐
Have you supported citizen science efforts through easier availability of public sector data and transparency?	☐	☐	☐
Have you used open digital infrastructure and tools to enable access to environmental data and facilitate collaboration?	☐	☐	☐
Have you used shared data standards and APIs to boost integration of citizen generated data with official datasets?	☐	☐	☐
Have you encouraged use of open-source software in citizen science projects to increase transparency and document all practices?	☐	☐	☐

Do you use calibration and cleaning methods to ensure data is robust while integration?	☐	☐	☐